§ 1.1 Sets. By a set a mathematician means an arbitrary but well-defined collection of objects. Sets will be denoted by bold capital letters. The objects in the collection are called elements.

If \( A \) is a set, and \( B \) is a set whose elements are some (but not necessarily all) of the elements of \( A \), then we say that \( B \) is a subset of \( A \), symbolized as \( B \subseteq A \). If the two sets have exactly the same elements, then we say that they are equal, i.e. \( A = B \). Thus \( A = B \) if and only if \( A \subseteq B \) and \( B \subseteq A \). If \( B \) is a subset of \( A \) and is not equal to \( A \), then we say that it is a proper subset, and write \( B \subset A \). If \( A \) and \( B \) have no element in common, we say that they are disjoint.

Very frequently we will deal with a given set of objects, and discuss various subsets of it. The entire set will be called the universe, \( U \). A particularly interesting subset is the set with no elements, the empty set \( E \).

Given a set, there are a number of ways of getting new subsets from old ones. If \( A \) and \( B \) are both subsets of \( U \), then we define the following operations:

1. The complement of \( A \), \( \tilde{A} \), has as elements all the elements of \( U \) which are not in \( A \).
2. The union of \( A \) and \( B \), \( A \cup B \), has as elements all the elements of \( A \) and all the elements of \( B \).
3. The intersection of \( A \) and \( B \), \( A \cap B \), has as elements all the elements that \( A \) and \( B \) have in common.
4. The difference of \( A \) and \( B \), \( A - B \), has as elements all the elements of \( A \) that are not in \( B \).

To illustrate these operations, we will list some easily provable relations between these sets:

\[
\begin{align*}
\tilde{U} &= E \\
\tilde{A} &= A \\
A - B &= A \cap \tilde{B} \\
A \cup B &= \tilde{A} \cap \tilde{B} \\
A \cap B &= B \cap A \\
A \cup E &= A \\
A \cap E &= E
\end{align*}
\]
1.5.2 Theorem. A probability measure \( m \) assigned to a possibility set \( U \) has the following properties:

1. For any subset \( P \) of \( U \), \( 0 \leq m(P) \leq 1 \).
2. If \( P \) and \( Q \) are disjoint subsets of \( U \), then \( m(P \cup Q) = m(P) + m(Q) \).
3. For any subsets \( P \) and \( Q \) of \( U \), \( m(P \cup Q) = m(P) + m(Q) - m(P \cap Q) \).
4. For any set \( P \) in \( U \), \( m(P^c) = 1 - m(P) \).

1.5.3 Definition. Let \( p \) be a statement relative to a set \( U \) having truth set \( P \). The probability of \( p \) relative to the probability measure \( m \) is defined as \( m(P) \).

In any discussion where there is a fixed probability measure we shall refer simply to the probability of \( p \) without mentioning each time the measure. From Theorem 1.5.2 and the relation of the connectives to the set operations, we have the following theorem:

1.5.4 Theorem. Let \( U \) be a set of possibilities for which a probability measure has been assigned. The probabilities of statements determined by this measure have the following properties:

1. For any statement \( p \), \( 0 \leq \Pr[p] \leq 1 \).
2. If \( p \) and \( q \) are inconsistent then \( \Pr[p \lor q] = \Pr[p] + \Pr[q] \).
3. For any two statements \( p \) and \( q \), \( \Pr[p \lor q] = \Pr[p] + \Pr[q] - \Pr[p \land q] \).
4. For any statement \( p \), \( \Pr[\neg p] = 1 - \Pr[p] \).

1.5.5 Example. Given any finite set having \( s \) elements we can determine a probability measure by assigning weight \( 1/s \) to each element of \( U \). This measure is called the equiprobable measure. For any set \( A \) with \( r \) elements, \( m(A) = r/s \). For example, this is the measure which would normally be assigned to the outcomes for the roll of a die. In this case \( U = \{1, 2, 3, 4, 5, 6\} \) and a weight of \( 1/6 \) is assigned to each.

1.5.6 Example. As an example of a situation where different weights would be assigned consider the following: A man observes a race between three horses \( a, b, \) and \( c \). He feels that \( a \) and \( b \) have the same chance of winning but that \( c \) is twice as likely to win as \( a \). We take the possibility set to be \( U = \{a, b, c\} \) and assign weights \( w(a) = 1/4, w(b) = 1/4 \) and \( w(c) = 1/2 \).

It is occasionally necessary to extend the above concepts to include the case of an experiment with an infinite sequence of possible outcomes. For example, consider the experiment of tossing a coin until the first
time that a head turns up. The possible outcomes would be \( U = \{1, 2, 3, \ldots\} \). The above definitions and theorems apply equally well to this possibility set. We will have an infinite number of weights assigned but we still must require that they have sum 1. In the example just mentioned we would assign weights \( 1/2, 1/4, 1/8, \ldots \). These weights form a geometric progression having sum 1.

§ 1.6 Conditional probability. It often happens that a probability measure has been assigned to a set \( U \) and then we learn that a certain statement \( q \) relative to \( U \) is true. With this new information we change the possibility set to the truth set \( Q \) of \( q \). We wish to determine a probability measure on this new set from our original measure \( m \). We do this by requiring that elements of \( Q \) should have the same relative weights as they had under the original assignment of weights. This means that our new weights must be the old weights multiplied by a constant to give them sum 1. This constant will be the reciprocal of the sum of the weights of all elements in \( Q \), i.e. \( 1/m(Q) \). (See FM Chapter IV or FMS Chapter III.)

