VII

*Applications to behavioral science problems

1. COMMUNICATION AND SOCIOMETRIC MATRICES

Matrices having only the entries 0 and 1 are useful in the analysis of graphs and networks. We shall not attempt to give a complete treatment of the subject here, but merely illustrate some of its more interesting applications.

A communication network consists of a set of people, $A_1, A_2, \ldots, A_n$, such that between some pairs of persons there is a communication link. Such a link may be either one-way or two-way. A two-way communication link might be made by telephone or radio, and a one-way link by sending a messenger, lighting a signal light, setting off an explosion, etc. We shall use the symbol $\gg$ to indicate such a connection; $A_i \gg A_j$ shall mean that that individual A_i can communicate with A_j (in that direction). The only requirement that we put on the symbol is

(i) It is false that $A_i \gg A_i$ for any i; that is an individual cannot (or need not) communicate with himself.

It is convenient to use directed graphs to represent communication networks. Two such graphs are drawn in Figure 1. Individuals are represented on the graph as (lettered) points and a communication relation between two individuals as a directed line segment (line segment with an arrow) connecting the two individuals.

We can also represent communication networks by means of square matrices C having only 0 and 1 entries, which we call *communication*

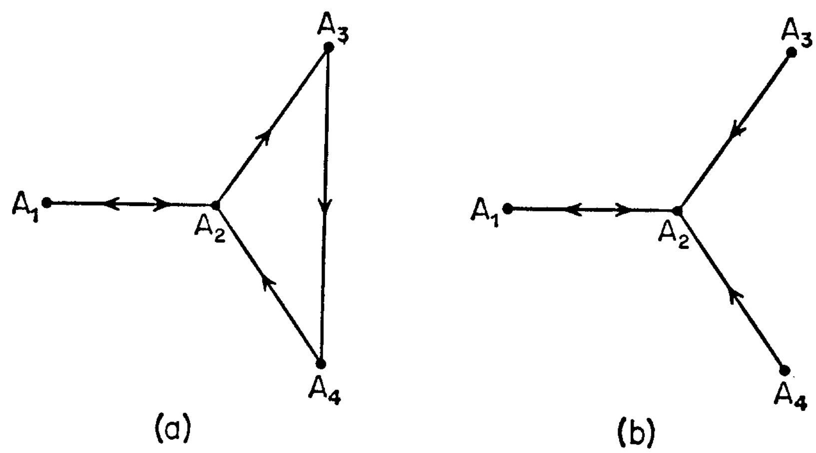


Figure 1

matrices. The entry in the ith row and jth column of C is equal to 1 if A_i can communicate with A_j (in that direction) and otherwise equal to 0. Thus the communication matrices corresponding to the com-

$$C = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{pmatrix} \qquad C = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}$$

(a) (b)

Figure 2

munication networks of Figure 1 are shown in Figure 2.

Notice that the diagonal entries of the matrices in Figure 2 are all equal to 0. This is true in general for a communication matrix, since the matrix restatement of condition (i) is

(i) For all i, $c_{ii} = 0$.

It is not hard to see that any matrix having only 0 and 1 entries, and with all zeros down the main diagonal, is the communication matrix of some network.

By a *dominance relation* we shall mean a special kind of communication relation in which, besides (i), the following condition holds.

(ii) For each pair i, j, with $i \neq j$, either $A_i \gg A_j$ or $A_j \gg A_i$, but not both; that is, in every pair of individuals, there is exactly one who is dominant.

It has been observed that in the pecking order of chickens a dominance relation holds. Also, in the play of one round of a round robin contest among athletic teams, if ties are not allowed (as in baseball), then a dominance relation holds.

The reader may have been surprised that we did not assume that if $A_i \gg A_j$ and $A_j \gg A_k$ then $A_i \gg A_k$. This is the so-called transitive law for relations. A moment's reflection shows that the transitive law need not hold for dominance relations. Thus if team A beats team B and team B beats team C (in football, say), then we cannot assume that team A will necessarily beat team C. In every football season there are instances in which "upsets" occur.

Dominance relations may also be depicted by means of directed graphs. Two such are shown in Figure 3. The graph in Figure 3a

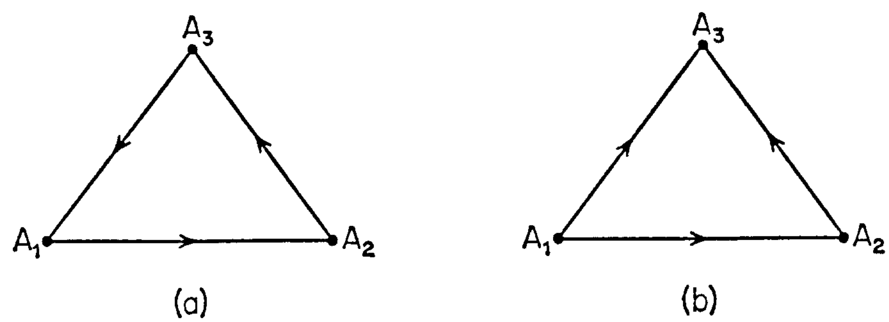


Figure 3

represents the situation: A_1 dominates A_2, A_2 dominates A_3, and A_3 dominates A_1. Similarly, the graph in Figure 3b represents the situation: A_1 dominates A_2 and A_3, and A_2 dominates A_3. These graphs represent the two essentially different dominance relations that are possible among three individuals (cf. Exercise 1).

Dominance relations may also be defined by means of matrices, called *dominance matrices,* defined as for communication matrices. In Figure 4 we have shown the two dominance matrices corresponding to the directed graphs of Figure 3.

Since a dominance matrix is derived from a dominance relation, we can investigate the effects of conditions (i) and (ii) above on the entries in the matrix. Condition (i) simply means that all entries on the main diagonal (the one which slants downward to the right) of the matrix must be zero. Condition (ii) means that, whenever an entry above the main diagonal of the matrix is 1, the corresponding entry of the matrix which is placed symmetrically to it through the main diagonal is 0, and

$$D = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}, \qquad D = \begin{pmatrix} 0 & 1 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}$$

(a) (b)

Figure 4

vice versa. To state these conditions more precisely, suppose that there are n individuals, and let D be a dominance matrix with entries d_{ij}. Then the conditions above are

(i) $d_{ii} = 0$ for $i = 1, 2, \ldots, n$.
(ii) If $i \neq j$, then $d_{ij} = 1$ if and only if $d_{ij} = 0$.

Every dominance relation is also a communication relation, hence we shall concentrate on the latter, and what we say about them will also be true for the former.

Since a communication matrix C is square, we can compute its powers, C^2, C^3, etc. Let $E = C^2$, and consider the entry in the ith row and jth column of E. It is

$$e_{ij} = c_{i1}c_{1j} + c_{i2}c_{2j} + \ldots + c_{in}c_{nj}.$$

Now a term of the form $c_{ik}c_{kj}$ can be nonzero only if both factors are nonzero, that is, only if both factors are equal to 1. But if $c_{ik} = 1$, then individual A_i communicates with A_k; and if $c_{kj} = 1$, then individual A_k communicates with A_j. In other words, $A_i \gg A_k \gg A_j$. We shall call a communication of this kind a *two-stage communication*. (To keep ideas straight, let us call $A_i \gg A_j$ a *one-stage* communication.) We can now see that the entry e_{ij} gives the total number of two-stage communication paths there are between A_i and A_j (in that direction). For example, let C be the matrix

$$C = \begin{pmatrix} 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

Then C^2 is the matrix

$$C^2 = \begin{pmatrix} 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

Thus we see that in this example A_1 has one two-stage communication path with A_3 and two two-stage communicat ons with A_4; similarly, A_2 has one two-stage communication with A_4. These can be written down explicitly as

$$A_1 \gg A_2 \gg A_3,$$
$$A_1 \gg A_2 \gg A_4,$$
$$A_1 \gg A_3 \gg A_4,$$
$$A_2 \gg A_3 \gg A_4.$$

The directed graph for this (dominance) situation is given in Figure 5. The reader should trace out on the graph of Figure 5 the two-stage communication paths given above.

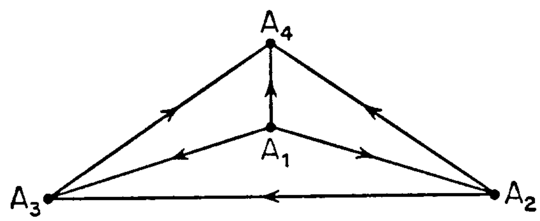


Figure 5

Theorem. Let a communication network of n individuals be such that, for every pair of individuals, at least one can communicate in one stage with the other. Then there is at least one person who can communicate with every other person in either one or two stages. Similarly, there is at least one person who can be communicated with in one or two stages by every other person.

Stated in matrix language, the above theorem is: Let C be the communication matrix for the network described above; then there is at least one row of $S = C + C^2$ which has all its elements nonzero, except possibly the entry on the main diagonal. Similarly, there is at least one column having this property.

Notice that every dominance relation satisfies the hypotheses of the theorem, but there are communication networks, not dominance relations, that also satisfy these hypotheses.

Proof. We shall prove only the first statement since the proof of the second is analogous.

First we shall prove the following statement: If A_1 cannot communicate in either one or two stages with A_i, where $i \neq 1$, then A_i can communicate in one stage with at least one more person than can A_1. We prove this in two steps. First by the hypothesis of the theorem, we see that:

(a) If it is false that $A_1 \gg A_i$, then $A_i \gg A_1$.

Second we can prove that:

(b) Suppose that for all k it is false that $A_1 \gg A_k \gg A_i$; it follows that, if $A_1 \gg A_k$, then also $A_i \gg A_k$.

For if $A_1 \gg A_k$, it is false that $A_k \gg A_i$; hence, by the hypothesis of the theorem, it is true that $A_i \gg A_k$.

Now (b) says that every one-stage communication possible for A_1 is also possible for A_i. From this and (a), it then follows that A_i can make at least one more (one-stage) communication than can A_1.

We now return to the proof of the theorem. Let $r_1, r_2, \ldots, r_n$ be the row sums of the matrix C. By renaming the individuals, if necessary, we can assume that the largest row sum is r_1, that is, $r_1 \geq r_k$ for $k = 1, 2, \ldots, n$. We shall show that A_1 can communicate with everyone else in one or two stages. (The proof is based on the indirect method.) Suppose, on the contrary, that there is an individual A_i, where $i > 1$, with whom A_1 cannot so communicate. By the statement proved above, A_i can communicate in one stage with at least one more person than A_1 can. But this implies that $r_i > r_1$, which contradicts the fact that we have named the individuals so that $r_1 \geq r_i$. This contradiction establishes the theorem.

An additional conclusion which can be made from the proof of the theorem is that the individual or individuals having the *largest* row sum in the matrix C can communicate with everyone else in one or two stages. Similarly, the individuals having the *largest* column sum can be communicated with by everyone in one or two stages.

The network shown in Figure 6 satisfies the hypothesis of the theorem, hence its conclusion. The communication matrix for this network is

$$\begin{pmatrix} 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 \end{pmatrix}.$$

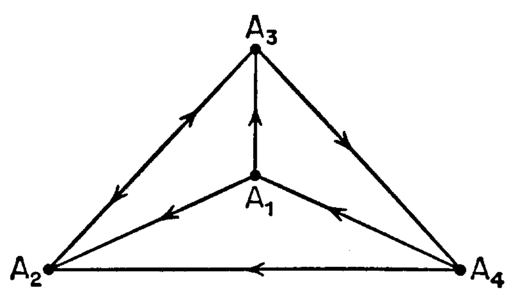


Figure 6

Here the maximum row sum of 2 occurs in rows one, three, and four, so that A_1, A_3, and A_4 can communicate with everyone else in one or two stages. (Find the necessary communication paths in Figure 6.) However, it requires three stages for A_2 to communicate with A_1. The maximum sum of 3 occurs in column two so that A_2 can be communicated with by everyone else in one or two stages (actually one stage is enough). It happens also that A_3 and A_4 can also be communicated with in one or two stages; however, as observed above, A_1 cannot be.

Neither of the networks in Figure 1 satisfies the hypothesis of the theorem. It happens that the network in Figure 1a does satisfy the conclusion of the theorem, while the network in Figure 1b does not. (See Exercise 7.)

As a final application of dominance matrices, we shall define the power of an individual. By the *power* of an individual in a dominance situation, we mean the total number of one-stage and two-stage dominances which he can exert. Since the total number of one-stage dominances exerted by A_i is the sum of the entries in row i of the matrix D, and the total number of two-stage dominances exerted by A_i is the sum of the entries in row i of the matrix D^2, we see that the power of A_i can be expressed as follows:

The power of A_i is the sum of the entries in row i of the matrix $S = D + D^2$.

In the example of Figure 7 it is easy to check that the powers of the various individuals are the following.

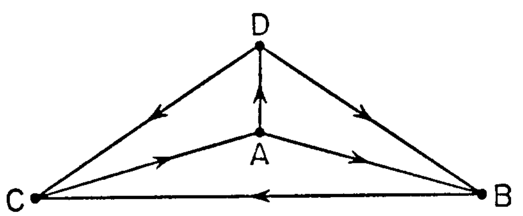


Figure 7

The power of A is 5.
The power of B is 2.
The power of C is 3.
The power of D is 4.

Example. (Athletic contest). The idea of the power of an individual can be used to judge athletic events. For example, the result of a single round of a round robin athletic event results in the following data.

Team A beats teams B and D.
Team B beats team C.
Team C beats team A.
Team D beats teams C and B.

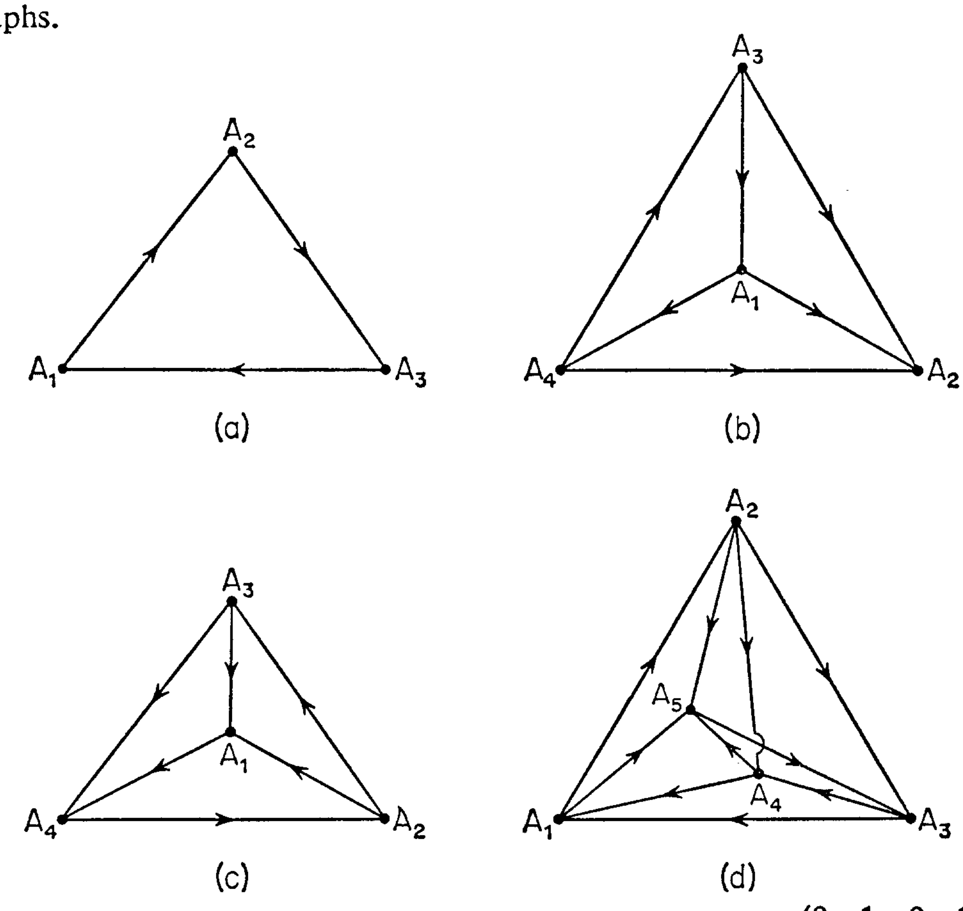
Then it is easy to check that this is precisely the dominance situation shown in Figure 7. By the analysis given above we can rate the teams in the following order according to their respective powers: A, D, C, and B.

It should be remarked that the above definition of the power of an individual is not the only one possible. In Exercise 13 below we suggest another definition of power which gives different results. Before using one or the other of these definitions, a sociologist should examine them carefully to see which (if either) fits his needs.

EXERCISES

1. Show that there are only two essentially different pecking orders possible among three chickens, namely, those given in Figure 3. [*Hint:* Use directed graphs.]

2. Find the dominance matrices D corresponding to the following directed graphs.

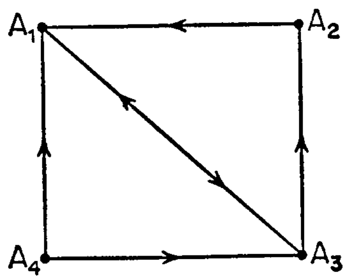


$$[\textit{Ans.}\ (b)\ \begin{pmatrix} 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \end{pmatrix}.]$$

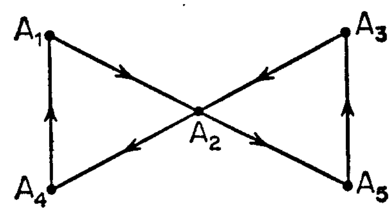
3. Compute the matrices D^2 and $S = D + D^2$ and determine the powers of each of the individuals in the examples of Exercise 2.

$$\left[\textit{Ans.}\ \text{(b)}\ D^2 = \begin{pmatrix} 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 \end{pmatrix};\ S = \begin{pmatrix} 0 & 2 & 1 & 1 \\ 0 & 0 & 0 & 0 \\ 1 & 2 & 0 & 1 \\ 1 & 2 & 1 & 0 \end{pmatrix};\ 4, 0, 4, 4.\right]$$

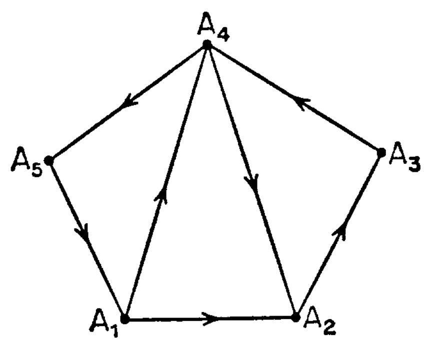
4. Find the communication matrices for the following communication networks.



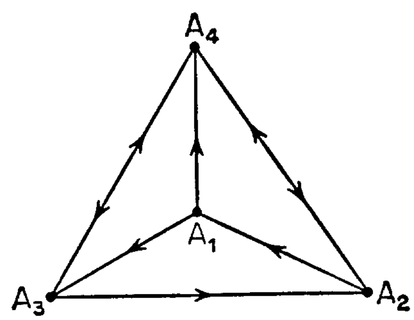
(a)



(b)



(c)



(d)

$$\left[\textit{Ans.}\ \text{(a)}\ \begin{pmatrix} 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \end{pmatrix}.\right]$$

5. Draw the directed graphs corresponding to the following communication matrices.

(a) $\begin{pmatrix} 0 & 1 & 1 \\ 0 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}$. (b) $\begin{pmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix}$.

(c) $\begin{pmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{pmatrix}$. (d) $\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}$.