1.6.1 Definition. Let \( U = \{a_1, a_2, \ldots, a_r\} \) be a possibility set for which a measure has been assigned, determined by weights \( w(a_j) \). Let \( q \) be a statement relative to \( U \) (not a self-contradiction). The conditional probability measure given \( q \) is a probability measure defined on \( Q \) the truth set of \( q \), determined by weights

\[
\tilde{w}(a_j) = \frac{w(a_j)}{m(Q)}.
\]

1.6.2 Definition. Let \( p \) and \( q \) be two statements relative to a set \( U \) (\( q \) not a self-contradiction). The conditional probability of \( p \) given \( q \), denoted by \( \Pr[p|q] \) is the probability of \( p \) computed from the conditional probability measure given \( q \).

1.6.3 Theorem. Let \( p \) and \( q \) be two statements relative to \( U \) (\( q \) not a self-contradiction). Assume that a probability measure \( m \) has been assigned to \( U \). Then

\[
\Pr[p|q] = \frac{\Pr[p \land q]}{\Pr[q]}
\]

where \( \Pr[p \land q] \) and \( \Pr[q] \) are found from the measure \( m \).

1.6.4 Example. In Example 1.5.6 assume that the man learns that horse \( b \) is not going to run. This causes him to consider the new possibility space \( Q = \{a, c\} \). The new weights which determine the conditional measure are \( \tilde{w}(a) = \frac{1/4}{1/4 + 1/2} = \frac{1}{3} \) and \( \tilde{w}(c) = \frac{1/2}{1/4 + 1/2} = \frac{2}{3} \).
the path $t_7$ corresponds to outcome A on the first stage, T on the second, and 5 on the third. The weight assigned to this path is

$$\frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{6} = \frac{1}{24}.$$ 

This procedure assigns a weight to each path of the tree and the sum of the weights assigned is 1. The set $U$ of all paths may be considered a suitable possibility space for the consideration of any statement whose truth value depends on the outcome of the total experiment. The measure assigned by the path weights is the appropriate probability measure.

![Diagram of a tree with probabilities and outcomes](attachment:image.png)

**Figure 1-1**

The above procedure can be carried out for any experiment that takes place in stages. We require only that there be a finite number of possible outcomes at each stage and that we know the probabilities for any particular outcome at the $j$-th stage, given the knowledge of the outcome for the first $j-1$ stages. For each $j$ we obtain a tree $U_j$. The set of paths of this tree serves as a possibility space for any statement relating to the first $j$ experiments. On this tree we assign a measure to the set of all paths. We first assign branch probabilities. Then the weight assigned to a path is the product of all branch probabilities on the path. The tree measures are consistent in the following sense. A statement whose truth value depends only on the first $j$ stages may be considered a statement relative to any tree $U_i$ for $i \geq j$. 
If \( A_1, A_2, \ldots, A_r \) are subsets of \( U \), and every element of \( U \) is in one and only one set \( A_j \), then we say that \( A = \{ A_1, A_2, \ldots, A_r \} \) is a *partition* of \( U \).

If we wish to specify a set by listing its elements, we write the elements inside curly brackets. Thus, for example, the set of the first five positive integers is \( \{1, 2, 3, 4, 5\} \). The set \( \{1, 3, 5\} \) is a proper subset of it. The set \( \{2\} \), which is also a subset of the five-element set, is called a *unit set*, since it has only one element.

In the course of this book we will have to deal with both finite and infinite sets, i.e. with sets having a finite number or an infinite number of elements. The only infinite sets that are used repeatedly are the set of integers \( \{1, 2, 3, \ldots\} \) and certain simple subsets of this set.

For a more detailed account of the theory of sets see FM Chapter II or FMS Chapter II.†

§ 1.2 *Statements.* We are concerned with a process which will frequently be a scientific experiment or a game of chance. There are a number of different possible outcomes, and we will consider various statements about the outcome.

We form the set \( U \) of all logically possible outcomes. These must be so chosen that we are assured that exactly one of these will take place. The set \( U \) is called the *possibility space*. If \( p \) is any statement about the outcome, then it will (in general) be true according to some possibilities, and false according to others. The set \( P \) of all possibilities which would make \( p \) true is called the *truth set* of \( p \). Thus to each statement about the outcome we assign a subset of \( U \) as a truth set.

The choice of \( U \) for a given experiment is not unique. For example, for two tosses of a coin we may analyze the possibilities as \( U = \{HH, HT, TH, TT\} \) or \( U = \{0H, 1H, 2H\} \). In the first case we give the outcome of each toss and in the second only the number of heads which turn up. (For a more detailed discussion of this concept see FM Chapter II or FMS Chapter II.)

Given two statements \( p \) and \( q \) having the same subject matter (i.e. the same \( U \)), we have a number of ways of forming new statements from them. (We will assume that the statements have \( P \) and \( Q \) as truth sets:)

1. The statement \( \neg p \) (read "not \( p \)"") is true if and only if \( p \) is false. Hence it has \( P \) as truth set.
2. The statement \( p \lor q \) (read "\( p \) or \( q \)"") is true if either \( p \) is true or \( q \) is true or both. Hence it has \( P \cup Q \) as truth set.


the main diagonal”) and 0’s elsewhere is denoted by $I_r$. The subscript is often omitted. The role that these matrices play can be seen as follows. Let $A$, $I$, and $O$ be $r \times r$, let $\alpha$ be an $r$-component row vector, and $\beta$ an $r$-component column vector. Then:

\begin{align*}
A + O &= O + A = A \\
A + (-A) &= (-A) + A = O \\
AI &= IA = A \\
\alpha I &= \alpha \\
I \beta &= \beta \\
AO &= OA = O \\
O \beta &= O \\
\alpha O &= O.
\end{align*}

Thus the matrices $O$ and $I$ play somewhat the same role as the numbers 0 and 1.

(6) In analogy to the reciprocal of a number we define the inverse of a matrix. The $r \times r$ matrix $B$ is said to be the inverse of the $r \times r$ matrix $A$ if $AB = I$. If such an inverse exists, it is denoted by $A^{-1}$. The inverse can be found by solving $r^2$ simultaneous equations. Of course, these equations may fail to have a solution. But when they do have a solution, the solution is unique, and we can show that $AA^{-1} = A^{-1}A = I$.