6. Which of the communication networks whose matrices are given in Exercise 5 satisfy the hypothesis of the theorem of this section?
[*Ans.* (a) and (c).]

7. Show that the network in Figure 1a satisfies the conclusion of the theorem, while the network in Figure 1b does not.

8. By computing the matrix S in each case, find the persons who can communicate with everyone else in one or two stages and those who can be communicated with in one or two stages, for the communication matrices in Exercise 5. (In some cases such persons need not exist.)
[*Ans.* (a) Everyone; (b) everyone; (d) neither type of person exists.]

9. Find all the essentially different pecking orders that are possible among four chickens. [*Ans.* There are four essentially different ones.]

10. If C is any communication matrix, give the interpretation of the entries in the columns of the matrix $S = C + C^2$. Also give the interpretation for the column sums of S.

11. Find all communication networks among three individuals which satisfy the hypothesis of the theorem of this section. How many of these are essentially different? [*Ans.* There are seven.]

12. A round robin tennis match among four people has produced the following results.

Smith has beaten Brown and Jones.
Jones has beaten Brown.
Taylor has beaten Smith, Brown, and Jones.

By finding the powers of each player, rank them into first, second, third, and fourth place. Does this ranking agree with your intuition?
[*Ans.* Taylor has power = 6, Smith has power = 3, Jones has power = 1, and Brown has power = 0.]

13. Let the power$_1$ of an individual be the power as defined in the text above. Define a new power, called power$_2$, of an individual as follows: If D is the dominance matrix for a group of n individuals, then the power$_2$ of A_i is the sum of row i of the matrix

$$S' = D + \tfrac{1}{2}D^2.$$

Find the power$_2$ of each of the teams in the athletic team example in the text. Show that the power$_2$ of a team need not equal his power$_1$. Comment on the result.

14. Find the power$_2$ of the players in Exercise 12. Discuss its relation with the power$_1$ of each of the players.
[*Ans.* Taylor has power$_2$ = $\frac{9}{2}$, Smith has power$_2$ = $\frac{5}{2}$, Jones has power$_2$ = 1, Brown has power$_2$ = 0.]

15. If C is a communication matrix, give an interpretation for the entries of the matrix C^3. Do the same for the matrix C^4.

[*Ans.* The entry in row i and column j of C^3 gives the number of three-stage communications from i to j; the same entry of C^4 gives the number of four-stage communications from i to j.]

16. If C is a communication matrix, give an interpretation for the entries of the matrix $S = C + C^2 + C^3 + \ldots + C^m$.

17. Prove the second statement of the theorem of the present section.

18. Prove that the following statement is true: In a communication network involving three individuals, it is possible for a message starting from any person to get to any other person if and only if the following condition is satisfied: each individual can send a message to at least one person and can receive a message from at least one person.

19. Show that the matrix form of the condition in Exercise 18 is: Every row and column of the communication matrix must have at least one nonzero entry.

20. Is the statement in Exercise 18 true for a communication network involving two individuals? For four or more individuals? [*Ans.* Yes; no.]

2. *EQUIVALENCE CLASSES IN COMMUNICATION NETWORKS*

When considering communication networks, it becomes obvious that the various members of the network play different roles. Some members can only send messages, some can only receive them, and others can both send and receive. Subsets of members are also important. We shall consider subsets of members having the following two properties: (a) every member of the subset can both send and receive messages (not necessarily in one step) to and from every other member in the subset; and (b) the subset having property (a) is as large as possible. We shall show that it is possible to partition the set of all people in the network into subsets (called equivalence classes) having these two properties, and that between such equivalence classes there is at most a one-way communication link. We then apply our results to three different problems, (i) putting any nonnegative matrix into canonical form, (ii) the classification of states in a Markov chain, and (iii) the solution of an archeological problem.

As in the previous section, let $A_1, \ldots, A_n$, be the members of the communication network. We define a relation, R, between some pairs of these members as follows: let $A_i\mathrm{R}A_j$ mean "A_i can send a message to A_j (in that direction and not necessarily in one step) or else $i = j$." Then it is easy to show that the relation R has the following two properties:

(1) $A_i\mathrm{R}A_i$ for every i. (Reflexive axiom)
(2) $A_i\mathrm{R}A_j$ and $A_j\mathrm{R}A_k$ implies $A_i\mathrm{R}A_k$. (Transitive axiom)

To see this, note that property (1) follows from the definition of R, and (2) follows since if A_i can send a message to A_j and A_j can send a message to A_k then A_i can send a message to A_k by routing it through A_j.

If S is *any* set and R is *any* relation defined for members of S that satisfies axioms (1) and (2), then R is called a *weak ordering* on S.

We next define another relation on the states of the network. Let $A_i\mathrm{T}A_j$ hold if and only if $(A_i\mathrm{R}A_j) \wedge (A_j\mathrm{R}A_i)$, that is, $A_i\mathrm{T}A_j$ holds if and only if "A_i has a two-way communication with A_j or else $i = j$." It is easy to show that the relation T has the following three properties:

(3) $A_i\mathrm{T}A_i$. (Reflexive axiom)
(4) $A_i\mathrm{T}A_j$ if and only if $A_j\mathrm{T}A_i$. (Symmetric axiom)
(5) $A_i\mathrm{T}A_j$ and $A_j\mathrm{T}A_k$ implies $A_i\mathrm{T}A_k$. (Transitive axiom)

In Exercise 1 the reader is asked to establish these three axioms.

If S is any set and T is any relation defined for members of S that satisfies axioms (3), (4), and (5), then T is called an *equivalence relation* on S. The principal result about equivalence relations defined over a set S is that they partition S into equivalence classes.

DEFINITION. We say that A_i and A_j are *equivalent* if $A_i\mathrm{T}A_j$. For any A_i the *equivalence class* E_i that it determines is the truth set of the statement $A_i\mathrm{T}A_k$, i.e., it is the set of all A_k such that $A_i\mathrm{T}A_k$ is true.

Theorem 1. The equivalence classes of T partition S, the set of members of the communication network.

Proof. We must show that every member A_i of S belongs to one and only one equivalence class. Let S' be the equivalence class of A_i. Since $A_i\mathrm{T}A_i$ [from (3) above], we know that A_i belongs to S', which shows that A_i belongs to some equivalence class, and also that S' is not empty.

Now let A_i and A_j be any two members of S, and let S' and S'', respectively, be their equivalence classes. We shall show that either $S' \cap S'' = \mathcal{E}$ or else $S' = S''$. If $S' \cap S'' = \mathcal{E}$ then we are done.

Hence, suppose that there is an element X of S in $S' \cap S''$. Since X is in S', we have $A_i\text{T}X$; and since X is in S'', we have $A_j\text{T}X$. Using (4) we have $X\text{T}A_j$. But, by virtue of transitivity (5), $A_i\text{T}X$ and $X\text{T}A_j$ implies $A_i\text{T}A_j$, hence A_j is in S'. Let Y be any element is S'' so that $A_j\text{T}Y$. Using transitivity again, we have $A_i\text{T}A_j$ and $A_j\text{T}Y$ so that Y is in S'. We have thus shown that every element of S'' is in S', i.e., $S'' \subset S'$. In the same manner, one can show that $S' \subset S''$. Hence $S' = S''$.

Since we have shown that every member of S belongs to an equivalence class, and that every pair of equivalence classes are either identical or else disjoint, we have shown that they partition S, completing the proof of the theorem.

We now define a relation R on the equivalence classes of S. Namely, we let $S'\text{R}S''$ mean, "either $S' = S''$ or else some member of S' can send a message to some member of S''. We leave it to the reader in Exercise 6 to show that R is a weak ordering of the set of equivalence classes of S.

Theorem 2. Let S' and S'' be two equivalence classes; then, if $S'\text{R}S''$, it is false that $S''\text{R}S'$. In other words, at most one-way communication is possible between equivalence classes.

Proof. Suppose, on the contrary, that S' and S'' are two equivalence classes such that $S'\text{R}S''$ and $S''\text{R}S'$. Then there is an element X in S' that can communicate with some element Y in S''; and there is an element Z in S'' that can communicate with some element U in S'. Since Y and Z are in S'', two-way communication is possible between them; and since X and U are in S' they also have two-way communication. Hence Y can communicate with Z, Z can communicate with U and U can communicate with X. Therefore X and Y are in the same equivalence class, contradicting the assumption that they were in different (and hence disjoint) equivalence classes. This completes the proof.

For applications it is important to be able to find the equivalence classes for a given communication network. We develop an iterative method that constructs the following sets.

(6) T_k, the set of states A_k can *send* a message *to* (not necessarily in one step),

(7) F_k, the set of states that A_k can *receive* a message *from* (not necessarily in one step),

(8) E_k, the equivalence class of A_k.

It is easily seen (see Exercise 7) that $E_k = T_k \cap F_k$, so that we develop a method for interatively (that is, step by step) constructing the sets T_k and F_k. We illustrate the method with an example.

Example 1. We wish to get in contact with five alumni of a certain college, but do not know all their addresses. However, we have information of the form, "Jones knows where Brown is," "Smith knows where Doe is," etc. We summarize this information in the communication matrix of Figure 8. In that figure for $i \neq j$ we put 1 in the i,jth

	Brown	Jones	Smith	Adams	Doe
Brown	0	0	0	0	0
Jones	1	0	0	1	0
Smith	0	0	0	0	1
Adams	0	1	1	0	0
Doe	0	0	0	0	0

Figure 8

entry if the ith person knows where the jth one is. What is the smallest number of people that we must contact in order to send a message to all of them?

In order to solve this problem we first find the "send-to" lists for each person. We start by listing all the persons a person can contact in zero or one steps; these data come directly out of the communication matrix. These people form the "first stage approximation" to the "send-to" lists. Next we go down the list of persons and add to his "send-to" list all the people who can be contacted by people already on his first-stage approximate "send-to" list. The results are the "second-stage approximation to the send-to lists." We continue this process, step by step, until for the first time we go through the list and do not add any member to any person's "send-to" list. We then have the actual "send-to" sets for each person, since going through the process again would not change any list. The computations for the example in Figure 8 are shown in Figure 9.

The first-stage approximation to the "send-to" list is shown in the second column of Figure 9. On the first pass through the list we add 3 to Jones's list, which is indicated by bold-face in the third column. We also add 1 and 5 to Adams' list, also indicated by bold-faced numerals. On the next pass through the computation we add 5 to Jones's list and make no other changes. The next pass through the computation produces no further changes so that the final lists shown in the third column of Figure 9 is the complete "send-to" list for each person.

Person	Zero- or One-Stage Communication	Send-to List
1 Brown	1	1
2 Jones	1, 2, 4	1, 2, 4, 3, 5
3 Smith	3, 5	3, 5
4 Adams	2, 3, 4	2, 3, 4, 1, 5
5 Doe	5	5

Figure 9

We see that we have solved the problem posed above, for by contacting either Jones or Adams, we can relay a message to each of the five alumni members.

Let us go further and find the "receive-from" lists and the equivalence classes for each person in the network. The "receive-from" lists are easy, for we simply go down the "send-to" list and if we find member k on the ith person's "send-to" list, we put i on the kth person's "receive-from" list. And we compute the equivalence classes from the relationship $E_k = T_k \cap F_k$. These computations are shown in Figure 10.

Person	Send-to List	Receive-from List	Equivalence Class
1 Brown	1	1, 2, 4	{1}
2 Jones	1, 2, 4, 3, 5	2, 4	{2, 4}
3 Smith	3, 5	2, 3, 4	{3}
4 Adams	2, 3, 4, 1, 5	2, 4	{2, 4}
5 Doe	5	2, 3, 4, 5	{5}

Figure 10

It is interesting to draw the graph of the weak ordering relation R on the equivalence classes. To find the graph we simply check whether one-way communication is possible between each pair of equivalence classes. Then we connect two equivalence classes in the graph if such one-way communication is possible and if there is no intermediate class in the communication path. The graph of the weak ordering relation for the matrix of Figure 8 is shown in Figure 11. Note that equivalence class {2, 4} can communicate directly to {1} and {3} and to {5} through {3}. This graph shows very clearly the fact, noted above, that in order to contact all members of the group it is sufficient to contact either member of the equivalence class {2, 4}.

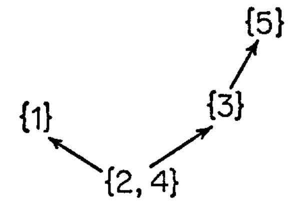


Figure 11

We can use the weak ordering of equivalence classes to put the matrix of Figure 8 in a canonical form, which is characterized by the following definition.

DEFINITION. Let C be any communication matrix, and let S', S'', . . . , be the equivalence classes of its states. Then, by a *canonical form* of C, we shall mean a reordering of the rows and columns of C so that the following two properties are satisfied.

(i) Members of a given equivalence class are listed next to each other.

(ii) No equivalence class S' is listed until all classes S'' "above" it in the graph of the equivalence classes have already been listed, i.e., S' is not listed until all classes such that $S'RS''$ have already been listed.

Example 1 (continued). We illustrate this definition in terms of the matrix A of Figure 8. Using the weak ordering diagram of Figure 11, we see that the following listing of the states (row indices) of A will satisfy the definition: 1, 5, 3, 2, 4. The resulting matrix is shown in Figure 12. In that figure dotted lines appear along the main diagonal, indicating the equivalence classes. Note that above the main diagonal blocks the only entries are zeros. Matrices having this property are called *block triangular.*

The same kind of canonical form is possible for *any* nonnegative matrix A, if we let $C(A)$ be the communication matrix derived from A by putting zeros on the main diagonal, and replacing positive off-diagonal entries by ones. We discuss this for Markov chain transition matrices. When the matrix under consideration is the transition matrix of a Markov chain, the classification of the states is extremely important in the study of the behavior of the chain, as the following definition indicates.

	Brown	Doe	Smith	Jones	Adams
1 Brown	0	0	0	0	0
5 Doe	0	0	0	0	0
3 Smith	0	1	0	0	0
2 Jones	1	0	0	0	1
4 Adams	0	0	1	1	0

Figure 12

DEFINITION. Let P be the transition matrix of a Markov chain, and let $C(P)$ be the matrix obtained from P by replacing each diagonal entry by 0 and replacing each positive off-diagonal entry by 1. Let $S', S'', \ldots$ be the equivalence classes of the states of $C(P)$; then

(i) The maximal equivalence classes, that is, those classes that cannot send to other classes, are called *ergodic sets*. Members of ergodic sets are called *ergodic states*. If an ergodic set contains a single state, that state is an *absorbing state*.

(ii) All equivalence classes that can send messages to other classes are called *transient sets*. Members of transient sets are called *transient states*.

Example 2. Consider the transition matrix

$$P = \begin{pmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{4} & \frac{1}{4} & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & \frac{1}{3} & 0 & \frac{2}{3} \end{pmatrix}.$$

Changing the diagonal entries to zeros and the positive off-diagonal entries to ones gives

$$C(P) = \begin{matrix} & \begin{matrix} 1 & 2 & 3 & 4 & 5 \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{matrix} & \begin{pmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{pmatrix} \end{matrix}.$$

In Exercise 8 the reader will be asked to show that the equivalence classes of $C(P)$ are $\{4\}$, $\{1, 3\}$, $\{5\}$, and $\{2\}$. Moreover, the graph of the weak ordering relation on these classes is as shown in Figure 13. As before, the graph is obtained by checking whether or not one-way communication is possible between each pair of equivalence classes. From this diagram and the above definition we see that $\{4\}$ and $\{1, 3\}$ are ergodic sets and that $\{4\}$ is an absorbing state; also $\{2\}$ and $\{5\}$ are transient sets. A canonical form of the matrix found by listing the states in the order 4, 1, 3, 5, 2 is

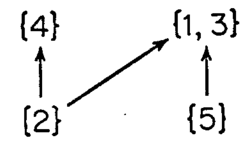


Figure 13

$$P = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{3} & \frac{2}{3} & 0 \\ \frac{1}{4} & 0 & \frac{1}{4} & 0 & \frac{1}{2} \end{pmatrix}.$$

Note again that it is block triangular, as indicated by the dotted lines. There are other orders in which to list the states, which lead to slightly different canonical forms for the matrix (see Exercise 9).

We conclude this section with an application of the above theory to an archeological problem.

Example 3. Recent archeological investigations in Asia Minor, between the Mediterranean and Black Seas, have disclosed the existence of an ancient Assyrian civilization dating back to at least the nineteenth century B.C. This civilization came to light when peasants working in fields turned up clay tablets having written inscriptions. Upon being translated, these tablets turned out to be letters written between merchants located at various cities and towns of the ancient civilization. The letters contained the name of the sender, the name of the receiver, and an order to buy, sell, or transport goods, to pay money, etc. But the *date* of the letter was not included. In addition, merchants in different villages sometimes had the same name, and the location of the merchant was not always made clear in each of the letters. More than 2500 such tablets have been discovered; their contents give rise to two different problems. The first problem is to try to order the merchants according to their chronological dates. A second problem is to try to determine when the same name refers to more than one person. By studying the communication network that can be set up from the data of the tablets, we shall illustrate with small examples methods of trying to get partial answers to these questions.

To illustrate an approach to the first problem, suppose that we set up a (hypothetical) communication matrix for a group of ten merchants, as indicated in the matrix of Figure 14. In that matrix an entry of 1 is

	1	2	3	4	5	6	7	8	9	10
1	0	0	0	0	0	1	0	0	0	0
2	0	0	1	0	1	0	0	0	0	0
3	0	1	0	0	1	1	0	0	0	0
4	0	0	0	0	0	0	0	1	0	1
5	0	1	1	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	1	0	0
8	0	0	1	0	0	0	0	0	1	0
9	0	0	0	0	0	0	1	0	0	0
10	1	0	0	0	0	0	0	0	0	0

Figure 14

made in the i,jth entry if merchant i sent a letter to merchant j. Carrying out the same analysis as in Example 1 the equivalence classes are found to be $\{6\}$, $\{1, 10\}$, $\{2, 3, 5\}$, $\{7, 8, 9\}$, and $\{4\}$. The graph of the weak ordering relation on these classes is shown in Figure 15. It was determined, as before, by seeing whether there is one-way communication between each pair of equivalence classes. It is clear that members of a given equivalence class are contemporaries. But it is not clear which of the equivalence classes is earlier, merely from the one-way communication between them. However, further analysis of the content of the messages might help to establish this. For instance, if one of the messages exchanged among merchants 7, 8, and 9 were related to one of the messages exchanged among merchants 2, 3, and 5, then it would be reasonable to assume that they are all contemporaries. We see that here is a case in which mathematics cannot furnish the complete answer

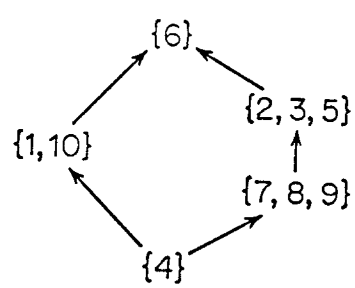


Figure 15

to the problem, but merely indicate directions in which to search for more information.