The various arithmetical operations on matrices, whenever they are defined, obey the usual laws of arithmetic. The one major exception to this is that matrix multiplication is not commutative, i.e. that $AB$ need not equal $BA$. One important case where matrices commute is the case of powers of a given matrix. Let $A^n$ be $A$ multiplied by itself $n$ times. Then $A^n \cdot A^m = A^m \cdot A^n$ for every $n$ and $m$. We define $A^0 = I$.

It is convenient to introduce row vector $\eta_r$ and the column vector $\xi_r$ having all components equal to 1. The subscript is again omitted when possible. These vectors are convenient for summing vectors or rows and columns of matrices. The product $\alpha \xi$ is a number (or more precisely a matrix with a single entry) which is the sum of the components of $\alpha$. Similarly for $\eta \beta$. The product $A \xi$ is a column vector whose $i$-th component gives the sum of the components in the $i$-th row of $A$ (or the $i$-th row sum of $A$). Similarly $\eta A$ gives the column sums of $A$. We shall denote by $E$ a square matrix with all entries 1. Note that $E = \xi \eta$. 
Let us give some examples of these operations and relations.

\[
\begin{pmatrix} 2 & 1 \\ 0 & -1 \end{pmatrix} = \begin{pmatrix} 6 & 3 \\ 0 & -3 \end{pmatrix}
\]

\[
\begin{pmatrix} 2 & 1 \\ 0 & -1 \end{pmatrix} + \begin{pmatrix} -1 & 0 \\ 0 & -2 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 0 & -3 \end{pmatrix}
\]

\[(1, 2, 3) + (2, 1, 0) = (3, 3, 3)\]

\[
(1, 2, 3) \begin{pmatrix} 2 \\ 0 \\ 1 \end{pmatrix} = (5, -1)
\]

\[
\begin{pmatrix} 2 & 1 & 0 \\ 0 & 1 & 2 \end{pmatrix} \begin{pmatrix} 2 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 4 \\ 2 \end{pmatrix}
\]

\[
\begin{pmatrix} 2 & 1 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 3 \\ -1 \end{pmatrix}
\]

\[
\begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} \begin{pmatrix} 1/2 \\ 0 \\ -2 \end{pmatrix}
\]

\[
\begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = I.
\]

Therefore,

\[
\begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix} = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} - 1.
\]

For a square matrix \(A\) we introduce its transpose \(A^T\). The \(ij\)-th entry of \(A^T\) is the \(ji\)-th entry of \(A\). We also define the matrix \(A_{dg}\) which agrees with \(A\) on the main diagonal, but is 0 elsewhere. The matrix \(A_{sq}\) is formed from \(A\) by squaring each entry. This, of course, will not normally be the same as \(A^2\). (But \(D^2 = D_{sq}\) for a diagonal matrix \(D\), i.e. a matrix whose only non-zero entries are on the main diagonal.) Similarly we define \(a_{sq}\) for a vector \(a\).

It is often convenient to give a matrix or a vector in terms of its components. We thus write \(\{a_{ij}\}\) for the matrix whose \(ij\) component
is \(a_{ij}\). Similarly we write \(\{a_i\}\) for a row-vector, and \(\{a_i\}\) for a column vector. The following relations will illustrate this notation.

\[
\{a_{ij}\} + \{b_{ij}\} = \{a_{ij} + b_{ij}\},
\]

\[
O = \{0\},
\]

\[
\xi \eta = E = \{1\},
\]

\[
\{a_{ij}\}_{sq} = \{a^2_{ij}\},
\]

\[
\{a_{ij}\}^T = \{a_{ji}\},
\]

\[
3\{a_i\} = \{3a_i\},
\]

\[
\{a_i\}\{b_j\} = \{a_i b_j\}.
\]

The last example shows that the product of a column vector and a row vector (each with \(r\) components) is a matrix (with \(r \times r\) components). This must be contrasted with the product in the reverse order, which is a single component. For example, if \(\alpha\) is a row vector, then \(\alpha \xi\) gives the sum of its components. However, \(\xi \alpha\) gives an \(r \times r\) matrix with \(\alpha\) for each row.

Suppose that we have a sequence of matrices \(A_k\), with entries \(a^{(k)}_{ij}\). We will say that the series \(A_0 + A_1 + A_2 + \cdots\) converges if each series of entries converges, i.e. if \(a^{(0)}_{ij} + a^{(1)}_{ij} + a^{(2)}_{ij} + \cdots\) converges for every \(i\) and \(j\). And if the sum of this series of components is \(a_{ij}\), for each \(i\) and \(j\), and if \(A\) is the matrix with these entries as components, then we say that \(A\) is the sum of the infinite series of matrices. In brief, we define an infinite sum of matrices by forming the sum for each component.

**1.11.1 Theorem.** If \(A^n\) tends to \(O\) (zero matrix) as \(n\) tends to infinity, then \((I - A)\) has an inverse, and

\[
(I - A)^{-1} = I + A + A^2 + A^3 + \cdots = \sum_{k=0}^{\infty} A^k.
\]

**Proof.** Consider the identity

\[
(I - A) \cdot (I + A + A^2 + \cdots + A^{n-1}) = I - A^n,
\]

which is easily verified by multiplying out the left side. By hypothesis we know that the right side tends to \(I\). This matrix had determinant 1. Hence for sufficiently large \(n\), \(I - A^n\) must have a non-zero determinant. But the determinant of a product of two matrices is the product of the determinants, hence \(I - A\) cannot have a zero determinant. The determinant not being equal to zero is a sufficient condition for a matrix to have an inverse. Hence \(I - A\) has an inverse. Since this inverse exists, we may multiply both sides of the identity by it:

\[
I + A + A^2 + \cdots + A^{n-1} = (I - A)^{-1} \cdot (I - A^n).
\]
Let $N_{ij}$ be the set of $n$ such that a message starting from member $i$ can be in member $j$'s hands at the end of $n$ steps. We will first consider $N_{ii}$, the possible times at which a message can return to its originator. It is clear that if $a \in N_{ii}$ and $b \in N_{ii}$, then $a + b \in N_{ii}$ after all the message can return in $a$ steps and can be sent out again and be received back after $b$ more steps. So the set $N_{ii}$ is closed under addition. The following number-theoretic result will be useful. Its proof is given at the end of the section.