To illustrate the second problem mentioned above, we use some actual data (see p. 865 of the second reference listed at the end of the chapter) summarized in the communication matrix of Figure 16. The

	1	2	3	4	5	6	7	8	9	10
1 ASSUR-TAB	0	1	0	0	0	0	0	1	0	1
2 PUSHU-KIN	1	0	1	1	1	1	0	1	1	0
3 LAQIPUM	0	1	0	1	0	1	0	1	1	0
4 AMUR-ISHTAR	0	1	1	0	1	1	1	1	1	1
5 ASSUR-TAKLAKU	0	1	0	1	0	1	0	0	0	1
6 ASSUR-NA'DA	0	1	1	1	1	0	1	0	0	0
7 ASSUR-IMITTI	0	0	0	1	0	1	0	0	0	0
8 IM(I)D-ILUM	1	1	1	1	0	0	0	0	1	0
9 HINA	0	1	1	1	0	0	0	1	0	0
10 TARAM-KUBIM	1	0	0	1	1	0	0	0	0	0

Figure 16

matrix is symmetric, indicating that either there is a two-way (direct) communication between two individuals or else no (direct) communication at all. All the merchants belong to the same equivalence class, so that the previous analysis does not shed any light on their relative dates, except that they are contemporaries. But is it possible that some names really stand for two different individuals? No definite answer can be provided to this question, but some indications can be provided by finding the *cliques* in the communication network.

DEFINITION. A *clique* of a communication network is a subset C of individuals containing at least three members, with the following two properties.

(i) Every pair of members of the clique has two-way communication.

(ii) The subset C is as large as possible with every pair of members having property (i).

The problem of finding all cliques has been solved but is too lengthy to describe here. We content ourselves with listing all the maximal cliques for the data of Figure 16. They are

$$\{1, 2, 8\}, \quad \{2, 3, 4, 6\}, \quad \{2, 3, 4, 8, 9\}, \quad \{2, 4, 5, 6\}, \quad \{4, 6, 7\}.$$

From this list we can derive the frequency with which each merchant occurs in a clique, as shown in Figure 17. From that table it is evident that merchants 2 PUSHU-KIN and 4 AMUR-ISHTAR occur most frequently in cliques, and hence these names are most likely to be homonyms for two different people. Here again, mathematics does not completely

Merchant	1	2	3	4	5	6	7	8	9	10
Number of times in a clique	1	4	2	4	1	3	1	2	1	0

Figure 17

solve the problem, but merely indicates the direction in which to look for further evidence.

The above calculations, though oversimplified, are illustrative of the kinds of calculations that must be done n order to study the complete communication network revealed by the 2500 tablets so far found at the archeological site.

EXERCISES

1. Show that the relation T satisfies (3), (4), and (5).

2. Show that the relation "$\geq$" is a weak ordering relation on the set of integers. [*Hint:* Show that $x \geq y$, for x and y integers, satisfies (1) and (2).]

3. Show that the relation "$=$" is an equivalence relation on the set of all rational numbers (fractions). What are the equivalence classes it determines?

4. Let x and y be any two words and let $x\text{R}y$ mean "Word x occurs no later than word y in the dictionary." Show that R is a weak order on the set of words.

5. Let x and y be people and let $x\text{T}y$ mean "x is the same height as y." Show that T is an equivalence relation. What are the equivalence classes it determines? Show that the relation "at least as tall as" is a weak ordering relation on these equivalence classes.

6. Let R and T be the relations defined in the text; let S', S'', . . . be the equivalence classes determined by T; and let $S'\text{R}S''$ be as defined in the text. Show that R satisfies properties (1) and (2), that is, it is a weak ordering on the set of equivalence classes.

7. Let E_k, T_k, and F_k be as defined in the text. Show that $E_k = T_k \cap F_k$.

8. Find the equivalence classes of the communication matrix given in Example 2.

9. Show that there are ten different canonical forms for the transition matrix of Example 2.

10. Show that if A can communicate with B in a communication network having n persons, then it must be possible to do this in not more than $n - 1$ steps.

11. Suppose that there are six different individuals each of whom knows the location of certain others. This information is summarized in the following communication matrix.

$$\begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{array} \begin{array}{c} \begin{array}{cccccc} 1 & 2 & 3 & 4 & 5 & 6 \end{array} \\ \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 \end{pmatrix} \end{array}$$

(a) Find the equivalence classes of T.
(b) Draw the graph of the weak ordering relation on the equivalence classes.
(c) Suppose you know where 3 is and you want to find out where 1 is. What is the shortest communication path from 3 to 1?
[*Partial Ans.* It has ength 3.]
(d) What is the longest such communication path?
[*Partial Ans.* It has length 5.]

12. Classify each of the states of the Markov chain whose transition matrix is given below, and put the matrix into a canonical form. [*Hint:* Use some of the results of Exercise 11.]

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{2} & 0 & 0 & 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & \frac{3}{4} & 0 & 0 & 0 & \frac{1}{4} \\ \frac{2}{5} & 0 & 0 & 0 & \frac{3}{5} & 0 \\ 0 & 0 & 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{pmatrix}$$

[*Ans.* One canonical form is

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ \frac{2}{5} & \frac{3}{5} & 0 & 0 & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{3}{4} & 0 & \frac{1}{4} & 0 \\ 0 & \frac{1}{3} & 0 & \frac{1}{3} & \frac{1}{3} & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}.$$

State 1 is absorbing; all other states are transient.]

13. If a matrix M can be put into the form

$$M = \begin{pmatrix} A & 0 \\ B & C \end{pmatrix},$$

where 0 is the zero matrix then M is said to be *reducible* or *decomposable*. If A and C are square and nonsingular show that

$$M^{-1} = \begin{pmatrix} A^{-1} & 0 \\ -C^{-1}BA^{-1} & C^{-1} \end{pmatrix}.$$

14. Use the results of Exercise 13 to show how a canonical form of a non-negative matrix can be used to simplify the work of finding its inverse.

15. (a) Show that the Markov chain in Exercise 12 is an absorbing Markov chain.

(b) Find the matrix Q in canonical form. Show that the matrix $I - Q$ is block triangular.

(c) Use the results of Exercises 13 and 14 to find $N = (I - Q)^{-1}$. [*Ans.* With the canonical form of the answer to Exercise 12, the inverse is

$$N = (I - Q)^{-1} = \begin{pmatrix} \frac{5}{2} & 0 & 0 & 0 & 0 \\ \frac{5}{4} & 1 & 0 & 0 & 0 \\ \frac{10}{7} & \frac{6}{7} & \frac{8}{7} & \frac{3}{7} & 0 \\ \frac{55}{28} & \frac{3}{7} & \frac{4}{7} & \frac{12}{7} & 0 \\ \frac{55}{28} & \frac{3}{7} & \frac{4}{7} & \frac{12}{7} & 1 \end{pmatrix}.]$$

16. Draw the graph of a three-person clique. Also that of a four-person clique. Describe the graph of a clique containing n persons ($n \geq 3$).

17. Verify that the cliques given in Example 3 satisfy the two properties given in the definition of a clique.

18. Let C_1 and C_2 be any two *distinct* cliques of the same communication network.

(a) Show by examples that $C_1 \cap C_2$ may or may not be empty.

(b) Prove that the sets $C_1 - C_2$ and $C_2 - C_1$ are *never* empty.

3. *STOCHASTIC PROCESSES IN GENETICS*

The simplest type of inheritance of traits in animals occurs when a trait is governed by a pair of genes, each of which may be of two types, say G and g. An individual may have a GG combination or Gg (which is genetically the same as gG) or gg. Very often the GG and Gg types are indistinguishable in appearance, and then we say that the G gene *dominates* the g gene. An individual is called *dominant* if he has GG genes, *recessive* if he has gg, and *hybrid* with a Gg mixture.

In the mating of two animals, the offspring inherits one gene of the pair from each parent, and the basic assumption of genetics is that these genes are selected at random, independently of each other. This assumption determines the probability of every type of offspring. Thus the offspring of two dominant parents must be dominant, of two recessive parents must be recessive, and of one dominant and one recessive parent must be hybrid. In the mating of a dominant and a hybrid animal, the offspring must get a G gene from the former and has probability $\frac{1}{2}$ for getting G or g from the latter, hence the probabilities are even for getting a dominant or a hybrid offspring. Again in the mating of a recessive and a hybrid, there is an even chance of getting either a recessive or a hybrid. In the mating of two hybrids, the offspring has probability $\frac{1}{2}$ for getting a G or a g from each parent. Hence the probabilities are $\frac{1}{4}$ for GG, $\frac{1}{2}$ for Gg, and $\frac{1}{4}$ for gg.

Example 1. Let us consider a process of continued crossings. We start with an individual of unknown genetic character, and cross it with a hybrid. The offspring is again crossed with a hybrid, etc. The resulting process is a Markov chain. The states are "dominant," "hybrid," and "recessive." The transition probabilities are

$$P = \begin{array}{c} \\ D \\ H \\ R \end{array} \begin{array}{c} \begin{array}{ccc} D & H & R \end{array} \\ \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ 0 & \frac{1}{2} & \frac{1}{2} \end{pmatrix} \end{array} \tag{1}$$

as can be seen from the previous paragraph. The matrix P^2 has all entries positive (see Exercise 1), hence we know from Chapter V, Section 7, that there is a unique fixed point probability vector, i.e., a vector p such that $pP = p$. By solving three equations, we find the fixed vector to be $p = (\frac{1}{4}, \frac{1}{2}, \frac{1}{4})$. Hence, no matter what type the original animal was, after repeated crossing we have probability nearly $\frac{1}{4}$ of having a dominant, $\frac{1}{2}$ of having a hybrid, and $\frac{1}{4}$ of having a recessive offspring.

In Example 1 we may ask a more difficult question. Suppose that we have a regular matrix P (as in Example 1), with states $a_1, \ldots, a_n$. The process keeps going through all the states. If we are in a_i, how long, on the average, will it take for the process to return to a_i? We can even ask the more general question of how long, on the average, it takes to go from a_i to a_j.

The average here is taken in the sense of an expected value. There is a probability p_1 that we reach a_j for the first time in one step, p_2 that we reach it first in two steps, etc. The expected value is $p_1 \cdot 1 + p_2 \cdot 2 + \ldots$. (See Chapter IV, Section 12.) This, in general, requires a difficult computation. However, there is a much simpler way of finding the expected values. Let the expected number of steps required to go from state a_i to a_j be m_{ij}. How can we go from a_i to a_j? We go from a_i to a_k with probability p_{ik} in one step. If $k = j$, we are there. If $k \neq j$, it takes an average of m_{kj} steps more to get to a_j. Hence m_{ij} is equal to 1 plus the sum of $p_{ik}m_{kj}$ for all $k \neq j$. To state this as a matrix equation we define the matrix $\overline{M}$ to be the matrix M but with all the diagonal entries m_{ii} being replaced by 0; also let C be the square matrix having all entries equal to 1. Then the equations for m_{ij} can be written in matrix form as

$$M = P\overline{M} + C. \tag{2}$$

To see that this is so let us concentrate on the i,jth entry of equation (2). On the left-hand side it is m_{ij}. On the right-hand side it is the i,jth entry of $P\overline{M}$ which is the sum of all products $p_{ik}m_{kj}$ for $k \neq j$ (since the main diagonal of $\overline{M}$ is zero) plus the i,jth entry in C, which is 1. This is the same as before. Let us now multiply (2) by p, the fixed vector of P. Recalling that p is a probability vector we obtain

$$pM = p\overline{M} + (1, \ldots, 1) \tag{3}$$

or

$$p(M - \overline{M}) = (1, \ldots, 1). \tag{4}$$

But all components of $M - \overline{M}$ except the diagonal ones are 0. Hence our equation simply states that $p_i m_{ii} = 1$ for each i. This tells us that $m_{ii} = 1/p_i$. *The average time it takes to return from a_i to a_i is the reciprocal of limiting probability of being in a_i.* In Example 1 this means that if we have a dominant offspring we will have another dominant in an average of four steps, after a hybrid we have another hybrid in an average of two steps, and a recessive follows a recessive on the average in four steps.

Example 2. A more interesting, and also more complex, process is obtained by crossing a given population with itself, and then crossing the offspring with offspring, etc. Let us suppose that our population has a fraction d of dominants, h hybrids, and r recessives. Then $d + h + r = 1$. If the population is very large and they are mated at random, then (by the law of large numbers) we can expect d^2 to be the fraction of matings in which both parents are dominant, $2dh$ the fraction of mating a dominant with a hybrid, etc. The tree of logical possibilities with branch probabilities marked on it is shown in Figure 18. We use it to compute the fraction of each type. To do this we

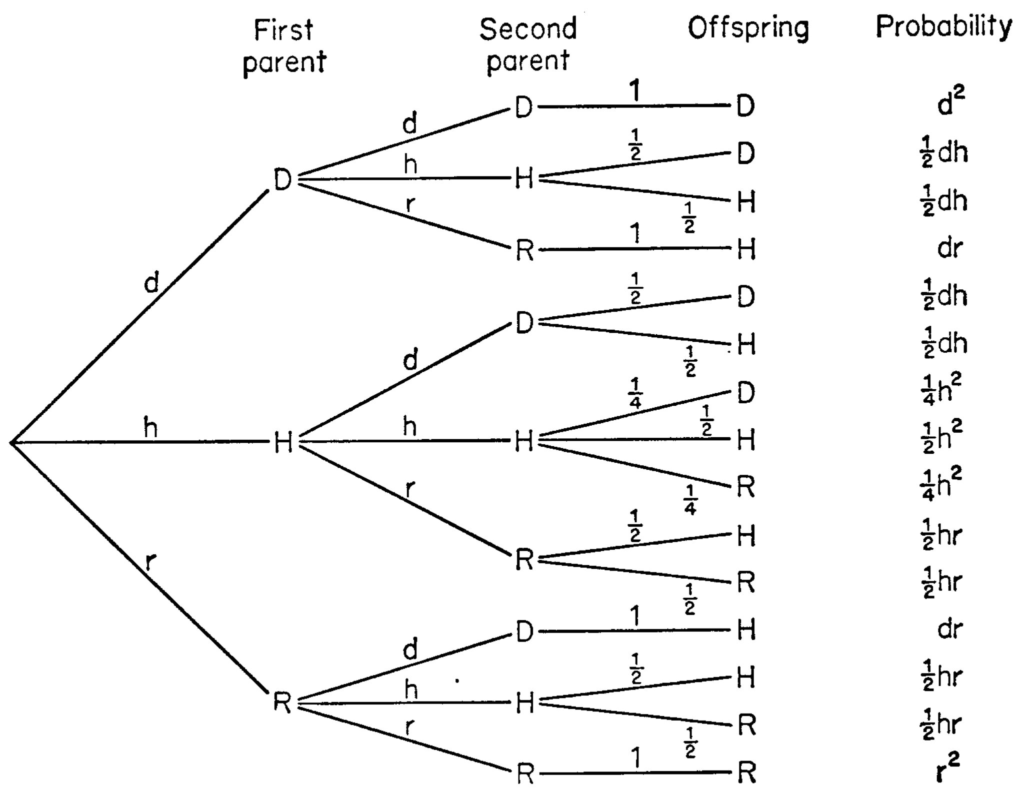


Figure 18

simply add together the path weights of the paths ending in D, in H, and in R. The results are:

$$D: \quad d^2 + 2\cdot\tfrac{1}{2}dh + \tfrac{1}{4}h^2 = d^2 + dh + \tfrac{1}{4}h^2$$

$$H: \quad 2\cdot\tfrac{1}{2}dh + 2dr + \tfrac{1}{2}h^2 + 2\cdot\tfrac{1}{2}hr = dh + rh + 2dr + \tfrac{1}{2}h^2$$

$$R: \quad \tfrac{1}{4}h^2 + 2\cdot\tfrac{1}{2}hr + r^2 = r^2 + hr + \tfrac{1}{4}h^2.$$

If we represent the fractions in a given generation by a row vector, the process may be thought of as a transformation T which changes a row vector into another row vector.

$$(5) \quad (d, h, r)\cdot T = (d^2 + dh + \tfrac{1}{4}h^2, dh + rh + 2dr + \tfrac{1}{2}h^2, r^2 + rh + \tfrac{1}{4}h^2).$$

The trouble is that (see Exercise 2) the transformation T is not linear. Nevertheless, we know that after n crossings the distribution will be

$(d, h, r)T^n$, so that, if we can get a simple formula for T^n, we can describe the results simply. And here luck is with us.

Let us compute T^2, i.e., find what happens if we apply twice the transformation specified above. The first generation of offspring is distributed according to the formula (5). We now take the first component on the right side as d, the second as h, and the third as r, and compute $d^2 + dh + \frac{1}{4}h^2$, etc. Here we find to our surprise that $T^2 = T$. Hence $T^n = T$.

This means that $(d, h, r)T = (d, h, r)T^n$, which in turn means that the distribution after many generations is the same as in the first generation of offspring. Hence we say that the process reaches an *equilibrium* in one step. It must, however, be remembered that our fractions are only approximate, and are a good approximation only for very large populations.

For the geneticist, this result is very interesting. It shows that, in a population in which no mutations occur and selection does not take place, "evolution" is all over in a single generation.

To the mathematician the process is interesting since it is an example of a quadratic transformation, a transformation more complex than the linear ones we have heretofore studied.

The next two examples give applications of absorbing Markov chains to genetics.

Example 3. If we keep crossing the offspring with a dominant animal, the result is quite different. The transition matrix is easily found to be

$$P' = \begin{array}{c} \\ D \\ H \\ R \end{array} \begin{array}{c} \begin{array}{ccc} D & H & R \end{array} \\ \begin{pmatrix} 1 & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 1 & 0 \end{pmatrix} \end{array}. \tag{6}$$

This is an absorbing Markov chain with one absorbing state, D. Using the results of Chapter V, Section 8, we have

$$Q' = \begin{pmatrix} \frac{1}{2} & 0 \\ 1 & 0 \end{pmatrix}, \qquad R' = \begin{pmatrix} \frac{1}{2} \\ 0 \end{pmatrix}$$

so that

$$I - Q' = \begin{pmatrix} \frac{1}{2} & 0 \\ -1 & 1 \end{pmatrix}$$

and

$$N' = \begin{pmatrix} 2 & 0 \\ 2 & 1 \end{pmatrix}.$$

The absorption probabilities are

$$B = N'R' = \begin{pmatrix} 2 & 0 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} \frac{1}{2} \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

as was to be expected, since there is only one absorbing state. This means that if we keep crossing the population with dominants, then after sufficiently many crossings we can expect only dominants. The mean number of steps to absorption are found by

$$t = N'c = \begin{pmatrix} 2 & 0 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 2 \\ 3 \end{pmatrix}$$

Hence we expect the process to be absorbed in two steps starting from state H, and three steps starting from state R.

Example 4. Let us construct a more complicated example of an absorbing Markov chain. We start with two animals of opposite sex, cross them, select two of their offspring of opposite sex and cross those, etc. To simplify the example we will assume that the trait under consideration is independent of sex.

Here a state is determined by a pair of animals. Hence the states of our process will be: $a_1 = (D, D)$, $a_2 = (D, H)$, $a_3 = (D, R)$, $a_4 = (H, H)$, $a_5 = (H, R)$, and $a_6 = (R, R)$. Clearly, states a_1 and a_6 are absorbing, since if we cross two dominants or two recessives we must get one of the same type. The rest of the transition probabilities are easy to find. We illustrate their calculation in terms of state a_2. When the process is in this state, one parent has GG genes, the other Gg. Hence the probability of a dominant offspring or a hybrid offspring is $\frac{1}{2}$ for each. Then the probability of transition to a_1 (selection of two dominants) is $\frac{1}{4}$, the transition to a_2 is $\frac{1}{2}$, and to a_4 is $\frac{1}{4}$. The complete transition matrix is (listing the absorbing states first)

$$P'' = \begin{array}{c} \\ a_1 \\ a_6 \\ a_2 \\ a_3 \\ a_4 \\ a_5 \end{array} \begin{array}{c} \begin{array}{cccccc} a_1 & a_6 & a_2 & a_3 & a_4 & a_5 \end{array} \\ \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ \frac{1}{4} & 0 & \frac{1}{2} & 0 & \frac{1}{4} & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ \frac{1}{16} & \frac{1}{16} & \frac{1}{4} & \frac{1}{8} & \frac{1}{4} & \frac{1}{4} \\ 0 & \frac{1}{4} & 0 & 0 & \frac{1}{4} & \frac{1}{2} \end{pmatrix} \end{array}.$$

Calculating the fundamental quantities for an absorbing chain, we obtain

$$Q'' = \begin{array}{c} \\ a_2 \\ a_3 \\ a_4 \\ a_5 \end{array}\!\!\begin{array}{c} \begin{array}{cccc} a_2 & a_3 & a_4 & a_5 \end{array} \\ \begin{pmatrix} \frac{1}{2} & 0 & \frac{1}{4} & 0 \\ 0 & 0 & 1 & 0 \\ \frac{1}{4} & \frac{1}{8} & \frac{1}{4} & \frac{1}{4} \\ 0 & 0 & \frac{1}{4} & \frac{1}{2} \end{pmatrix} \end{array}, \qquad R'' = \begin{pmatrix} \frac{1}{4} & 0 \\ 0 & 0 \\ \frac{1}{16} & \frac{1}{16} \\ 0 & \frac{1}{4} \end{pmatrix}$$

and

$$I - Q'' = \begin{pmatrix} \frac{1}{2} & 0 & -\frac{1}{4} & 0 \\ 0 & 1 & -1 & 0 \\ -\frac{1}{4} & -\frac{1}{8} & \frac{3}{4} & -\frac{1}{4} \\ 0 & 0 & -\frac{1}{4} & \frac{1}{2} \end{pmatrix},$$

and

$$N'' = (I - Q'')^{-1} = \begin{pmatrix} \frac{8}{3} & \frac{1}{6} & \frac{4}{3} & \frac{2}{3} \\ \frac{4}{3} & \frac{4}{3} & \frac{8}{3} & \frac{4}{3} \\ \frac{4}{3} & \frac{1}{3} & \frac{8}{3} & \frac{4}{3} \\ \frac{2}{3} & \frac{1}{6} & \frac{4}{3} & \frac{8}{3} \end{pmatrix}.$$

The absorption probabilities are found to be

$$B'' = N''R'' = \begin{pmatrix} \frac{3}{4} & \frac{1}{4} \\ \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \\ \frac{1}{4} & \frac{3}{4} \end{pmatrix}.$$

The genetic interpretation of absorption is that after a large number of inbreedings either the G or the g gene must disappear. It is also interesting to note that the probability of ending up entirely with G genes, if we start from a given state, is equal to the proportion of G genes in this state.

The mean number of steps to absorption are

$$t = N'' \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 4\frac{5}{6} \\ 6\frac{2}{3} \\ 5\frac{2}{3} \\ 4\frac{5}{6} \end{pmatrix}.$$

Hence we see that, if we start in a state other than (D, D) or (R, R), we can expect to reach one of these states in about five or six steps. The exact expected times are given by the entries of t. The matrix N'' provides more detailed information, namely how many times we can expect to have offspring of the types (D, H), (D, R), (H, H), and (H, R), starting from a given nonabsorbing state. And the matrix B'' gives the probabilities of ending up in a_1 or a_6. These quantities jointly give us an excellent description of what we can expect of our process.

EXERCISES

1. From (1) compute P^2, P^3, P^4, and P^5. Verify that $P^2 > 0$ and that the powers approach the expected form (see Chapter V, Section 7).

2. Prove that T is not a linear transformation. [*Hint:* Check the conditions on linearity given in Chapter V, Section 9, and show by means of an example that T does not have one of these properties.]

3. Compute T^2 by taking the first component of (5) as d, the second as h, the third as r, and substituting into the formula (5). Making use of the fact that $d + h + r = 1$, show that $T^2 = T$.

4. A fixed point of T is a vector such that $(d, h, r)T = (d, h, r)$. Write the conditions that such a vector must satisfy, and give three examples of such fixed vectors. What is the genetic meaning of such a distribution?

[*Ans.* For example, $(\frac{1}{9}, \frac{4}{9}, \frac{4}{9})$.]

5. In the matrix P the second row is equal to the fixed point vector. What significance does this have?

6. For Example 1 write the matrix M with unknown entries m_{ij}. Write M by replacing m_{11}, m_{22}, and m_{33} by zeros. Then solve the nine simultaneous equations given by (3), to find the m_{ij}. Check that $m_{ii} = 1/p_i$.

[*Ans.* $m_{11} = 4$, $m_{12} = 2$, $m_{13} = 8$.]

7. From the definition of a stochastic matrix (Chapter V, Section 7), prove that $PC = C$.

8. Prove that, if P is a regular $n \times n$ stochastic matrix having column sums equal to 1, then it takes an average of n steps to return from any state to itself. (Cf. Chapter V, Section 7, Exercise 8.)

9. It is raining in the Land of Oz. In how many days can the Wizard of Oz expect to go on a picnic? (Cf. Chapter V, Section 7, Exercise 13.) [*Ans.* 4.]

Exercises 10–15 develop a simpler method of treating the nonlinear transformation T, in the text above.

10. Let p be the ratio of G genes in the population, and $q = 1 - p$ the ratio of g genes. Express p and q in terms of d, h, and r.

[*Ans.* $p = d + \frac{1}{2}h$, $q = r + \frac{1}{2}h$.]

11. Suppose that we take all the genes in the population, mix them thoroughly, and select a pair at random for each offspring. Show, using the result of Exercise 10, that the resulting distribution of dominant, hybrid, and recessive individuals is precisely that given in (5).

[*Ans.* $(d, h, r) \cdot T = (p^2, 2pq, q^2)$.]

12. If we write $(d, h, r) \cdot T = (d', h', r')$, show, using the result of Exercise 11, that $h'^2 = 4d'r'$.

13. Show that for equilibrium it is necessary that $h^2 = 4dr$.

14. Show that if $h^2 = 4dr$, then $p^2 = d$, $q^2 = r$, and $2pq = h$. Hence show that this condition is also sufficient for equilibrium.

15. Use the results of Exercises 12–14 to show that the population reaches equilibrium in one generation.

16. Prove that in an absorbing Markov chain

(a) The probability of reaching a given absorbing state is independent of the starting state if and only if there is only one absorbing state.

(b) The expected time for reaching an absorbing state is independent of the starting state if and only if every state is absorbing.

17. Suppose that hybrids have a high mortality rate; say that half of the hybrids die before maturity, while only a negligible number of dominants and recessives die before maturity.

(a) In Example 4 above, modify the matrix P'' to apply to this situation.

(b) What are the absorbing states?

(c) Verify that it is an absorbing chain.

(d) Find the vectors d representing the probabilities of absorption in the various absorbing states.

$$[\textit{Ans. } \text{For } a_1, d = \begin{pmatrix} 1 \\ 0 \\ \frac{9}{10} \\ \frac{1}{2} \\ \frac{1}{2} \\ \frac{1}{10} \end{pmatrix}.]$$

(e) Find N, and interpret.

(f) Find t, and interpret.

$$[\textit{Ans. } t = \begin{pmatrix} \frac{65}{26} \\ \frac{117}{26} \\ \frac{91}{26} \\ \frac{65}{26} \end{pmatrix}.]$$

The remaining problems concern the inheritance of color-blindness, which is a sex-linked characteristic. There is a pair of genes, C and S, of which the former tends to produce color-blindness, the latter normal vision. The S gene is dominant. But a man has only one gene, and if this is C, he is color-blind. A man inherits one of his mother's two genes, while a woman inherits one gene from each parent. Thus a man may be of type C or S, while a woman may be of type CC or CS or SS. We will study a process of inbreeding similar to that of Example 4.

18. List the states of the chain. [*Hint:* There are six.]

19. Compute the transition probabilities.

20. Show that the chain is absorbing, and interpret the absorbing states. [*Ans.* In one, the S gene disappears; in the other, the C gene is lost.]

21. Prove that the probability of absorption in the state having only C genes, if we start in a given state, is equal to the proportion of C genes in that state.

22. Find N, and interpret.

23. Find t, and interpret.

[*Ans.* $\begin{pmatrix} 5 \\ 6 \\ 6 \\ 5 \end{pmatrix}$; if we start with both C and S genes, we can expect one of these to disappear in five or six crossings.]

4. THE ESTES LEARNING MODEL

In this section we shall discuss a mathematical model for learning proposed by W. K. Estes. We shall not give the most general theory, but only some special cases.

The theory was developed to explain certain kinds of learning which can be illustrated by experiments of the following kind. Suppose for example that a rat is put in a T maze and goes either right or left. The experimenter places food on one side, and if the rat goes to the correct side he is rewarded. This experiment is then repeated many times, using some particular feeding schedule. The interest here lies in trying to predict the behavior of the rat under the different feeding schedules. For instance, if the food is always placed on the right side, will the rat eventually learn this and always go right?

A similar experiment, performed with a human subject, is the following. A subject is given a sequence of heads and tails and each time is asked to guess what the next choice will be. He is to try to get as many right as possible. Again there are various ways that the experimenter can produce his sequences of H's and T's, and the interest lies in how the subject will react to different choices.

In the Estes model it is assumed that there are a finite number of elements, called "stimulus elements." At any given time each of these elements is connected either to a response A_1 or to a response A_2. These connections are allowed to change from experiment to experiment.

In a single experiment there is a certain probability θ $(0 < \theta < 1)$ that any particular stimulus element will be sampled by the subject. To say that an element is sampled is the same as to say that it has an effect upon the subject on that experiment. It is assumed that elements sampled and connected to A_1 influence the subject in the direction of producing an A_1 response, and those sampled and connected to A_2 tend to produce an A_2 response.

The samplings of the various elements are assumed to be an independent trials process (see Chapter IV, Section 8). Thus, for example, if there are three stimulus elements a, b, and c, the probability that a is sampled, b is not sampled, and c is sampled would be $\theta(1 - \theta)\theta$.

We also assume that the experimenter takes one of two possible "reinforcing" actions, E_1 or E_2. This action may be taken before or after the subject's choice, but we assume that the subject learns of the choice of the experimenter only after he has made his own choice. The subject would like to make A_1, if the experimenter makes E_1, and A_2 if the experimenter chooses E_2. We shall say that the subject "guesses correctly" if he matches the choice of the experimenter, i.e., does A_1 when the experimenter does E_1, or A_2 when the experimenter does E_2. In some experiments (e.g., the rat experiment above), he is rewarded if he does guess correctly.

The following two basic assumptions are made.

Assumption A. The probability that the subject makes response A_1 is equal to the proportion of elements *in the set sampled* that are connected to A_1. If no elements are sampled, the responses are assumed to be the same as if all elements are sampled.

Assumption B. If, in a given experiment, the experimenter chooses E_1, then all the elements that were sampled on this experiment, and that were connected to A_2 have their connections changed to A_1. If the experimenter chooses E_2, then all the elements sampled and connected to A_1 have their connections changed to A_2.

Note that in a single experiment only the set of elements that are actually sampled play a role, and these are the only elements whose connections can be changed by this experiment. In general, however, a different set will be sampled on each experiment, so that all the elements will at some time have an effect.

By assumptions A and B it is clear that the future choices of the subject are going to depend upon the choice of the experimenter. Therefore we must describe the method that the experimenter uses to determine his E's. Typical schemes that have been used in actual experiments are the following.

(i) Choose E_1 with probability p, independent of the choice of the subject.

(ii) Make the same choice as the subject made (i.e., choose E_1 if he chose A_1, E_2 if he chose A_2).

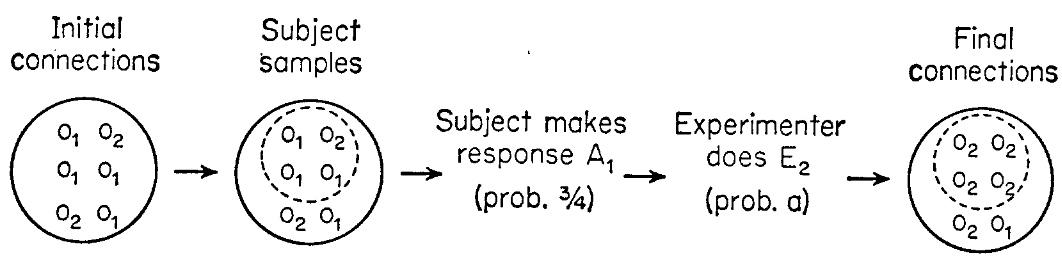
(iii) Choose E_1 if the response of the subject on the previous experiment was A_1. Choose E_2 and E_1 with equal probabilities if his response was A_2.

We can describe a general class of schemes of the above kind as follows: We assume that the experimenter chooses E_2 with probability a, if the subject made response A_1 on the previous experiment, and chooses E_1 with probability b, if the subject made response A_2 on the last experiment. We can represent the choices of the experimenter for each choice of the subject by the matrix

$$\begin{array}{c} \\ A_1 \\ A_2 \end{array} \begin{array}{c} \begin{array}{cc} E_1 & \quad E_2 \end{array} \\ \begin{pmatrix} 1 - a & a \\ b & 1 - b \end{pmatrix} \end{array}.$$

Thus in the above examples, (i) is the case $1 - a = b = p$, (ii) is the case $a = 0$, $b = 0$, and (iii) is the case $a = 0$, $b = \frac{1}{2}$.

In Figure 19 we illustrate a typical sequence of actions that might occur in a single trial for the case of six stimulus elements. An O_1



A Typical Sequence on One Trial

Figure 19

indicates a stimulus element which is connected to A_1 and an $\bigcirc_2$ indicates a stimulus element which is connected to A_2.

We note in this case that the subject sampled four of the six stimulus elements. The probability that this particular set of four elements is sampled is $\theta^4(1 - \theta)^2$. Since three of the four elements in the set sampled are connected to A_1, the subject makes response A_1 with probability $\frac{3}{4}$. We assume the subject made response A_1. Then the experimenter chooses E_2 with probability a. We have assumed that the experimenter did choose E_2. All four of the stimulus elements in the set sampled then become connected to response A_2. The final connections for this trial then become the initial connections for the next trial.

We shall now develop a method for studying the response process. We do this by introducing a Markov chain. The states of the chain will be the number of elements connected to response A_1. If there are six stimulus elements, then there are seven possible states: 0, 1, 2, 3, 4, 5, 6. We compute the transition probabilities from the above assumptions.

We shall consider throughout the rest of this section and the next the case of two stimulus elements. The analysis for a larger number of elements is similar but more complicated. Many of the results do not depend upon the number of stimulus elements assumed.

Our states are numbered 2, 1, and 0, indicating the number of stimulus elements connected to an A_1 response. We will illustrate the computation of transition probabilities. For example, let us compute $p_{0,1}$. Since the chain is in state 0, both stimulus elements are connected to response A_2. To change to state 1, exactly one stimulus element must be sampled. The probability for this is $2\theta(1 - \theta)$. If this stimulus element is to change to A_1, the experimenter must do E_1. The probability of this is b. Hence $p_{0,1} = 2\theta(1 - \theta)b$.

A more complicated computation is needed for $p_{1,0}$. In state 1, one stimulus element is connected to A_1 and one to A_2. The former must be sampled, the latter may also be sampled. The response of the experimenter must be E_2, to effect a change to A_2. There are three cases. (1) Only one element is sampled, with response A_1. (2) Both are sampled, with response A_1. (3) Both are sampled, with response A_2. These yield the three terms

$$p_{1,0} = \theta(1 - \theta)a + \tfrac{1}{2}\theta^2 a + \tfrac{1}{2}\theta^2(1 - b).$$

Proceeding in this manner, we obtain the transition matrix

$$P = \begin{matrix} 2 \\ 1 \\ 0 \end{matrix} \overset{\begin{matrix} 2 & \qquad\qquad 1 & \qquad\qquad 0 \end{matrix}}{\begin{pmatrix} (1-\theta)^2 a + 1 - a & 2\theta(1-\theta)a & \theta^2 a \\ \frac{1}{2}\theta^2(1-a) + \frac{1}{2}\theta(2-\theta)b & (1-\theta)^2 + \theta(1-\theta)(1-a) + \theta(1-\theta)(1-b) & \frac{1}{2}\theta^2(1-b) + \frac{1}{2}\theta(2-\theta)a \\ \theta^2 b & 2\theta(1-\theta)b & (1-\theta)^2 b + (1-b) \end{pmatrix}}.$$

In the next section we shall study this Markov chain in more detail.

EXERCISES

1. Construct a tree to show the possibilities for the connections after an experiment if the two stimulus elements are both connected to A_1 at the beginning of the experiment. Do the same for the case of no elements connected to A_1 at the beginning of the experiment.

2. Using the trees in Exercise 1, verify that the transition probabilities $p_{0,j}$ and $p_{2,j}$ given above are correct.

3. What is the probability that the subject will make response A_1 if at the beginning of the experiment one element is connected to each response? What is this probability if at the beginning of the experiment both elements are connected to response A_1? [*Ans.* $\frac{1}{2}$, 1.]

In the following exercises, find the matrix of transition probabilities under the special assumptions given in the problem. State whether the resulting Markov chain is absorbing or regular. Give an interpretation for each of the special cases in terms of the actual experiment. If the process is regular, find the limiting probabilities. If the process is absorbing, find the expected number of steps before absorption for each possible starting state. (See Chapter V, Section 8.)

4. $a = 1, b = 1, \theta = \frac{1}{2}$. [*Ans.* Regular; (.3, .4, .3).]

5. $a = 1, b = 0$. [*Ans.* Absorbing; $t_2 = (3 - 2\theta)/(2\theta - \theta^2)$; $t_1 = 1/\theta$.]

6. $a = \frac{1}{4}, b = \frac{3}{4}, \theta = .1$.

7. $a = 0, b = \frac{1}{2}, \theta = \frac{1}{2}$.

8. $a = 1, b = \frac{1}{2}, \theta = \frac{1}{2}$.

9. $a = 0, b = 0$.

10. Work out the transition matrix of the Markov chain for the model having a single stimulus element.

11. Assume that $a > 0$ and $b > 0$ for the one-element model. Show that the chain is regular, and find the limiting probabilities.

[*Ans.* $b/(a + b)$, $a/(a + b)$.]

12. Assume that $a > 0$ and $b = 0$ for the one-element model. Find the expected number of steps to absorption. [*Ans.* $1/\theta a$.]

5. *LIMITING PROBABILITIES IN THE ESTES MODEL*

We wish now to study the limiting probabilities that the subject and that the experimenter will choose each of the possible alternatives.

If our process is in state 0 on a given experiment, then the probability that the subject will make response A_1 is (by assumption A) equal to 0. If it is in state 1, then by symmetry this probability is $\frac{1}{2}$. If it is in state 2, it is (by assumption A) equal to 1.

The matrix P will be regular if and only if the quantities a and b are not zero (see Exercise 1). If the matrix is regular, then there will be a limiting probability for being in each of the states. These probabilities can be represented by a vector $p = (p_0, p_1, p_2)$ and found by solving the equations

$$pP = p.$$

If these equations are solved, we obtain

$$p_2 = \frac{b\theta + 2b^2(1 - \theta)}{(a + b)\theta + 2(a + b)^2(1 - \theta)},$$

$$p_1 = \frac{4ab(1 - \theta)}{(a + b)\theta + 2(a + b)^2(1 - \theta)},$$

$$p_0 = \frac{a\theta + 2a^2(1 - \theta)}{(a + b)\theta + 2(a + b)^2(1 - \theta)}.$$

From these probabilities we can find that the limiting probability that the subject will make response A_1 is

$$1 \cdot p_2 + \tfrac{1}{2}p_1 + 0 \cdot p_0 = \frac{b}{a + b},$$

and that the limiting probability that the subject makes response A_2 is $a/(a + b)$.