1.4.1 Theorem. A set of positive integers that is closed under addition contains all but a finite number of multiples of its greatest common divisor.

If the greatest common divisor of the elements of $N_{ii}$ is designated $d_i$, it is clear that the elements of $N_{ii}$ are all multiples of $d_i$. But Theorem 1.4.1 tells us in addition that all sufficiently high multiples of $d_i$ are in the set.

Since each member can contact every other member in its equivalence class, the $N_{ij}$ are non-empty. We next prove that for $i$ and $j$ in the same equivalence class, $d_i = d_j = d$, and that the elements of a given $N_{ij}$ are congruent to each other modulo $d$ (their difference is a multiple of $d$). Suppose that $a \in N_{ij}$, $b \in N_{ij}$, and $c \in N_{ji}$.

First of all, member $i$ can contact himself by sending a message to member $j$ and getting a message back. Hence $a + c \in N_{ii}$. The message could also go to member $j$, come back to member $j$, and then go to member $i$. This could be done in $a + kd + c$ steps, where $k$ is sufficiently large. Hence $d_j$ must be a multiple of $d_i$. But in exactly the same way we can prove that $d_i$ is a multiple of $d_j$. Hence $d_i = d_j = d$.

Or again, the message could go to member $j$ in $b$ steps, and then back to member $i$. Hence $b + c \in N_{ii}$. Hence $a + c$ and $b + c$ are both divisible by $d$, and thus we see that $a \equiv b \ (\text{mod } d)$. Thus the elements of a given $N_{ij}$ are congruent to each other modulo $d$. We can thus introduce numbers $t_{ij}$, with $0 \leq t_{ij} < d$, so that any element of $N_{ij}$ is congruent to $t_{ij}$, modulo $d$. It is also easy to see that $N_{ij}$ contains all but a finite number of the numbers $t_{ij} + kd$.

In particular we see that $t_{ii} = 0$ in each case, and hence $t_{ij} + t_{ji} \equiv 0 \ (\text{mod } d)$. Also $t_{ij} + t_{jm} \equiv t_{im} \ (\text{mod } d)$. From this it is easily seen that $t_{ij} = 0$ is an equivalence relation. Let us call such an equivalence class a cyclic class.

Since $t_{ij} + t_{jm} \equiv t_{im} \ (\text{mod } d)$, we see that $t_{ij} = t_{im}$ if and only if $t_{jm} = 0$, hence if and only if members $j$ and $m$ are in the same cyclic class. Let $n$ be any integer. If $n \equiv t_{ij} \ (\text{mod } d)$, then the message originating from member $i$ can only be in this one cyclic class after $n$ steps. From
this it immediately follows that there are exactly $d$ cyclic classes, and that the message moves cyclically from class to class, with cycle of length $d$. It is also easily seen that after sufficient time has elapsed, it can be in the hands of any member of the one cyclic class appropriate for $n$.

While this description of an equivalence class of the communication network holds in complete generality, the cycle degenerates when $d = 1$. In this case there is a single "cyclic class," and after sufficient time has elapsed the message can be in the hands of any member at any time.

In particular, it is worth noting that if any member of the equivalence class can contact himself directly, then $d = 1$. This is immediately seen from the fact that $d$ is a divisor of any time in which a member can contact himself, and here $d$ has to divide 1.

The number-theoretic result, §1.4.1, is of such interest that its proof will be given here.

First of all we note that if the greatest common divisor $d$ of the set is not 1, then we can divide all elements by $d$, and reduce the problem to the case $d = 1$. Hence it suffices to treat this case. Here we have a set of numbers whose greatest common divisor is 1, and we must have a finite subset with this property. Hence, by a well-known result, there is a linear combination, $a_1n_1 + a_2n_2 + \cdots + a_kn_k$ of the elements (with positive or negative integers $a_i$) which is equal to 1. If we collect all the positive and all the negative terms separately, and remember that the set is closed under addition, we note that there must be elements $m$ and $n$ in the set, such that $m - n = 1$ ($m$ being the sum of the positive terms, and $-n$ the sum of the negative terms).

Let $q$ be any sufficiently large number, or more precisely $q \geq n(n-1)$. We can write $q = an + b$, where $a \geq (n-1)$ and $0 \leq b \leq (n-1)$. Then we see that $q = (a-b)n + bm$, and hence $q$ must be in the set.

§ 1.5 Probability measures. In making a probability analysis of an experiment there are two basic steps. First, a set of logical possibilities is chosen. This problem was discussed in §1.2. Second a probability measure is assigned. The way that this second step is carried out will be discussed in this section. We consider first a finite possibility space. (For a more detailed discussion see FM Chapter IV or FMS Chapter III.)

1.5.1 Definition. Let $U = \{a_1, a_2, \ldots, a_r\}$ be a set of logical possibilities. A probability measure for $U$ is obtained by assigning to each element $a_j$ a positive number $w(a_j)$, called a weight, in such a way that the weights assigned have sum 1. The measure of a subset $A$ of $U$, denoted by $m(A)$, is the sum of the weights assigned to elements of $A$. 
1.7.3 Definition. Let $U$ be a possibility space, and $f$ and $g$ be two functions with domain $U$, each having as range a set of numbers. The function $f + g$ is the function with domain $U$ which assigns to $a_j$ the number $f(a_j) + g(a_j)$. The function $f \cdot g$ is the function with domain $U$ which assigns to $a_j$ the number $f(a_j) \cdot g(a_j)$. For any number $c$ the constant function $c$ is the function which assigns the number $c$ to every element of $U$.

Let $U$ be a possibility space for which a measure has been assigned. Then if $f$ and $g$ are two numerical functions with domain $U$, $f + g$ and $f \cdot g$ will be functions with domain $U$, and as such have induced measures. In general there is no simple connection between the induced measures of these functions and the induced measure for $f$ and $g$.

1.7.4 Example. In Example 1.6.6 let $g$ be a function having the value 1 if a head turns up on the first toss and 0 otherwise. Let $h$ be a function having the value 1 if a head turns up on the second toss and 0 if a tail turns up. Then the range and induced measures for $g$, $h$, $g + h$, and $g \cdot h$ are

\[
g: \begin{bmatrix} 0 & 1 \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}
\]

\[
h: \begin{bmatrix} 0 & 1 \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}
\]

\[
g + h: \begin{bmatrix} 0 & 1 & 2 \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \end{bmatrix}
\]

\[
g \cdot h: \begin{bmatrix} 0 & 1 \\ \frac{3}{4} & \frac{1}{4} \end{bmatrix}.
\]

1.7.5 Definition. Let $f$ be a function defined on $U$. Let $p$ be a statement relative to $U$ having truth set $P$. Assume that a measure $m$ has been assigned to $U$. Let $f'$ be the function $f$ considered only on the set $P$. Then the induced measure for $f'$ calculated from the conditional measure given $p$ is called the conditional induced measure for $f$ given $p$.

1.7.6 Definition. Let $f$ and $g$ be two functions defined on a space $U$ for which a probability measure has been assigned. Then $f$ and $g$ are independent if, for any $r_k$ in the range of $f$ and $s_j$ in the range of $g$, the statements $f = r_k$ and $g = s_j$ are independent statements.

An equivalent way to state the condition for independence of two
We observe that it is still twice as likely that \( c \) will win than it is that \( a \) will win.

1.6.5 Definition. Two statements \( p \) and \( q \) (neither of which is a self-contradiction) are independent if \( \Pr[p \land q] = \Pr[p] \cdot \Pr[q] \).

It follows from Theorem 1.6.3 that \( p \) and \( q \) are independent if and only if \( \Pr[p|q] = \Pr[p] \) and \( \Pr[q|p] = \Pr[q] \). Thus to say that \( p \) and \( q \) are independent is to say that the knowledge that one is true does not affect the probability assigned to the other.

1.6.6 Example. Consider two tosses of a coin. We describe the outcomes by \( U = \{ HH, HT, TH, TT \} \). We assign the equiprobable measure. Let \( p \) be the statement “a head turns up on the first toss” and \( q \) the statement “a head turns up on the second toss.” Then \( \Pr[p \land q] = \frac{1}{4}, \Pr[p] = \Pr[q] = \frac{1}{2} \). Thus \( p \) and \( q \) are independent.

§ 1.7 Functions on a possibility space. Let \( U = \{ a_1, a_2, \ldots, a_r \} \) be a possibility space. Let \( f \) be a function with domain \( U \) and range \( R = \{ r_1, r_2, \ldots, r_s \} \). That is, \( f \) assigns to each element \( U \) a unique element of \( R \). If \( f \) assigns \( r_k \) to \( a_j \), we write \( f(a_j) = r_k \). We write \( f = r_k \) for the statement “the value of the function is \( r_k \)” This is a statement relative to \( U \), since its truth value is known when the outcome \( a_j \) is known. Hence it has a truth set which is a subset of \( U \). (See FMS Chapters II, III, or M4 Vol. II, Unit 1.)

1.7.1 Definition. Let \( f \) be a function with domain \( U \) and range \( R \). Assume that a measure has been assigned to \( U \). For each \( r_k \) in \( R \) let \( w(r_k) = \Pr[f = r_k] \). The weights \( w(r_k) \) determine a probability measure on the set \( R \), called the induced measure for \( f \). The weights are called the induced weights.

We shall normally indicate the induced measure by giving both the range values and the weights in the form:

\[
\begin{align*}
\text{f:} & \quad \{ r_1, r_2, \ldots, r_s \} \\
& \quad \{ w(r_1), w(r_2), \ldots, w(r_s) \}
\end{align*}
\]

Thus the induced weight of \( r_k \) in \( R \) is the measure of the truth set of \( f = r_k \) in \( U \).

1.7.2 Example. In Example 1.6.6 let \( f \) be the function which gives the number of heads which turn up. The range of \( f \) is \( R = \{ 0, 1, 2 \} \). The \( \Pr[f = 0] = \frac{1}{4}, \Pr[f = 1] = \frac{1}{2}, \) and \( \Pr[f = 2] = \frac{1}{4} \). Hence the range and induced measure is:

\[
\begin{align*}
\text{f:} & \quad \begin{bmatrix} 0 & 1 & 2 \\ 1/4 & 1/2 & 1/4 \end{bmatrix}
\end{align*}
\]
functions is to say that the induced measure for one function is not changed by the knowledge of the value of the other.

§ 1.8 Mean and variance of a function. Throughout this section we shall assume that the functions considered are functions whose range set is a set of numbers. (A detailed discussion of the concepts introduced in this section is given in FMS Chapter III, or M⁴ Vol. II, Unit 1.)

1.8.1 Definition. Let $f$ be a function defined on a possibility space $U = \{a_1, a_2, \ldots, a_r\}$, for which a measure determined by weights $w(a_j)$ has been assigned. Then the mean value of $f$ denoted by $M[f]$ is

$$M[f] = \sum_j f(a_j) \cdot w(a_j).$$

The term expected value is often used in place of mean value.