To find the probability that the experimenter makes the choice E_1, we must multiply the probabilities for each of the choices of the subject, by the probabilities that the experimenter does E_1 if the subject made the particular choice. Thus the limiting probability that the experimenter makes choice E_1 is

$$\frac{b(1-a)}{a+b} + \frac{ab}{a+b} = \frac{b}{a+b}.$$

Thus we see that the limiting probability that the subject will make response A_1 is equal to the limiting probability that the experimenter will choose E_1. From the limiting probabilities we can also find the limiting probability that the subject will guess correctly (see Exercise 3).

If we assume that the experimenter makes response E_1 with probability p independent of the choice of the subject, the subject can maximize the expected number of correct responses by always making response A_1 if $p > \frac{1}{2}$ and always making A_2 if $p < \frac{1}{2}$. (See Exercise 5.) The model predicts a less rational choice on the part of the subject. This would not seem disturbing in the case of the rat, but it would be hoped humans would do better. Unfortunately, experiments have borne out that the model's predictions are approximately correct even with human subjects.

The following interesting experiment was performed by W. K. Estes and others with many types of subjects. If the subject does A_1, he is rewarded half the time; if he does A_2 he is never rewarded. One might expect that the subject will learn to do A_1, but this is not the case. What does the theory predict? If A_1 is chosen, reward follows half the time. Hence $a = \frac{1}{2}$. If A_2 is chosen, reward never follows. Hence $1 - b = 0$ or $b = 1$. The theory predicts a limiting probability of $b/(a + b) = \frac{2}{3}$ for the subject to choose A_1, which is in good agreement with experimental results.

We next consider an absorbing case. Specifically, we consider the case $a = 0$ and $b = 1$. This means that the experimenter always does E_1. The matrix of transition probabilities here is

$$P = \begin{matrix} \\ 2 \\ 1 \\ 0 \end{matrix} \begin{matrix} \begin{matrix} 2 & \quad 1 & \quad\quad 0 \end{matrix} \\ \begin{pmatrix} 1 & 0 & 0 \\ \theta & 1-\theta & 0 \\ \theta^2 & 2\theta(1-\theta) & (1-\theta)^2 \end{pmatrix} \end{matrix}.$$

We shall use the methods developed in Chapter V to study this Markov chain. We have one absorbing state, namely, 2. Thus we know that the process will eventually enter this state and remain there. Being in this state means, by assumption A of the previous section, that the subject is sure to make response A_1. Thus being absorbed can be interpreted as the subject "learning" that the experimenter always does E_1.

We have seen that in an absorbing Markov chain it is possible to find the expected number of times that the process will be in each of the states before being absorbed, assuming some given starting state. Let n_{ij} be the expected number of times the process will be in state j if it starts in state i. Before calculating n_{ij} we consider what the knowledge of these quantities would tell us about the experiment. We observe that every time the process is in state 1, the subject chooses A_2 with probability $\frac{1}{2}$ and hence makes a wrong response with probability $\frac{1}{2}$. Every time the process is in state 0, the subject is sure to make response A_2, that is, to make a wrong response. Thus the expected number of wrong responses that the subject will make before learning is

$$\tag{1} \tfrac{1}{2}n_{i1} + n_{i0} \quad \text{for} \quad i = 0, 1$$

assuming that the process starts in state i.

We find the n_{ij} as in Chapter V. We first form the truncated matrix Q obtained from P by omitting the column and the row corresponding to the absorbing state.

$$Q = \begin{pmatrix} 1-\theta & 0 \\ 2\theta(1-\theta) & (1-\theta)^2 \end{pmatrix}.$$

We then find $(I - Q)^{-1}$ to be

$$N = (I - Q)^{-1} = \begin{array}{c} \\ 1 \\ 0 \end{array} \begin{array}{cc} 1 & 0 \\ \left(\begin{array}{cc} \dfrac{1}{\theta} & 0 \\ \dfrac{2(1-\theta)}{\theta(2-\theta)} & \dfrac{1}{\theta(2-\theta)} \end{array}\right) \end{array}.$$

Then from (1) we obtain $1/2\theta$ as the expected number of wrong responses if the process begins in state 1, and $1/\theta$ as the expected number of wrong responses if the process begins in state 2.

Of course it is true that in an actual experiment the starting state would not be known. However, it is not unreasonable to assume that on the first experiment the stimuli elements are connected at random. This would mean that the process starts at state 0 with probability $\frac{1}{4}$, at state 1 with probability $\frac{1}{2}$, and at state 2 with probability $\frac{1}{4}$. Thus under this assumption the expected number of wrong responses before learning is

$$\tag{2} \frac{1}{2}\cdot\frac{1}{2\theta} + \frac{1}{4}\cdot\frac{1}{\theta} = \frac{1}{2\theta}.$$

EXERCISES

1. Prove that the matrix P in Section 4 is regular if and only if a and b are different from zero. [*Hint:* Show that if either quantity is 0 the chain is not regular.]

2. Verify that the probability that the subject makes response A_2 is $a/(a + b)$ by finding $1 \cdot p_0 + \frac{1}{2} \cdot p_1 + 0 \cdot p_2$.

3. Show that the limiting probability that the subject's choice agrees with that of the experimenter is

$$\frac{a(1 - b) + b(1 - a)}{a + b}.$$

4. Assume that the experimenter always chooses E_1 with a fixed probability p, independent of the choice of the subject. What proportion would the subject expect to guess correctly? [*Ans.* $1 - 2p + 2p^2$.]

5. Suppose under the conditions of Exercise 4 that the subject were always to make response A_1. Show that if $p > \frac{1}{2}$, then on the average the subject will do better by this method than by the method predicted by the model.

6. Consider the case $a = \frac{1}{2}$, $b = 0$, and $\theta = \frac{1}{2}$. For each possible starting state find the expected number of times that the process will be in each of the states before being absorbed. [*Ans.* $n_{22} = 3$; $n_{21} = 2$; $n_{12} = \frac{1}{2}$; $n_{11} = 3$.]

7. Do the same as in Exercise 6, for the case $a = 0$, and $b = 0$.

8. In Exercises 6 and 7 find the expected number of incorrect responses that the subject will make, assuming each possible starting state. [*Ans.* 0, 2, 4; 0, 0, 0.]

9. In Exercises 6 and 7 find the expected number of incorrect responses that the subject will make assuming random connections for the stimuli elements on the first experiment, as in (2).

10. If the subject chooses A_1, he is rewarded with probability p. If he chooses A_2, he is never rewarded. (See the example with $p = \frac{1}{2}$ in the text above.) Find a and b. What is the limiting probability that the subject chooses A_1? How often is he rewarded? How often would he be rewarded if he always chose A_1? Compare these two values for $p = \frac{3}{4}, \frac{1}{2}, \frac{1}{4}$. [*Ans.* $1/(2 - p)$; $p/(2 - p)$; p.]

11. Compute p_0, p_1, p_2 for the cases given in Section 4, Exercises 4–9. For the regular matrices verify that these are the limiting probabilities there obtained. What do p_0, p_1, p_2 mean for the absorbing chains?

6. MARRIAGE RULES IN PRIMITIVE SOCIETIES

In some primitive societies there are rigid rules as to when marriages are permissible. These rules are designed to prevent very close relatives from marrying. The rules can be given precise mathematical formulation in terms of permutation matrices. Our discussion is based, in part, on the work of André Weil and Robert R. Bush.

The marriage rules found in these societies are characterized by the following axioms.

Axiom 1. Each member of the society is assigned a marriage type.

Axiom 2. Two individuals are permitted to marry only if they are of the same marriage type.

Axiom 3. The type of an individual is determined by the individual's sex and by the type of his parents.

Axiom 4. Two boys (or two girls) whose parents are of different types will themselves be of different types.

Axiom 5. The rule as to whether a man is allowed to marry a female relative of a given kind depends only on the kind of relationship.

Axiom 6. In particular, no man is allowed to marry his sister.

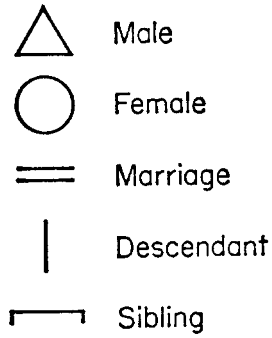
Axiom 7. For any two individuals it is permissible for some of their descendants to intermarry.

Example. Let us suppose that there are three marriage types, t_1, t_2, t_3. Two parents in a given family must be of the same type, since only then are they allowed to marry. Thus there are only three logical possibilities for marriages. For each case we have to state what the type of a son or a daughter will be.

Type of both parents	*Type of their son*	*Type of their daughter*
t_1	t_2	t_3
t_2	t_3	t_1
t_3	t_1	t_2

We must verify that all the axioms are satisfied. Some of the axioms are easy to check (see Exercise 1), others are harder to verify. We will prove a general theorem which will show that this rule satisfies all the axioms.

In order to give a complete treatment to this problem, we must have a simple systematic method of representing relationships. For this we use family trees, as drawn by anthropologists. The following symbols are commonly used.



In Figure 20 we draw four family trees, representing the four kinds of first-cousin relationships between a man and a woman.

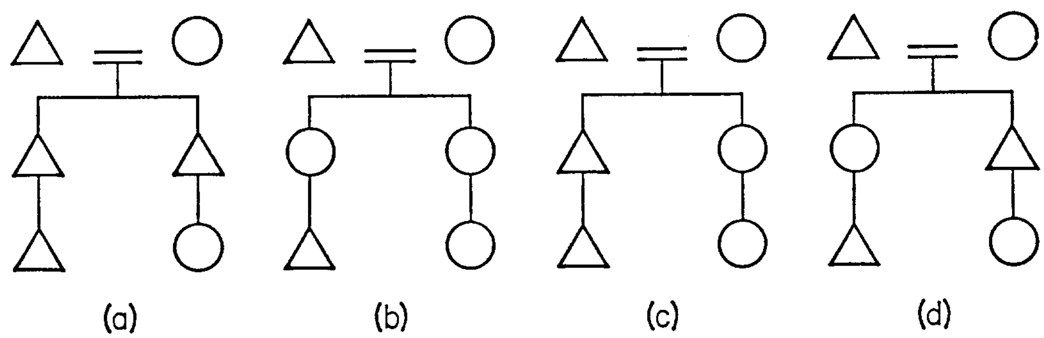


Figure 20

Example (continued). Does our rule allow marriage between a man and his father's brother's daughter? This is the relationship in Figure 20a. There are three possible types for the original couple (the grandparents) and in Figure 21 we work out the three cases. We find in each

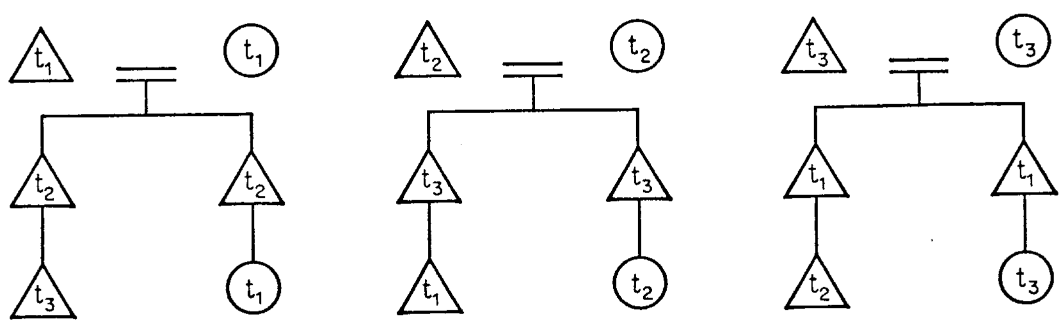


Figure 21

case that the man and woman are of different type, hence such marriages are *never* allowed. Can a man marry his mother's brother's daughter? This is the relationship in Figure 20d. The three cases for this relationship are found in Figure 22. We find that such marriages are *always* allowed.

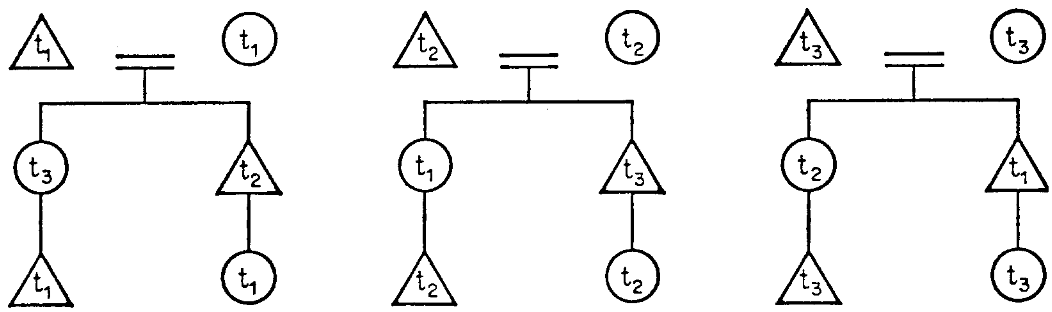


Figure 22

We are now ready to give the rules a mathematical formulation. The society chooses a number, say n, of marriage types (Axiom 1). We call these $t_1, t_2, \ldots, t_n$. Our rule has two parts, one concerning sons, one concerning daughters. Let us consider the marriage type of sons. The parents must be of the same marriage type (Axiom 2). We must assign to a boy a type which depends only on the common type of his parents (Axiom 3). If his parents are of type t_i, he will be of type t_j. Furthermore, if some other boy has parents of a type different from t_i, then the boy will be of type different from t_j (Axiom 4). This defines a *permutation* of the marriage types (see Chapter V, Section 10); the type of a son is obtained from the type of his parents by a permutation specified by the rule of the society. Hence we form the type vector $t = (t_1, \ldots, t_n)$ and represent the permutation in question by the $n \times n$ permutation matrix S. If the type of the parents is component i of t, the type of their sons is component i of tS. By a similar argument we arrive at the permutation matrix D giving the type of daughters.

We have shown that the mathematical form of the first four axioms is to introduce the row vector t and the two permutation matrices S and D. The last three axioms restrict the choice of S and D. This will be considered in the next section.

We have repeatedly seen how the vector and matrix notation allows us to replace a series of equations by a single one. In the present problem this notation allows us to work out a given kind of relationship for *all* marriage types in a single diagram. As a matter of fact, this can be done without knowing how many types there are in the given society, or knowing what the rules are. Let us illustrate this in terms of Figure 22. The couple at the top of the tree is of a given type, represented by our vector t. Their son is of type tS, their daughter of type tD. Then the son of a son is of type tSS, the son's daughter is of type tSD, etc. We arrive at the single vector diagram of Figure 23. If in this figure we take t to have three components, then the diagram is a shorthand for the three diagrams of Figure 22.

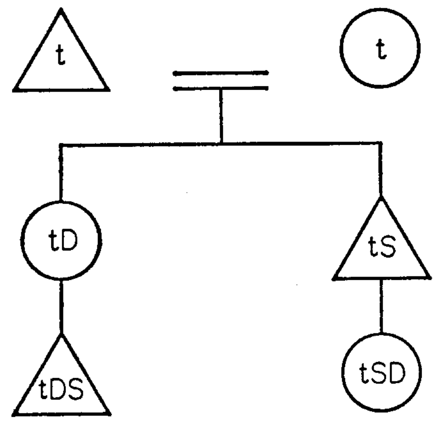


Figure 23

Example (continued). Our t vector is (t_1, t_2, t_3) and

$$D = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}, \qquad S = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}.$$

We know from Figure 22 that a man is always allowed to marry his mother's brother's daughter. Can we see this in Figure 23? The marriage will always be permitted if tDS always equals tSD, which is equivalent to the matrix equation $DS = SD$. It so happens for our S and D that this equation is correct. But we can see more from Figure 23. No matter how many types there are, this kind of marriage will be permitted if and only if $SD = DS$, i.e., if the two matrices commute.

We have now seen one example of how the nature of S and D determines which kinds of relatives are allowed to marry. This question will be the subject of the next section.

EXERCISES

1. In the example above, verify that the rule satisfies Axioms 1, 3, and 4.

2. In the example above, verify that the matrices S and D given represent the rule given.

3. Construct a diagram for the brother-sister relationship.

4. Using the diagram of Exercise 3, show that, in the above example, brother-sister marriages are never permitted.

5. Find the condition on S and D that would always allow brother-sister marriages. [*Ans.* $S = D$.]

In the *Kariera* society there are four marriage types, assigned according to the following rules:

Parent type	*Son type*	*Daughter type*
t_1	t_3	t_4
t_2	t_4	t_3
t_3	t_1	t_2
t_4	t_2	t_1

Exercises 6–11 refer to this society.

6. Find the t, S, and D of the Kariera society.

7. Show that brother-sister marriages are never allowed in the Kariera society.

8. Show that S and D commute. What does this tell us about first-cousin marriages in the Kariera society?

9. Show that first cousins of the kinds in Figures 20(a) and (b) are never allowed to marry in the Kariera society.

10. Show that first cousins of the kind in Figure 20(c) are always allowed to marry in the Kariera society.

11. Find the group generated by S and D of the Kariera society. (See Chapter V, Section 11.)

In the *Tarau* society there are also four marriage types. A son is of the same type as his parents. A daughter's type is given by:

Parent type	*Daughter type*
t_1	t_4
t_2	t_1
t_3	t_2
t_4	t_3

Exercises 12–17 refer to this society.

12. Find the t, S, and D of the Tarau society.

13. Show that brother-sister marriages are never allowed in the Tarau society.

14. Show that S and D commute. What does this tell us about first-cousin marriages in the Tarau society?

15. Show that first cousins of the kinds in Figures 20(a) and (b) are never allowed to marry in the Tarau society.

16. Show that first cousins of the kind in Figure 20(c) are never allowed to marry in the Tarau society.

17. Find the group generated by S and D of the Tarau society. (See Chapter V, Section 11.)

7. THE CHOICE OF MARRIAGE RULES

In the last section we saw that the marriage rules of a primitive society are determined by the vector t and the matrices S and D. The axioms make no mention of the number of types, and indeed, we will find that we can have any number of types, as long as $n > 1$. But we will find that the choice of S and D are severely limited. This shows that the rules of existing primitive societies required considerable ingenuity for their construction.

We must now consider the last three axioms. For Axiom 5 we need a simple way of describing a kind of relationship. The family tree is our basic tool, but we want to replace the family tree by a suitable matrix.

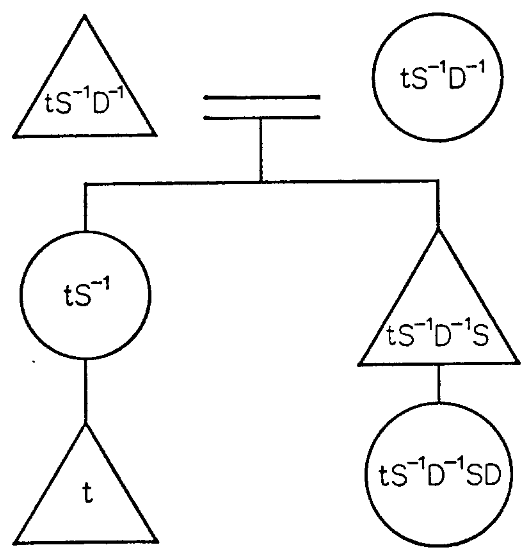


Figure 24

Let us consider Figure 23. Instead of starting with the grandparents and finding the types of the grandson and the granddaughter, we could start with the grandson, work up to the grandparents, and then down to the granddaughter. For this we must consider how we work "up." If a parent is of type t, the son is of type tS. Hence, if the son is of type t, then the parent is of type tS^{-1} (see Chapter V, Section 10). Similarly, if a daughter has type t, her parents have type tD^{-1}. In Figure 24 we find the new version of Figure 23.