1.8.2 Theorem. Let $f$ be a function defined on $U$. Assume that for a probability measure $m$ defined on $U$, the function $f$ has induced measure

$$f: \left\{ \begin{array}{c} r_1, \\ r_2, \\ \ldots, \\ r_s \\ \end{array} \right\} \mapsto \left\{ \begin{array}{c} w(r_1), \\ w(r_2), \\ \ldots, \\ w(r_s) \end{array} \right\}.$$

Then

$$M[f] = \sum_j r_j \cdot w(r_j).$$

1.8.3 Example. In Example 1.6.6 let $f$ be the number of heads which turn up. From the definition of mean value we have

$$M[f] = f(HH) \cdot \frac{1}{4} + f(HT) \cdot \frac{1}{4} + f(TH) \cdot \frac{1}{4} + f(TT) \cdot \frac{1}{4}$$

$$= 2 \cdot \frac{1}{4} + 1 \cdot \frac{1}{4} + 1 \cdot \frac{1}{4} + 0 \cdot \frac{1}{4}$$

$$= 1.$$

We can also calculate the mean of $f$ by making use of Theorem 1.8.2. The range and induced measure for $f$ is

$$f: \left\{ \begin{array}{c} 0, 1, 2 \\ \end{array} \right\} \mapsto \left\{ \begin{array}{c} 0.1/4, 1.1/2, 2.1/4 \\ \end{array} \right\}.$$

Thus by Theorem 1.8.2,

$$M[f] = 0.1/4 + 1.1/2 + 2.1/4 = 1.$$

1.8.4 Definition. Let $f$ be a function defined on a possibility space $U$ for which a measure has been assigned. Let $M[f] = m$ be the mean of this function. Then the variance of $f$, denoted by $\text{Var}[f]$, is the mean of the function $(f - m)^2$. The standard deviation denoted by $\text{sd}[f]$, is the square root of the variance.
1.8.5 **Theorem.** Let \( f \) be a function having mean value \( m \). Then
\[
\]

1.8.6 **Example.** Let \( f \) be the function in Example 1.8.3. We found that \( M[f] = 1 \). Thus
\[
\text{Var}[f] = (2 - 1)^2 \cdot \frac{1}{4} + (1 - 1)^2 \cdot \frac{1}{4} + (1 - 1)^2 \cdot \frac{1}{4} + (0 - 1)^2 \cdot \frac{1}{4} = \frac{1}{2}.
\]

An alternative way to compute the variance is to make use of Theorem 1.8.5. Using this result we find
\[
M[f^2] = 4 \cdot \frac{1}{4} + 1 \cdot \frac{1}{4} + 1 \cdot \frac{1}{4} + 0 \cdot \frac{1}{4} = \frac{3}{2}.
\]
Since \( M[f] = 1 \), we have \( \text{Var}[f] = \frac{3}{2} - 1 = \frac{1}{2} \).

1.8.7 **Theorem.** Let \( f \) and \( g \) be any two functions for which means and variances have been defined. Then
\begin{align*}
(1) &\quad M[c] = c. & (4) &\quad \text{Var}[c \cdot f] = c^2 \cdot \text{Var}[f]. \\
(2) &\quad M[f + g] = M[f] + M[g]. & (5) &\quad \text{Var}[c + f] = \text{Var}[f]. \\
(3) &\quad M[c \cdot f] = c \cdot M[f]. & (6) &\quad \text{Var}[c] = 0.
\end{align*}

If \( f \) and \( g \) are independent functions then
\begin{align*}
(7) &\quad M[f \cdot g] = M[f] \cdot M[g]. \\
(8) &\quad \text{Var}[f + g] = \text{Var}[f] + \text{Var}[g].
\end{align*}

1.8.8 **Definition.** Let \( p \) be a statement relative to a possibility set \( U \) for which a measure has been assigned. Let \( f \) be a function with domain \( U \). The **conditional mean** and **variance** of \( f \) given \( p \) are the mean and variance of \( f \) computed from the conditional measure given \( p \). We denote these by \( M[f|p] \) and \( \text{Var}[f|p] \).

1.8.9 **Theorem.** Let \( p_1, p_2, \ldots, p_r \) be a complete set of alternatives relative to a set \( U \). Let \( f \) be a function with domain \( U \). Then
\[
M[f] = \sum_j M[f|p_j] \cdot \Pr[p_j].
\]

1.8.10 **Theorem.** If \( f_1, f_2, \ldots \) is a sequence of functions such that for some constant \( c \),
\[
M[(f_n - c)^2] \to 0
\]
as \( n \to \infty \), then
\[
M[f_n] \to c
\]
and for any \( \varepsilon > 0 \)
\[
\Pr[|f_n - c| > \varepsilon] \to 0
\]
as \( n \to \infty \).
Each of these trees has its own tree measure and the probability of the statement could be found from any one of these measures. However, in every case the same probability would be assigned.

Assume that we have a tree for an \( n \) stage experiment. Let \( f_j \) be a function with domain the set of paths \( U_n \) and value the outcome at the \( j \)-th stage. Then the functions \( f_1, f_2, \ldots, f_n \) are called *outcome functions*. The set of functions \( f_1, f_2, \ldots, f_n \) is called a *stochastic process*. (In Markov chain theory it is convenient to denote the first outcome by \( f_0 \) instead of \( f_1 \).)

In our example there are three outcome functions. We have indicated in Figure 1-1 the value of each function on each path.

There is a simple connection between the branch probabilities and the outcome functions. The branch probabilities at the first stage are,

\[
\Pr[f_1 = r_i]
\]

at the second stage

\[
\Pr[f_2 = r_j | f_1 = r_i]
\]

at the third stage

\[
\Pr[f_3 = r_k | f_2 = r_j \land f_1 = r_i]
\]

e tc.