It is easily seen that we can follow this procedure for any relationship. Given a kind of relationship, it determines a matrix M such that if the male of the relationship is of type t, then the female is of type tM. From Figure 24 we see that for "mother's brother's daughter" $M = S^{-1}D^{-1}SD$. We will speak of M as the *matrix of the relationship.* These matrices are all products of S, D, and their inverses, hence each matrix is an element of the group generated by S and D.

Let us consider Axiom 5. Given any kind of relationship between a man and a woman, we form the matrix of the relationship M. The man will be permitted to marry this relation of his if and only if his type is the same as hers, i.e., if a certain component of t is the same as the corresponding component of tM. This means that this component is left unchanged by the permutation M, which proves our first theorem. (See Chapter V, Section 11.)

Theorem 1. A man is allowed to marry a female relative of a certain kind if and only if his marriage type does not belong to the effective set of the matrix of the relationship.

A second result follows from this theorem easily.

Theorem 2. Marriage between relatives of a given kind is always permitted if the matrix of the relationship has an empty effective set; it is never permitted if the matrix has a universal effective set.

Theorem 3. Axiom 5 requires that in the group generated by S and D every element except I is a complete permutation.

Proof. The axiom states that for a given relationship the marriage must always be allowed or must never be allowed. Hence, by Theorem 2, the matrix of the relationship must have an empty effective set or a universal one. The former means that the matrix is I, the latter that it is a complete permutation (see Chapter V, Section 11). Hence the matrix of every relationship must either be I or a complete permutation matrix. The matrices are elements of the group generated by S and D. And given any element of this group, which can be written as a product of S's and D's, we can draw a family tree having this matrix. Hence the matrices of relationships are all the elements of the group. This means that all the elements of the group, other than the identity, must be complete permutations. This completes the proof.

Theorem 4. Axiom 6 requires that $S^{-1}D$ be a complete permutation.

This theorem is an immediate consequence of the fact that the matrix of the brother-sister relationship is $S^{-1}D$.

Theorem 5. Axiom 7 requires that for every i and j there be a permutation in the group which carries t_i into t_j.

Proof. Let us choose two individuals, one of type t_i and one of type t_j. There must be a descendant of the former who can marry a descendant of the latter. Hence the two descendants must have the same type. This means that we have permutations M_1 and M_2 such that t_i is carried by M_1 into the same type as t_j by M_2. Then $M_1M_2^{-1}$ carries t_i into t_j. Hence the theorem follows.

We have now translated Axioms 5–7 into the following three conditions on S and D: (1) The group generated by S and D consists of I and of complete permutations. (2) $S^{-1}D$ is a complete permutation. (3) For every pair of types there is a permutation in the group that carries one type into the other.

DEFINITION. A permutation group is called *regular* if (a) it is complete, i.e., every element of the group other than I is a complete permutation and if (b) for every pair from among the n objects there is a permutation in the group that carries one into the other.

Basic theorem. To satisfy the axioms we must choose two different $n \times n$ permutation matrices S and D which generate a regular permutation group.

Proof. Conditions (1) and (3) above state precisely that the group generated by S and D be regular. In a regular group every element other than I is a complete permutation; hence condition (2) requires only that $S^{-1}D$ be different from I. Since $S^{-1}D = I$ is equivalent to $D = S$, we need only require that $D \neq S$. This completes the proof.

It is important to be able to recognize regular permutation groups. Here we are helped by a very simple, well-known theorem: A subgroup of the group of permutations of degree n is regular if and only if it has n elements and is complete.

This leads to a relatively simple procedure. We choose n. Then we must pick a group of $n \times n$ permutation matrices which has n elements and is complete, and select two different elements which generate the group. This is always possible if $n > 1$ (see Exercise 11). One of these is chosen as S and one as D. Since there are not very many regular permutation groups for any n, the choice is very limited.

Example. Let us find all possibilities for a society having four marriage types. First of all we must find the regular subgroups of the symmetric group of degree 4, i.e., the groups of permutations on four objects that have four elements and are complete.

Among these we find cyclic groups. Any two of these groups have the same structure and hence lead to equivalent rules. Let us suppose that we choose the permutation group generated by

$$P = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{pmatrix}.$$

The group consists of P, P^2, P^3, and I. Either P or P^3 generates the group, and they play analogous roles. We may therefore assume that P is one of the two permutations chosen. This allows us (P, P^2), (P, P^3), and (P, I) as possibilities. We must still ask which is S and which is D. In the second case it makes no difference, since P and P^3 play analogous roles in the group, but there is a difference in the first two cases. This leads to five possibilities:

1. $S = P$, $D = P^2$
2. $S = P^2$, $D = P$
3. $S = P$, $D = P^3$
4. $S = P$, $D = I$
5. $S = I$, $D = P$. This is the Tarau society.

There is only one noncyclic complete subgroup with four elements, consisting of I and the three permutations which interchange two pairs of elements. In this group we have essentially only one case, since all three permutations play the same role.

6. The Kariera society. (See exercises after the last section.)

Two of these six possibilities are actually exemplified in known primitive societies.

EXERCISES

1. Figure 24 shows the matrix of one of the first-cousin relations. Find the matrices of the other three first-cousin relationships.

2. Prove that marriage between relations of a certain kind is permitted if and only if the matrix of the relation is I.

3. Use the result of Exercise 2 to prove that no society allows the marriage between cousins of the types in Figures 20(a) and (b).

4. Which of the six rules described above (in the example) allow marriage between a man and his father's sister's daughter? [*Ans.* 3, 6.]

5. Show that all six rules given in the example above allow marriages between a man and his mother's brother's daughter.

6. There are eight kinds of second-cousin relationships between a man and a woman. Draw their family trees.

7. Find the matrices of the eight second-cousin relationships.

8. Are there any second-cousin relationships for which marriage is forbidden by all possible rules? [*Ans.* Yes.]

9. Test the second-cousin relationships (other than those found in Exercise 8) for each of the six rules given in the example above.

10. For n objects, consider the permutation that carries object number i into position $i + 1$, except that the last object is put into first place. Show that the cyclic group generated by this permutation is regular.

11. Use the result of Exercise 10 to show that a society can have any number of marriage types, as long as the number is greater than one.

12. In the Example of Section 6, prove that S and D generate a regular permutation group.

13. Prove that the following matrices lead to a rule satisfying all axioms.

$$S = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix}, \qquad D = \begin{pmatrix} 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

14. Prove that the rule given in Exercise 13 allows no first-cousin marriages.

8. MODEL OF AN EXPANDING ECONOMY

The following model is a modification of a model proposed by John von Neumann. It is designed to study an economy which is expanding at a fixed rate, but which is otherwise in equilibrium. The model makes certain assumptions about how an economy behaves in equilibrium. These assumptions are idealizations, and it is to be expected that the model will eventually be replaced by a better model. For the present many economists consider the von Neumann model to be a reasonable approximation of reality. Our interest in the model is purely to illustrate how finite mathematics is used in an economic problem.

The economy is described by n *goods* and m *processes*. A good may be steel, coal, houses, shoes, etc. Goods are the materials of production in the economy. Each good may be measured in any convenient units, as long as the units are fixed once and for all. It is convenient to be able to talk of arbitrary multiples of these units; e.g., we will consider not only 2.75 tons of steel but also 2.75 houses. The latter may be interpreted as an average.

A manufacturing process needs certain goods as raw materials (the *inputs*) and produces one or more of our goods (the *outputs*). As a process we may, for example, consider the conversion of steel, wood, glass, etc. into a house. Of course this process may be used to manufacture more than one house, and hence we have the concept of the *intensity* with which a process is used. One of the basic assumptions is one of linearity, i.e., that k houses will require k times as much of each raw material. Thus we choose an arbitrary "unit intensity" for each process, and the process is completely described if we know the inputs necessary for this unit operation and the outputs produced.

Process number i when operating at unit intensity will require a certain amount of good j as an input. This amount will be called a_{ij}. (In particular, if good j is not needed for process i, then $a_{ij} = 0$.) We will call b_{ij} the amount of good j produced by process i. Here we allow a process to produce several different goods (e.g., a principal output and by-products). But, of course, we allow processes that produce only one good. Then all the b_{ij} for this i will be 0, except for one. The a_{ij} and b_{ij} are nonnegative numbers.

We define the matrix A to be the $m \times n$ matrix having components a_{ij}, and B to be the $m \times n$ matrix with components b_{ij}. Then the entire economy is described by these two matrices.

We must still consider the element of time. It is customary to think of the economy as working in stages or cycles. In one such stage there is just time enough for process i to convert the inputs a_{ij} to outputs b_{ij}. Then, in the next stage, these outputs may in turn be used as inputs. The length of this cycle may be any time interval convenient for the study of the particular economy. It may be a month, a year, or a number of years.

Example. Let us take as our economy a chicken farm. Our goods are chickens and eggs, with one chicken and one egg being the natural units. Our two processes consist of laying eggs and hatching them. Let us assume that in a given month a chicken lays an average of 12 eggs if we use it for laying eggs. If used for hatching, it will hatch an average of four eggs per month. From this information we can construct A and B.

Our cycle is of length one month. Good 1 is "chicken," good 2 is "egg," process 1 is "laying," and process 2 is "hatching." The unit of intensity of a process will be what one chicken can do on the average in a month. The input of process 1 is one chicken, i.e., one unit of good 1. The output will consist of a dozen eggs *plus* the original chicken. (We must not forget this, since the original chicken can be used again in the next cycle.) Hence the output is one unit of good 1 and 12 units of good 2. In process 2 the inputs are one chicken and four eggs, while the output consists of five chickens (the original one plus the four hatched). Hence our matrices are

$$\begin{array}{l} \text{Laying eggs:} \\ \text{Hatching eggs:} \end{array} \quad A = \overset{\quad\text{Chicken}\quad\text{Egg}}{\begin{pmatrix} 1 & 0 \\ 1 & 4 \end{pmatrix}}, \qquad B = \overset{\quad\text{Chicken}\quad\text{Egg}}{\begin{pmatrix} 1 & 12 \\ 5 & 0 \end{pmatrix}}.$$

Suppose that our farmer starts with three chickens and eight eggs ready for hatching. He will need two chickens for hatching the eight eggs, and this leaves him one for laying eggs. Hence he uses process 1 with intensity 1, process 2 with intensity 2. We symbolize this by the vector $x = (1, 2)$. Note that his inputs are the components of xA. His one laying chicken will lay 12 eggs. He will end up with his original three chickens plus eight new ones. Hence he will have an output of 11 units of good 1 and 12 units of good 2. These are the components of xB. Of his 11 chickens only three can be used for hatching, hence he will employ intensities (8, 3). The outputs will be $(8, 3)B = (23, 96)$, as can easily be checked (see Exercise 1). He now has 96 eggs and only 23 chickens, so that some eggs must go unhatched.

On the other hand, suppose that he starts with only two chickens and four eggs. He will then use intensity (1, 1). His laying chicken lays 12 eggs, and with four newly hatched chickens he has a total of six chickens. This result is also given by $(1, 1)B = (6, 12)$. He now has tripled both his chickens and his eggs. He can use intensity (3, 3) on the next cycle, yielding $(3, 3)B = (18, 36)$, which again triples both the chickens and the eggs. Thus he can continue to use the same proportion of the processes, and will continue to triple his output on every cycle. This economy operates in *equilibrium.*

As was seen in the example, the natural way to represent the intensities of our processes is by means of a row vector. Let x_i be the intensity with which process number i is operated, then the *intensity vector* x is $(x_1, \ldots, x_m)$. Matrix multiplication is then an easy way of finding the total amount of each good needed, and the totals produced. Component j of xA is the sum $x_1a_{1j} + \ldots + x_ma_{mj}$; where x_1a_{1j} is the amount of good j we are using in process 1, x_2a_{2j} the amount we use in process 2, etc. Hence the jth component of xA is the total amount of good j needed in the inputs. Similarly, xB gives the total amounts of the various goods in the outputs.

We must now introduce prices for the various goods. Let y_j be the price of a unit of good j; this must be nonnegative, but it may be zero. (The latter represents a good that is so cheap as to be "practically free.") It is assumed that k units of good j will cost ky_j. The *price vector* y is the column vector

$$\begin{pmatrix} y_1 \\ \cdot \\ \cdot \\ \cdot \\ y_n \end{pmatrix}.$$

Let us consider the products Ay and By. In Ay the ith element is $a_{i1}y_1 + \ldots + a_{in}y_n$; the product $a_{i1}y_1$ is the amount of good 1 needed for unit operation of process i multiplied by the per unit price of good 1, hence this is the cost of good 1 used in the process, $a_{i2}y_2$ is the cost of good 2 used, etc. Hence the ith component of Ay is the total cost of inputs for a unit intensity operation of process i. Similarly, By gives the cost (value) of the outputs.

Finally, we consider the products xAy and xBy. Since x is $1 \times m$, the matrices $m \times n$, and y is $n \times 1$, each product is 1×1—or a number. An analysis similar to those above shows that xAy is the total cost of inputs if the economy is operated at intensity x, with prices y, and xBy is the total value of all goods produced. (See Exercise 2.)

Example (continued). Suppose that a chicken costs ten monetary units, while an egg costs one unit; then $y = \begin{pmatrix} 10 \\ 1 \end{pmatrix}$. Here

$$Ay = \begin{pmatrix} 10 \\ 14 \end{pmatrix} \quad \text{and} \quad By = \begin{pmatrix} 22 \\ 50 \end{pmatrix}.$$

This means that process 1, laying eggs, multiplies our investment by a factor of 2.2; while process 2, hatching, brings in over \$3.50 for every \$1.00 invested. There will be pressure to use the hens just for hatching—which will create a shortage of eggs, bringing about a drastic change in prices. Suppose now that a chicken costs only six times as much as an egg, i.e., $y = \begin{pmatrix} 6 \\ 1 \end{pmatrix}$. Then

$$Ay = \begin{pmatrix} 6 \\ 10 \end{pmatrix} \quad \text{and} \quad By = \begin{pmatrix} 18 \\ 30 \end{pmatrix}.$$

In this case each process triples our investment, and there will be no undue monetary pressure. Hence the farmer can set up his processes so as to be in equilibrium, and the price structure will be stable.

The remaining factor to be considered is the expansion of the economy. We assume that everything expands at a constant rate, i.e., that there is a fixed *expansion factor* α such that if the processes operate at intensity x in this cycle, they operate at intensity αx during the next cycle, $\alpha^2 x$ after that, etc. There is also something similar to expansion for the money of the economy, namely, that through bearing interest, y units of money in this cycle will be worth βy units after the cycle. We again assume that the *interest factor* β is fixed once and for all in equilibrium. Usually these factors will be greater than 1, but this does not have to be the case. Thus $\alpha = 1$ represents a stationary economy, and $\alpha < 1$ represents a contracting economy.

This completes the survey of the basic concepts. We must now lay down our assumptions concerning the behavior of an economy which is in equilibrium. These assumptions serve as axioms for the system.

First of all, we must assure that we produce enough of each good in each cycle to furnish the inputs of the next cycle. If in a given cycle the economy functions at intensity x, it will function at αx next time. The outputs this time will be xB, while the inputs next time will be αxA; hence we must require:

Axiom 1. $$xB \geq \alpha xA.$$

(When we write a vector inequality, we mean that the inequality holds for every component.) We will of course have to require similar conditions for the future. For example, in the second cycle the outputs are αxB, and the inputs needed for the third cycle are $\alpha^2 xA$. But when we write the condition that the former be greater than the latter, an α cancels, and we have again the same condition as in Axiom 1. Hence this axiom serves for all cycles.

The first condition assures that it is possible for the economy to expand at the constant rate α. We must also assure that the economy is financially in equilibrium. Suppose that the output of some process was worth more than β times the input. Then we would be prepared to pay interest at a larger rate to someone willing to invest in our process. Hence β would increase. Thus, in equilibrium this must not be possible; no process can produce profits at a rate greater than that given by investment. If we operate processes at a unit intensity, then Ay gives the costs of inputs, while By gives the cost of outputs. The latter cannot exceed the former by more than a factor β for any process.

Axiom 2. $$By \leq \beta Ay.$$

The next assumption concerns surplus production. If we produce more of a given good than can be used by the total economy, the price drops sharply as merchants try to get rid of their produce. It is customary to assume, for the sake of simplicity, that such goods are free, i.e., to give them price zero. The vector difference $xB - \alpha xA = x(B - \alpha A)$ gives the amounts of overproduction, i.e., the jth component is positive if and only if good j is overproduced. If we assign price zero to these goods, then in the product of the above vector with y every nonzero factor of the former is multiplied by zero; hence the product of the two vectors will be 0.

Axiom 3. $$x(B - \alpha A)y = 0.$$

Now we turn to the question of whether a given process is worth undertaking. From Axiom 2 we know that no process can yield more profit than investment can. But if it yields any less, it is better not to use it, but rather to invest our money. Hence in Axiom 2 we form the difference $By - \beta Ay$; if the ith component of this is negative, process i should not be used; it must be assigned intensity 0. Similar to the argument used for Axiom 3, this shows that multiplying this vector difference by x must yield zero.

Axiom 4. $$x(B - \beta A)y = 0.$$

Our final assumption is that something worth while is produced in the economy, i.e., that the value of all goods produced is a positive amount.

Axiom 5. $$xBy > 0.$$

If for a given economy (given A and B) we find vectors x and y and numbers α and β which satisfy these five axioms, we say that we have found a *possible equilibrium solution* for the economy.

Example (continued). We have already seen that if $x = (1, 1)$, the economy expands at the fixed rate $\alpha = 3$. We can now check that Axiom 1 is satisfied. Actually, xB turns out to equal αxA. Similarly, we have noted a monetary equilibrium if $y = \begin{pmatrix} 6 \\ 1 \end{pmatrix}$, and each process multiplies the money put into it by a factor of $\beta = 3$. We can check that Axiom 2 holds. Actually By is equal to βAy in this case. From these two equations we also know that $x(B - \alpha A)$ and $(B - \beta A)y$ are identically 0; hence Axioms 3 and 4 hold. Finally, $xBy = 48$; the total value of goods produced is positive, so that Axiom 5 holds. Therefore these values of x, y, α, and β represent an equilibrium for the economy. It can also be shown that these are the only possible values of α and β, and that x and y must be proportional to those shown here (which may be thought of simply as a change in the units).

In our example we found one and only one equilibrium for the economy, and we found that $\alpha = \beta$. This raises several very natural questions: (1) Is there a possible equilibrium for every economy? (2) If yes, then is there only one? (3) Must the expansion factor always be the same as the interest factor? In the next section we will establish the following answers: (1) For every economy satisfying a certain restriction (which is certainly satisfied for all real economies) there is a possible equilibrium. (2) There may be more than one equilibrium, though the number of different possible expansion factors is finite. (In the example there is essentially only one possibility for x and y; however this is not true in general.) (3) The interest and expansion factors are always equal in equilibrium.

EXERCISES

1. In the example, for $x = (1, 2)$, verify for three cycles that xA and xB give the correct inputs and outputs.

2. Give an interpretation of xAy and xBy,
(a) Using the interpretations of xA and xB given above.
(b) Using the interpretations of Ay and By given above.
(c) And show that the results in (a) and (b) are the same.

3. In the example suppose that two chickens lay eggs and three hatch eggs. Find x, xA, and xB. Substitute these quantities into Axiom 1, and find the largest possible expansion factor. [*Ans.* $\alpha = 2$.]

4. In the example, suppose that chickens cost 80 cents and eggs cost five cents. Find y, Ay, and By. Substitute these quantities into Axiom 2, and find the smallest possible interest factor. [*Ans.* $\beta = 4$.]

5. Show that the x, y, α, and β found in two previous Exercises do *not* lead to equilibrium, by showing that Axioms 3 and 4 fail to hold.

6. Show that if $\alpha = \beta = 3$, then the only possible x's and y's are proportional to those given in the example. [*Hint*: Show that the axioms force us to choose $x_1 = x_2$ and $y_1 = 6y_2$.]

The remaining problems refer to the following economy: On a chicken farm there is a breed of chicken that lays an average of 16 eggs a month, and such that they can hatch an average of $3\frac{1}{5} = \frac{16}{5}$ eggs.

7. Set up the matrices A and B.

8. Suppose that three chickens lay and five chickens hatch. Find x, xA, and xB. What is α? [*Ans.* $x = (3, 5)$; $xA = (8, 16)$; $xB = (24, 48)$; $\alpha = 3$.]

9. Suppose that chickens cost 40 cents and eggs five cents. Find y, Ay, and By. What is β?

10. Verify that the x, y, α, and β found in the previous exercises represent an equilibrium for the economy, by substituting these into the five axioms.

11. Suppose that we start with 16 chickens and 32 eggs. Choose the intensities so that the economy will be in equilibrium, and find what happens in the first three months. [*Ans.* $x = (6, 10)$; 432 chickens, and 864 eggs.]

12. Suppose that with 16 chickens and 32 eggs (see Exercise 11) we start out by having only five hatching, the others laying. Show that we cannot have as many chickens after three months as we would have in the equilibrium solution.

9. *EXISTENCE OF AN ECONOMIC EQUILIBRIUM*

We must ask whether the axioms can always be satisfied, i.e., whether the model of the economy allows such an equilibrium.

Of course we are interested only in an economy that could really occur. That means that these goods must be goods that are somehow produced, and that they cannot be produced out of nothing. Hence every process must require at least one raw material and every good has at least one process that produces it. We summarize this:

Restriction. Every row of A and every column of B has at least one positive component.

Theorem. If A and B satisfy the restriction, then an equilibrium is possible.

We will sketch the proof of this theorem. From Axiom 3 we have that $xBy = \alpha xAy$, while from Axiom 4, $xBy = \beta xAy$. Hence $\alpha xAy = \beta xAy$. Furthermore, from Axiom 5 we know that xBy is not zero, hence xAy is not zero. Then $\alpha = \beta$. Hence *in equilibrium the rate of expansion equals the interest rate.*

If $\alpha = \beta$, then Axioms 3 and 4 are equivalent. We can also rewrite the first two axioms (using our result).

Axiom 1′. $x(B - \alpha A) \geq 0.$

Axiom 2′. $(B - \alpha A)y \leq 0.$

If we multiply the first inequality by y on the right, and the second by x on the left, we see that Axiom 3 (and hence 4) follows from these two axioms. Hence we need only worry about Axioms 1′, 2′, and 5.

The key to the proof is to reinterpret the problem as a game-theoretic one. This is done in spite of the fact that no game is involved in the model. We simply use the mathematical results of the theory of games as tools.

Axioms 1′ and 2′ suggest that we think of the matrix $B - \alpha A$ as a matrix game. We would then like to think of the vectors x and y as mixed strategies for the two players. The vectors are nonnegative, but the sum of their components need not be 1. However, we know that multiplying x by a constant can be thought of as a change in the units of intensities, and multiplying y by a constant is equivalent to a change in the units of the various goods. Hence, without loss of generality, we may assume that x and y have component sum 1, and think of them as mixed strategies. If we do this, the two axioms state precisely that the game has value zero, and that x and y form a pair of optimal strategies for the two players. Thus our first problem is to choose α so that the "game" $B - \alpha A$ has value zero.

Example 1. Let us set up the example of the last section as a game.

$$M = B - \alpha A = \begin{pmatrix} 1 - \alpha & 12 \\ 5 - \alpha & -4\alpha \end{pmatrix}.$$

If we choose $x = (\frac{1}{2}, \frac{1}{2})$ as a mixed strategy for the row player, then $xM = [3 - \alpha, 2(3 - \alpha)]$. If $\alpha < 3$, the components are both positive; hence the game has value greater than zero. If we choose $y = \begin{pmatrix} \frac{6}{7} \\ \frac{1}{7} \end{pmatrix}$ as a mixed strategy for the column player, then

$$My = \begin{bmatrix} \frac{6}{7}(3 - \alpha) \\ \frac{10}{7}(3 - \alpha) \end{bmatrix}.$$

If $\alpha > 3$, both components are negative, and hence the game has negative value. We thus see that the only value of α that could possibly give us a zero value of the game is $\alpha = 3$, and we see from the above that in this case the value really is zero, and x and y are optimal strategies. (See Exercise 1.)

We must now show that the above example is typical in that we can always find an α making the value of $B - \alpha A$ equal to zero. We may write this matrix as the sum $B + \alpha(-A)$, and think of our game as a combination of game B and game $-A$.

By our restriction, every column of B has a positive entry. The strategy vector y for the column player must have at least one positive component. Hence in the product By, one of the components at least must be positive. Hence the value of the game B is positive. Since every row of A has a positive entry, every row of the game $-A$ must have a negative entry. Hence at least one component of $x(-A)$ must be negative, and hence $-A$ has a negative value.

In the combination $B + \alpha(-A)$ the second term is negligible for very small α; hence for these the game has positive value. As α increases, we keep adding larger negative quantities to some of the entries of the game, i.e., we keep decreasing some of these entries. Hence the value of the game decreases steadily. For very large α the first term is negligible, and hence the combined game has negative value. For some intermediate value of α the game must have value zero.

Example 1 (continued). The value of the combined game M is plotted for various α in Figure 25. Since B has value $\frac{15}{4}$ and $-A$ has value -1 (see Exercise 2), at the beginning the game M has value nearly $\frac{15}{4}$, and near the end it has value nearly $2 - \alpha$, which is less than zero (see Exercise 3).

We know that there is at least one α for which the game $B - \alpha A$ has value zero. By choosing such an α together with a pair x, y of optimal strategies, we arrive at a set of quantities satisfying Axioms 1′ and 2′. This still leaves the question of Axiom 5.

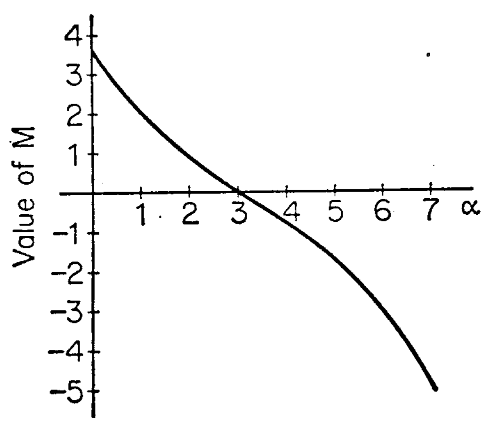


Figure 25

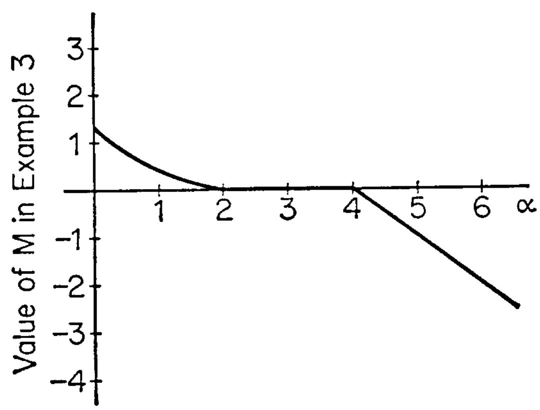


Figure 26

If there are two values of α, say $p < q$, for which the game has value zero, every value between p and q also has this property. This is because the value of the game cannot increase as α increases, as we saw above. Hence we must have a situation such as that shown in Figure 26. It can be shown, however, that most of these values represent methods of procedure where nothing worthwhile is produced, i.e., where Axiom 5 fails. For Axiom 5 to hold, different values of α can be achieved only by using at least one new process. Since there are only a finite number of processes, we can have only a finite number of different possible α's on the interval between p and q. If p is the smallest possible expansion rate and q the largest, then p and q are such that Axiom 5 can be satisfied, and there may be a limited number of additional ones in between.

Example 2. In the chemical industry we are interested in manufacturing compounds P, Q, and R. We assume that the basic chemicals are available in plentiful supply, and that their cost can be neglected for this analysis. But to manufacture compound P we must have a unit of both P and Q available, while to manufacture Q we must have P and R available. Compound R is a by-product of both manufacturing processes. The exact quantities are given by

$$\begin{array}{l} \text{Manufacture of } P\text{:} \\ \text{Manufacture of } Q\text{:} \end{array} \quad A = \overset{\begin{array}{ccc} P & Q & R \end{array}}{\begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}}, \qquad B = \overset{\begin{array}{ccc} P & Q & R \end{array}}{\begin{pmatrix} 6 & 0 & 1 \\ 0 & 3 & 2 \end{pmatrix}}.$$

Then

$$M = B - \alpha A = \begin{pmatrix} 6 - \alpha & -\alpha & 1 \\ -\alpha & 3 & 2 - \alpha \end{pmatrix}.$$

Let us choose

$$x = (\tfrac{1}{2}, \tfrac{1}{2}) \quad \text{and} \quad y = \begin{pmatrix} \frac{1}{6} \\ \frac{1}{3} \\ \frac{1}{2} \end{pmatrix}.$$

Then

$$xM = [3 - \alpha, \tfrac{1}{2}(3 - \alpha), \tfrac{1}{2}(3 - \alpha)] \quad \text{and} \quad My = \begin{bmatrix} \frac{1}{2}(3 - \alpha) \\ \frac{2}{3}(3 - \alpha) \end{bmatrix}.$$

From this we see that if $\alpha < 3$, then the row player has a guaranteed profit, while if $\alpha > 3$, the column player does. Thus $\alpha = 3$ is the only possibility, and for this case the value of the game is zero, and the vectors x and y are optimal strategies, as can be seen from the fact that xM and My have all components zero. Thus there is a unique equilibrium, with $\alpha = \beta = 3$.

We also find that the mixed strategy x is unique, which means that the two processes must be used with the same intensity. However, the strategy y is not unique. We may instead use

$$y' = \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \\ 0 \end{pmatrix} \quad \text{or} \quad y'' = \begin{pmatrix} 0 \\ \frac{1}{4} \\ \frac{3}{4} \end{pmatrix}$$

or any mixture $ty' + (1 - t)y''$, $0 \le t \le 1$. Our y is the case $t = \frac{1}{3}$. Hence we see that different price structures are possible, each leading to the same expansion rate.

Example 3. This "economy" is a schematic representation of the production of essentials and inessentials in a society. Goods are lumped together into two types, E (essential goods) and I (inessential goods or luxury items). For the manufacture of E we need only essential goods (since anything so needed is essential). For the manufacture of I we may need both types of raw materials. Let us suppose that our economy functions as follows.

$$\begin{matrix} \text{Manufacture of essentials:} \\ \text{Manufacture of luxuries:} \end{matrix} \quad A = \overset{\begin{matrix} E & I \end{matrix}}{\begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}}, \qquad B = \overset{\begin{matrix} E & I \end{matrix}}{\begin{pmatrix} 4 & 0 \\ 0 & 2 \end{pmatrix}}.$$

Then

$$M = B - \alpha A = \begin{pmatrix} 4 - \alpha & 0 \\ -\alpha & 2 - \alpha \end{pmatrix}.$$

With a little patience we can determine the values of M for various values of α, and we arrive at the curve in Figure 26. (See Exercise 4.) Hence α must be between 2 and 4. For $\alpha = 4$, we have the optimal strategies $x = (1, 0)$ and $y = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$, which satisfy all our axioms; while for $\alpha = 2$ we have

$$x = (\tfrac{1}{2}, \tfrac{1}{2}) \quad \text{and} \quad y = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

For in-between values of α we cannot satisfy Axiom 5. (See Exercises 5–7.) Hence there are two possible equilibria: (1) The society can decide to manufacture only essentials, in which case the production of these will increase rapidly. (2) By putting a high enough value on inessentials, it will arrive at an equilibrium in which both essentials and inessentials are produced, but then the rate of expansion is considerably decreased.

We have now provided complete answers for the three questions raised at the end of the last section, providing a mathematical solution to a series of economic problems.

EXERCISES

1. In Example 1 verify that for $\alpha = 3$ the game M has value 0, and that the x and y given are optimal strategies.

2. In Example 1 solve the 2×2 games B and $-A$, finding their values and pairs of optimal strategies.

3. In Example 1

(a) Show that the game M is nonstrictly determined for every α.
(b) Find the value of M for any α. [*Ans.* $(5 + \alpha)(3 - \alpha)/(4 + \alpha)$.]
(c) Show that the value for $\alpha = .01$ is very near $\frac{15}{4}$.
(d) Show that the value for $\alpha = 100$ is very near -98.
(e) Show that the value is 0 if and only if $\alpha = 3$.

4. Find the value of M in Example 3 for $\alpha = 0, 1, 2, 3, 4, 5$, and 6. [*Hint:* Some of these games are strictly determined.]

[*Ans.* 1.33, .60, 0, 0, 0, −1.00, −2.00.]

5. In Example 3, for $\alpha = 4$, verify that the strategies given are optimal, and that Axiom 5 is satisfied.

6. In Example 3, for $\alpha = 2$, verify that the strategies given are optimal, and that Axiom 5 is satisfied.

7. In Example 3, for $\alpha = 3$, find the unique optimal x and y, and show that Axiom 5 is *not* satisfied. Prove that the same happens for every α if $2 < \alpha < 4$.

The remaining problems refer to the following economy: There are four goods and five processes, and the economy is given by

$$A = \begin{pmatrix} 0 & 0 & 1 & 1 \\ 0 & 0 & 2 & 2 \\ 0 & 4 & 0 & 2 \\ 2 & 1 & 1 & 0 \\ 0 & 1 & 0 & 2 \end{pmatrix}, \qquad B = \begin{pmatrix} 0 & 0 & 4 & 2 \\ 0 & 0 & 5 & 7 \\ 6 & 5 & 4 & 0 \\ 0 & 4 & 0 & 3 \\ 3 & 0 & 6 & 0 \end{pmatrix}.$$

Also let $x = (\frac{1}{2}, \frac{1}{2}, 0, 0, 0)$, $\quad x' = (0, 0, \frac{2}{5}, \frac{3}{5}, 0)$,

$$y = \begin{pmatrix} \frac{1}{3} \\ \frac{2}{3} \\ 0 \\ 0 \end{pmatrix}, \qquad y' = \begin{pmatrix} 0 \\ 0 \\ \frac{1}{2} \\ \frac{1}{2} \end{pmatrix}.$$

8. Verify that A and B satisfy the restriction.

9. Compute $M = B - \alpha A$.

10. Compute xM, $x'M$, My, and My'.

11. When will $x'M$ have all positive entries? When will My' have all negative entries? What possibilities does this leave for α?

[*Ans.* $\alpha < 2$; $\alpha > 3$; $2 \le \alpha \le 3$.]

12. Show that for the remaining possible values of α the game M has value zero, and x and y are optimal strategies.

13. Show that for the largest possible α the vectors x and y' provide optimal strategies which satisfy Axiom 5.

14. Show that for the smallest possible α the vectors x' and y provide optimal strategies which satisfy Axiom 5.

15. If α is in between its two extreme values, show that
- (a) xM is positive in its last two components, and hence the second player can use only his first two strategies.
- (b) My is negative in its last three components, and hence the first player can use only his first two strategies.
- (c) For these cases it is impossible to satisfy Axiom 5.

16. Process number five is in a special position. Why? [*Ans.* Never used.]

17. Use the results of Exercises 8–16 to show that there are exactly two possible equilibriums for this economy. Interpret each equilibrium, and point out the differences between the two methods of operating the economy.

[*Ans.* At the price of reducing the expansion rate, the economy can produce a larger variety of goods. To achieve this, the additional types of goods must be valued (relatively) very high.]

10. COMPUTER SIMULATION

Probabilistic models prevail in the social sciences. While many of them can, in principle, be treated by the methods studied in this book, in practice they frequently are much too complicated to obtain precise theoretical results. In such cases, simulation by a high-speed computer may be a powerful tool.

Simulation is a process during which the computer acts out a situation from real life. Typically, the relevant facts about an experiment are supplied to the computer, and it is instructed to run through a large series of experiments, perhaps under varying conditions. This enables the scientist to carry out in an hour a series of experiments that would otherwise take years, and at the same time all the important information is automatically tabulated by the computer.

Of course, the computer cannot duplicate the exact circumstances of an experiment. The facts fed to it are based on a model (or theory) formed by the scientist, and the value of the simulation depends on the accuracy of the model. Thus the main significance of simulation is that it enables a scientist to study the kind of behavior predicted by his model. For very complicated models this may be the only procedure open to him.

In addition to the use of simulation for theoretical studies, there are two very important types of pragmatic uses of simulation: (1) It can be used as a planning device. If there are various alternative courses of action open, the computer is asked to try out the various alternatives under different conditions, and report the advantages and disadvantages of each course. (2) Simulation may be used as a training device. For example, business schools make increasing use of "business games" in which fledgling executives may try their skill at decision-making under realistic circumstances. Similarly, simulated "war-games" are used to train military leaders.

We will first discuss how machines simulate stochastic processes, and then illustrate the procedure in four examples. To avoid the necessity of lengthy introduction of new models, we shall use three of the games previously discussed, and a Markov chain model. Also, we will describe the simulation so that no previous knowledge of computers is necessary.

How does one introduce a probabilistic element into a high-speed computer? This is achieved by the generation of so-called *random numbers*. In a typical set-up, when an instruction contains the letters "RND," a real number between 0 and 1 is computed that gives rise to fairly good random results.

Actually, the computer is forced to cheat, in that it has only a finite capacity for expressing numbers. So that it may in reality divide the unit interval into a million (or more) numbers, and give them in a pretty random order. When its supply is exhausted, it will start giving the same numbers in the same order. However, if one needs only 100,000 numbers, or even a million numbers, the results are highly satisfactory.

One use of the RND device is to generate an independent trials process with two outcomes. For example, suppose that we wish to have probability .3 for success. Then on each trial we generate an RND, and ask:

$$\text{Is RND} < .3?$$

If yes, we mark it as success; if no, then it is a failure. Since a number picked at random from the unit interval has a .3 probability of being less than .3, we obtain an excellent approximation to the independent trials experiment.

In Figure 27 we show 30 RND's generated by the Dartmouth Computer. If we used these for the above mentioned simulation, we would have success in 8 of 30 trials, only one below the expected number of 9.

Suppose that we wish to simulate an independent trials experiment with more than two outcomes. If the outcomes are equally likely, then the generation of *random integers* is a very convenient device. In this we generate RND's as usual, but reinterpret them as integers.

For example, in Figure 28 we show the result of multiplying the RND's of Figure 27 by 6, and adding 1 to each. Now we have numbers picked at random between 1 and 7. Since such a number is just as likely to lie between 3 and 4 as between 4 and 5, saving the integer part of the number will result in equally likely random integers 1, 2, 3, 4, 5, and 6. This is shown in Figure 29.