In our example,

\[
\begin{align*}
\Pr[f_1 = A] &= w(t_1) + \cdots + w(t_8) = \frac{1}{2} \\
\Pr[f_2 = T | f_1 = A] &= \frac{\Pr[f_2 = T \land f_1 = A]}{\Pr[f_1 = A]} \\
&= \frac{w(t_3) + \cdots + w(t_8)}{w(t_1) + \cdots + w(t_8)} = \frac{1/4}{1/2} = \frac{1}{2} \\
\Pr[f_3 = 1 | f_2 = T \land f_1 = A] &= \frac{\Pr[f_3 = 1 \land f_2 = T \land f_1 = A]}{\Pr[f_2 = T \land f_1 = A]} \\
&= \frac{w(t_3)}{w(t_3) + \cdots + w(t_8)} = \frac{1/24}{1/4} = \frac{1}{6}.
\end{align*}
\]

**1.9.2 Example.** We shall often deal with experiments where we allow an arbitrary number of stages. For example, in considering the tosses of a coin, we can envision any number of tosses. The tree for three tosses and the path measure is shown in Fig. 1.2.

For any number of tosses we can construct a tree. It is even possible to consider continuing the tree indefinitely to obtain a tree with infinite paths. Our procedure for assigning a measure would not in this case be adequate since it would assign weight 0 to every path. We shall not, however, have to assign a measure to the infinite tree. This is the case because the statements about the process that interest us will depend only on a finite part of the tree, and for any finite number
of stages we have a method of assigning a measure. We shall, however, consider functions whose definition requires the infinite tree.

For example, in Example 1.9.2 let the value of \( f \) be the stage at which the first head occurs. Then \( f \) is defined for all paths with at least one head. This is a subset of paths in the infinite tree. We shall speak of the mean value of such a function when the following conditions are satisfied:

(a) There is a sequence of numerical range values \( r_1, r_2, \ldots \) such that the truth value of the statement \( f = r_j \) depends only on the outcomes of a finite number of stages and \( \sum_j \Pr[f = r_j] = 1 \).

(b) \( \sum_j r_j \Pr[f = r_j] < \infty \).

In case (a) and (b) hold, we say that \( f \) has a mean value given by

\[ M[f] = \sum_j r_j \Pr[f = r_j] \]

When \( f \) has a mean \( a \), we shall say that \( f \) has a variance if \( (f - a)^2 \) has a mean. If so, \( \text{Var}[f] = M[(f - a)^2] \).

All properties of means and variances given in § 8 hold for these
But the right side of this new identity clearly tends to \((I - A)^{-1}\), which completes the proof.

One can define the summability of matrix sequences and series exactly as in § 1.10, applying the averaging method to each component of the matrix. Then there is a generalization of the previous theorem: If the sequence \(A^n\) is summable to \(O\) by some averaging method, then the matrix \(I - A\) has an inverse, and the series \(I + A + A^2 + \cdots\) is summable by the same method to \((I - A)^{-1}\).

1.11.2 Definition. A square matrix \(A\) is positive semi-definite if for any column vector \(\gamma\), \(\gamma^T A \gamma \geq 0\).

1.11.3 Theorem. For any positive semi-definite matrix \(A\) there is a matrix \(B\) such that \(A = B^T B\).
(3) The statement $p \land q$ (read "p and q") is true if both $p$ and $q$ are true. Hence it has $P \cap Q$ as truth set.

Two special kinds of statements are among the principal concerns of logic. A statement that is true for each logically possible outcome, that is, a statement having $U$ as its truth set, is said to be logically true (such a statement is sometimes called a tautology). A statement that is false for each logically possible outcome, that is a statement having $E$ as its truth set, is logically false or self-contradictory.

Two statements are said to be equivalent if they have the same truth set. That means that one is true if and only if the other is true.

The statements $p_1, p_2, \ldots, p_k$ are inconsistent if the intersection of their truth sets is empty, i.e., $P_1 \cap P_2 \cap \cdots \cap P_k = E$. Otherwise they are said to be consistent. If the statements are inconsistent, then they cannot all be true. If they are consistent, then they could all be true.

The statements $p_1, p_2, \ldots, p_k$ are said to form a complete set of alternatives if for every element of $U$ exactly one of them is true. This means that the intersection of any two truth sets is empty, and the union of all the truth sets is $U$. Thus the truth sets of a complete set of alternatives form a partition of $U$. A complete set of alternatives provides a new way (and normally a less detailed way) of analyzing the possible outcomes.

§ 1.3 Order relations. We will need some simple ideas from the theory of order relations. A complete treatment of this theory will be found in M4, Vol. II, Unit 2.† We will take only a few concepts from that treatment.

Let $R$ be a relation between two objects (selected from a specified set $U$). We denote by $aRb$ the fact that $a$ holds the relation $R$ to $b$. Some special properties of such relations are of interest to us.

1.3.1 Definition. The relation $R$ is reflexive if $xRx$ holds for all $x$ in $U$.

1.3.2 Definition. The relation $R$ is symmetric if whenever $xRy$ holds, then $yRx$ also holds, for all $x, y$ in $U$.

1.3.3 Definition. The relation $R$ is transitive if whenever $xRy \land yRz$ holds, then $xRz$ also holds, for all $x, y, z$ in $U$.

1.3.4 Definition. A relation that is reflexive, symmetric, and transitive is an equivalence relation.

The fundamental property of an equivalence relation is that it partitions the set $U$. More specifically, let us suppose that $R$ is an

† M4 = Modern Mathematical Methods and Models, by the Dartmouth Writing Group. Mathematical Association of America, 1958.
equivalence relation defined on \( U \). We put elements of \( U \) into classes in such a manner that two elements \( a \) and \( b \) are in the same class if \( a \mathrel{\text{R}} b \). It can be shown that the resulting classes are well defined and mutually exclusive, giving us a partition of \( U \). These classes are the equivalence classes of \( R \).