.746489	.196691	.053368	.323690	.244322
.625169	.193130	.935845	.445447	.262310
.218802	.783032	.402600	.848350	.558119
.980484	.918514	.873523	.388814	.393435
.545924	.578063	.638623	.637121	.587565
.952204	.985279	.076776	.096170	.736181

Figure 27

5.47893	2.18015	1.32021	2.94214	2.46593
4.75101	2.15878	6.61507	3.67268	2.57386
2.31281	5.69819	3.41560	6.09010	4.34871
6.88290	6.51108	6.24114	3.33289	3.36061
4.27555	4.46838	4.83174	4.82273	4.52539
6.71322	6.91168	1.46066	1.57702	5.41708

Figure 28

5	2	1	2	2	4	2	6	3	2
2	5	3	6	4	6	6	6	3	3
4	4	4	4	4	6	6	1	1	5

Figure 29

Example 1. Craps. Let us simulate the game of craps on the computer. First of all, we must imitate the roll of a pair of dice. We may do this by choosing a pair of numbers from Figure 29, each number representing one die, and letting the sum represent the sum of the two dice. Then we proceed according to the rules of craps.

A flow-diagram for this simulation is shown in Figure 30. If we carry this out for three games, using the numbers in Figure 29 (reading from left to right in successive rows), we obtain the following results: (1) The player rolls 7, and wins. (2) The player rolls 3, and loses.

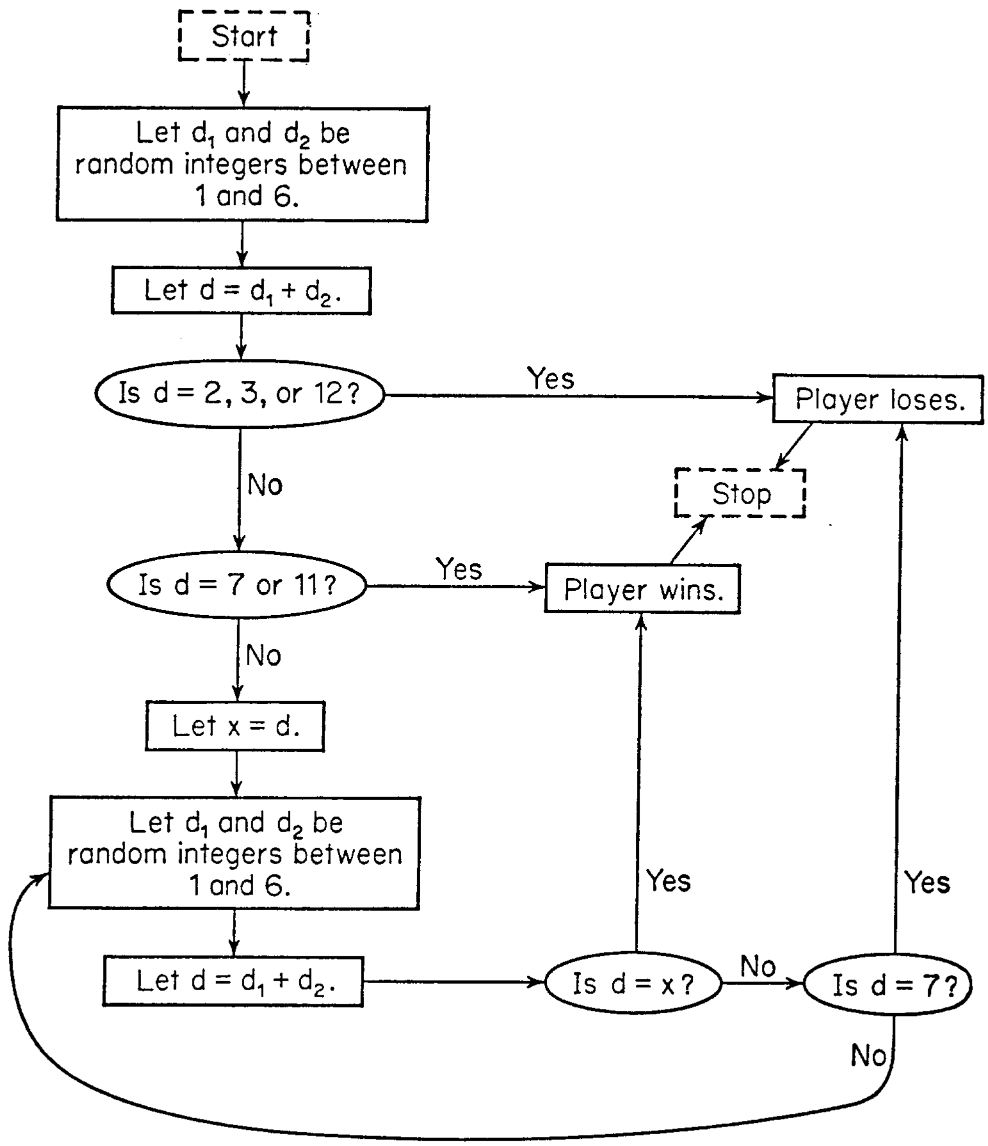


Figure 30

(3) The "point" is 6, but a 7 turns up before a 6, and the player loses again.

Let us use this simulation model to estimate the player's expected value. (In Chapter IV, Section 12, this was found to be $-.0141$.) After 10,000 simulated games the player was behind by \$312, yielding the rather poor estimate of $-.0312$. However, after 250,000 games the estimate was $-.0154$, in good agreement with theory. The total computing time required for 250,000 games of craps was 8 minutes, indicating that this is a practical procedure.

The difficult question to answer, in general, is: "When have we run enough simulated games?" However, in the simple case of craps this is not hard to determine. Since the probability of winning is nearly $\frac{1}{2}$, the standard deviation for n games is roughly $\sqrt{n \cdot \frac{1}{2} \cdot \frac{1}{2}}$ or $\sqrt{n}/2$. Suppose that the player wins that many more games than expected. This increases his winnings by $\sqrt{n}/2$ dollars, and decreases his losses by the same amount. Thus there is a net gain of $\sqrt{n}$ dollars. This changes our estimated expected value by $\sqrt{n}/n$, or by $1/\sqrt{n}$.

Thus, running the game 10,000 times will typically result in an error of about .01, which is quite significant compared to the correct answer of $-.0141$. In our simulation we were off by about .017, which is within two standard deviations (which is .020). However, an error of the same size in the opposite direction would actually have yielded a positive expected value. So we can have little confidence in 10,000 simulated games in estimating a number so near to 0.

However, after 250,000 games the typical error is only about .002. Thus from our computed estimate of $-.0154$ we can have considerable confidence that the game is not favorable to the player; and we would estimate that the correct expected value lies somewhere between $-.012$ and $-.019$.

Example 2. Poker. In the exercises of Chapter IV, Section 3, we computed the probabilities for various poker hands. Let us obtain estimates for the same by simulation.

Our problem here amounts to selecting 5 cards at random from a deck of 52 cards. We first of all number the cards from 1 to 52, in any convenient manner. Then we select one card by generating a random integer from the set 1 through 52. (This can be achieved by computing $52 \cdot \text{RND} + 1$.) Next we select one of the 51 remaining cards at random, etc. When we have five cards, we determine how good a hand we drew.

This simulation was carried out for 10,000 poker hands on the Dartmouth Computer, requiring about one hour of computing time. The results were as in Figure 31.

Type of hand.	Number of times.
Bust	5046
One pair	4169
Two pairs	508
Three of a kind	191
Straight	43
Flush	11
Full house	25
Four of a kind	6
Straight flush	1

Figure 31

You will be asked, in the exercises, to compare these figures with the expected values.

Example 3. Land of Oz. Models in the social sciences often depend on Markov chain processes. While there are powerful theoretical tools for treating Markov chains, sufficiently complex models may have to be simulated. We will illustrate this for a simple Markov chain, which we have already treated theoretically.

Consider the Land of Oz (Chapter V, Section 7, Exercise 13). Suppose that we wished to find the fraction of times that the weather is "nice," "rain," or "snow," by simulation. We would first pick a starting state, say "rain." We then know that the probability of "rain" is $\frac{1}{2}$, of "nice" $\frac{1}{4}$, and of "snow" $\frac{1}{4}$. We can achieve this by generating an RND; if it is less than $\frac{1}{2}$ we decide on "rain," if it is between $\frac{1}{2}$ and $\frac{3}{4}$ then "nice" is next, while if RND $> \frac{3}{4}$ then "snow" is next.

If we use the RND's in Figure 27, we obtain "nice" for the second day. From here we go to "rain" or "snow," with probability $\frac{1}{2}$ each. Since the next RND is less than $\frac{1}{2}$, we choose "rain." Proceeding in this manner, after the original "rain" we obtain

"nice," "rain," "rain," "rain," "rain," "nice," "rain," "snow."

We carried out this simulation for 10,000 times for each starting state, with the results shown in Figure 32. We note that the results are in excellent agreement with the .4, .2, and .4 long-run distribution predicted by theory, and that the results are pretty much independent of the starting state.

Starting state	Number of "rain"	Number of "nice"	Number of "snow"
"Rain"	4080	2002	3918
"Nice"	3976	1986	4038
"Snow"	3975	1984	4041

Figure 32

Example 4. Baseball. The game of baseball is a good example of a game having a model for which a complete theoretical treatment is not practical, and hence much can be gained from simulation.

How would we build a simulation model for a given team, in order to study the way they produce runs? Fortunately, some very detailed statistics are kept, over long periods, which are ideal for building such a model. Let us suppose that a given batter comes to bat. We know from past experience what the probabilities are for his making an out, getting a walk, or getting a hit of various kinds. We simply generate an RND, and use it to decide what the batter did.

For example, if he has probabilities .1 for a walk, .64 for an out, .2 for a single, .03 for a double, .01 for a triple, and .02 for a home-run, we can generate a random integer from 1 through 100, and interpret it as in Figure 33.

We can then bring the next batter to bat, and arrive at a result based on *his* past performance. The running on the bases may be simulated similarly. For example, we can feed into the machine the probability that a man on first reaches third on a single. Just how realistic we wish to make the model depends entirely on how much work we are willing to do.

Range	Result	Probability
1–10	Walk	.1
11–74	Out	.64
75–94	Single	.2
95–97	Double	.03
98	Triple	.01
99–100	Home-run	.02

Figure 33

It should be noted that we are simulating only the batting of *one* team. We do not here consider the batting of the other team, or questions of defensive play.

Such a model would be most useful in training young managers. The computer could make all decisions (many of them stochastic) having to do with the performance of the players, while the manager could make all decisions normally open to managers. For example, he could call for a hit-and-run play, and the machine would simulate the results. He could call for a steal, or send in a pinch-hitter, or tell a batter to try to hit a long fly ball.

By the use of a computer a new manager could gain an entire season's experience in a few days—and he would not be learning at the expense of his team.

The model is also useful for planning purposes, as we will illustrate here. One important task of the manager is to decide on his batting order. He could feed a variety of batting orders to the computer, have it try each for a season's games (or more), and report back the results.

This was actually done on the Dartmouth Computer.

The team used in the simulation was the starting line-up of the 1963 world champion Los Angeles Dodgers. The line-up of Figure 34 was used throughout.

Line-up	Batting average	Slugging average
1. Wills	.302	.349
2. Gilliam	.282	.383
3. W. Davis	.245	.365
4. T. Davis	.326	.457
5. Howard	.273	.518
6. Fairly	.271	.388
7. McMullen	.236	.339
8. Roseboro	.236	.351
9. Pitcher (average)	.117	.152

Figure 34

An entire season of 162 games was simulated, keeping detailed records for each player. Of course, this simulation differed from the normal year in a few respects. For instance, the first eight players played every inning of every game. Since only the batting was simulated, no allowance was made for defensive play, nor did the game stop after eight innings if the home team was ahead. Games were not called on account of rain, and there were no extra-inning games. But, many important features concerning batting were recreated quite realistically. We will cite a few of the more interesting results.

Seven of the batters ended up with batting averages close to their actual ones, but two did not. Tommy Davis, the league's leading hitter, had an even more spectacular year during simulation: He batted an even 350 (compared with 326 in 1963). On the other hand, Fairly who batted 271 in actuality, had a bad simulated year, batting only 250. This shows how much a batting average can change due to purely random factors.

Howard was far ahead in home runs, with 54. This is much higher than the 28 he had in actuality, but he was only used part time in 1963, while in the simulated year he played all the time. Two of the home runs were hit by pitchers—just as in real life. In one game Howard hit three home runs. But mostly it was the balance of the Dodger team that showed up; there were ten games in which three different players hit home runs.

There were no really spectacular slumps, though Gilliam once went 15 consecutive at-bats without getting a hit. The total number of runs scored was 652, in excellent agreement with the actual 640. On the other hand, the 1352 men left on base compared very poorly with the Dodgers' league-leading performance of leaving only 1034 men on base. Two factors in this were the absence of double-plays and pinch-hitters in the simulation model. But there is probably some other relevant attribute of the team that was missed in the model.

Perhaps the most interesting result is the number of shut-outs. There were 11 in the simulation, as compared to the league-leading performance of only eight shut-outs. In the simulation, two of the shut-outs occurred in the final two games. Thus, if the season ended in 160 games, the simulation would have been off by only one shut-out. This shows how hard it is to get an accurate estimate for a small probability through simulation! And there were four games late in the season, three of which ended in shut-outs. If this had happened in real life, all the Los Angeles papers would have carried headlines about a Dodger batting slump.

To compare various possible batting orders, several line-ups were simulated for ten entire seasons. The seven line-ups are shown in the first column of Figure 35, and the results in the second column. The standard deviation of the average number of runs per game was about .07. Since the difference between the best and the worst line-up is over three standard deviations, one is tempted to conclude that the batting order really makes a difference—though not very much of a difference.

Line-up	Average number of runs per game		
	10 seasons	7 × 10 seasons	Range
1, 2, 3, 4, 5, 6, 7, 8, 9	4.06	4.00	3.91–4.06
1, 4, 2, 5, 6, 3, 8, 7, 9	4.07	4.02	3.92–4.07
4, 5, 6, 1, 2, 3, 7, 8, 9	4.00	3.98	3.90–4.04
2, 1, 3, 5, 4, 6, 8, 7, 9	3.98	4.01	3.95–4.08
1, 4, 7, 2, 5, 8, 3, 6, 9	3.90	3.98	3.90–4.05
9, 8, 7, 6, 5, 4, 3, 2, 1	3.89	3.82	3.72–3.89
9, 6, 3, 8, 5, 2, 7, 4, 1	3.83	3.83	3.76–3.92

Figure 35

However, this simulation—though time-consuming—is not conclusive. We may still entertain the hypothesis that any line-up averages about 3.95 runs per game, and all seven outcomes are within two standard deviations of this. We are forced into an even more substantial simulation run.

The simulation was repeated; this time every line-up had seven sets of ten entire seasons simulated. The newly computed averages are shown in column three of Figure 35, while the maximum and minimum values obtained for a set of ten seasons are shown in the last column. Since we have simulated seven times as many games for each line-up, the standard deviation is reduced by a factor of $\sqrt{7}$, to less than .03. The differences in the averages now look more significant. Also we note that the ranges obtained for the first five line-ups don't overlap (or hardly overlap) the ranges for the last two line-ups. We may therefore conclude that we have five "good" and two "poor" line-ups. And this hypothesis stands up under more sophisticated tests.

What characterizes the poor line-ups? Most noticeably, the pitcher is first, rather than being last. But also we note that the Dodgers had three weak hitters (numbers 3, 7, and 8), and two of these are near the top of the bad line-ups. We therefore conclude that poor hitters should be near the end of the line-up. But little else can be concluded.

We should also note that the difference between best and worst is surprisingly little, and drastic changes in the "best" have practically no effect. Thus we conclude that the importance of the batting order has been greatly exaggerated.

One additional remark may be of interest: The first line-up in Figure 35 is, of course, the one chosen by the coach. The last five are simply permutations chosen according to simple patterns. However, the second line-up was chosen by one of the authors, a Dodger fan, as his attempt to "coach" the team. He was most pleased that it turned out best! Of course, .02 is only $\frac{2}{3}$ of a standard deviation, which represents about three runs per year, and is not significant.

EXERCISES

1. Use the RND in Figure 27 to simulate an independent trials process with probability .4 of success, for 30 trials. How many successes do you obtain? [*Ans.* 11.]

2. In Example 1 three games of craps were simulated, using Figure 29. Check these, and then simulate one more game.

3. From Chapter IV, Section 3, Exercises 18, 19 compute the expected number of bust, one pair, two pairs, and three-of-a-kind hands in 10,000 poker hands. Also compute the standard deviation for each. Do the figures given in Example 2 for the simulation look reasonable?

[*Partial Ans.* Bust: expect 5012; off by less than one standard deviation.]

4. Consider an independent trials process with probability p for success. Show that if p is very small then the standard deviation $\sqrt{npq}$ is very close to the square root of the expected number of successes.

5. Use the results of Chapter IV, Section 3, Exercise 11, together with the result of Exercise 4 (above), to check the simulated values for the rarer poker hands in Example 2.

6. Use the results of Exercises 3 and 5 to discuss how far one can rely on the simulated probabilities obtained from 10,000 poker hands.

7. Use the RND in Figure 27 to simulate 30 days' weather in the Land of Oz, following a rainy day. [*Ans.* "Rain" 10, "Nice" 7, "Snow" 13.]

8. Change the RND in Figure 27 to random integers from 1 through 100.

9. Suppose that we have a team each of whose batters performs according to the simulation scheme in Figure 33. Use the random integers obtained in Exercise 8 to simulate the performance of the first 30 batters on one team. How does the team stand after 30 men have come to bat?

[*Ans.* End of six innings, four runs scored.]

10. In 1951, Gil Hodges of the Brooklyn Dodgers was officially at bat 582 times, and hit 40 home runs. Estimate his probability of hitting a home run each time he was at bat. How large a fluctuation in his annual home-run output is attributable to pure chance?

11. From 1949 through 1959, Gil Hodges had the following number of home runs: 23, 32, 40, 32, 31, 42, 27, 32, 27, 22, 25. Is there a case for his having had "good" and "bad" years, or may we assign the differences entirely to chance fluctuations? [*Hint:* Estimate the expected value from the data, and use Exercise 10.] [*Ans.* Explainable as chance fluctuations.]

SUGGESTED READING

Harary, F., R. Z. Norman, and C. C. Cartwright, *Structural Models*, Wiley, New York, 1965.

Gardin, J. C., et P. Garelli, "Étude des établissements assyriens en Cappadoce par ordinateurs," Annales Économies, Sociétés, Civilisations, 16 (1961) pp. 837–876.

Neyman, J., *First Course in Probability and Statistics*, Holt, New York, 1950, Chap. III.

Estes, W. K., and C. J. Burke, "Application of a Statistical Model to Simple Discrimination Learning in Human Subjects," *J. Exptl. Psychol.*, 50 (1955), pp. 81–88.

Bush, R. R., and F. Mosteller, *Stochastic Models for Learning*, Wiley, New York, 1955.

Bush, R. R., and W. K. Estes, *Studies in Mathematical Learning Theory*, Stanford University Press, 1959.

Weil, André, "Sur L'Étude de Certains Types de Lois de Marriage (Système Murngin)," in Appendice à la Première Partie, *Les Structures Elementaire de la Parente*, by Claude Levi-Strauss, Presses Universitaires de France, Paris, 1949, pp. 278–285.

Kemeny, John G., Oskar Morgenstern, and Gerald L. Thompson, "A Generalization of the von Neumann Model of an Expanding Economy," *Econometrica*, 24 (1956), pp. 115–135.

Index

D

E

L

M

N