For example, let \( x \mathrel{\text{R}} y \) express that “\( x \) is the same height as \( y \),” where \( U \) is a set of human beings. Then the resulting partition divides these people according to their heights. Two men are in the same equivalence class if and only if they are the same height.

1.3.5 Definition. A relation \( T \) is said to be consistent with the equivalence relation \( R \) if, given that \( x \mathrel{\text{R}} y \), then if \( x \mathrel{\text{T}} z \) holds so does \( y \mathrel{\text{T}} z \), and if \( z \mathrel{\text{T}} x \) holds so does \( z \mathrel{\text{T}} y \).

1.3.6 Definition. A relation that is reflexive and transitive is known as a weak ordering relation.

A weak ordering relation can be used to order the elements of \( U \). Given a weak ordering \( T \), and given any two elements \( a \) and \( b \) of \( U \), there are four possibilities: (1) \( a \mathrel{\text{T}} b \land b \mathrel{\text{T}} a \); then the two elements are “alike” according to \( T \). (2) \( a \mathrel{\text{T}} b \land \neg (b \mathrel{\text{T}} a) \); then \( a \) is “ahead” of \( b \). (3) \( \neg (a \mathrel{\text{T}} b) \land b \mathrel{\text{T}} a \); then \( b \) is “ahead.” (4) \( \neg (a \mathrel{\text{T}} b) \land \neg (b \mathrel{\text{T}} a) \); then we are unable to compare the two objects.

For example, if \( x \mathrel{\text{T}} y \) expresses that “I like \( x \) at least as well as \( y \),” then the four cases correspond to “I like them equally,” “I prefer \( x \),” “I prefer \( y \),” and “I cannot choose,” respectively.

The relation of being alike acts as an equivalence relation. Indeed, it can be shown that if \( T \) is a weak ordering, then the relation \( x \mathrel{\text{R}} y \) that expresses that \( x \mathrel{\text{T}} y \land y \mathrel{\text{T}} x \) is an equivalence relation consistent with \( T \). Thus \( T \) serves both to classify and to order. Consistency assures us that equivalent elements of \( U \) have the same place in the ordering.

For example, if we choose “is at least as tall” as our weak ordering, this determines the equivalence relation “is the same height,” which is consistent with the original relation.

1.3.7 Definition. If \( T \) is a weak ordering, then the relation \( x \mathrel{\text{T}} y \land y \mathrel{\text{T}} x \) is the equivalence relation determined by it.

1.3.8 Definition. If \( T \) is a weak ordering, and the equivalence relation determined by it is the identity relation \( (x = y) \) then \( T \) is a partial ordering.

The significance of a partial ordering is that no two distinct elements are alike according to it. One simple way of getting a partial ordering is as follows: Let \( T \) be a weak ordering defined on \( U \). Define a new relation \( T^* \) on the set of equivalence classes by saying that \( U \mathrel{\text{T}} V \) holds if every element of \( U \) bears the relation \( T \) to every element of \( V \).
This is a partial ordering of the equivalence classes, and we call it the partial ordering induced by $T$.

1.3.9 Definition. An element $a$ of $U$ is called a minimal element if $aTx$ implies $xTa$ for all $x \in U$. If a minimal element is unique, we call it a minimum.

We can define "maximal element" and "maximum" similarly. If $U$ is a finite set, then it is easily shown that for any weak ordering there must be at least one minimal element. However, this minimal element need not be unique. Similarly, the weak ordering must have a maximal element, but not necessarily a maximum.

§ 1.4 Communication relations. An important application of order relations is the study of communication networks. Let us suppose that $r$ individuals are connected through a complex network. Each individual can pass a message on to a subset of the individuals. This we will call direct contact. These messages may be relayed, and relayed again, etc. This will be indirect contact. It will not be assumed that a member can contact himself directly. Let $aTb$ express that the individual $a$ can contact $b$ (directly or indirectly) or that $a=b$. It is easy to verify that $T$ is a weak ordering of the set of individuals. It determines the equivalence relation $xTy \land yTx$, which may be read as "$x$ and $y$ can communicate with each other, or $x=y$".

This equivalence relation may be used to classify the individuals. Two men will be in the same equivalence class if they can communicate, that is, if each can contact the other one. The induced partial ordering $T^*$ has a very intuitive meaning: The relation $uT^*v$ holds if all members of the class $u$ can contact all members of the class $v$, but not conversely unless $u=v$. Thus the partial ordering shows us the possible flow of information.

In particular, $u$ is a maximal element of the partial ordering if its members cannot be contacted by members of any other class, and $u$ is a minimal element if its members cannot contact members of other classes. Thus the maximal sets are message initiators, while the minimal sets are terminals for messages. (See M4 Vol. II, Unit 2.)

It is interesting to study a given equivalence class. Any two members of such a class can communicate with each other. Hence any member can contact any other member. But how long does it take to contact other members? As a unit of time we will take the time needed to send a message from any one member to any member he can contact directly. We call this one step. We will assume that member $i$ sends out a message, and we will be interested to know where the message could possibly be after $n$ steps.
extended mean values. In addition we shall need the following theorem.

1.9.3 Theorem. Let \( f_1, f_2, \ldots \) be functions such that the range of each \( f_j \) is a subset of the same finite set of numbers. Let \( s = f_1 + f_2 + \cdots \). Then if the mean of \( s \) exists,

\[
\mathbb{E}[s] = \sum_{j} \mathbb{E}[f_j].
\]

A stochastic process for which the outcome functions all have ranges which are subsets of a given finite set is called a finite stochastic process. Thus Theorem 1.9.3 states that in a finite stochastic process the mean of the sum of the functions (if this mean exists) is the sum of the means of the functions.

§ 1.10 Summability of sequences and series. It may occur that for a divergent sequence \( s_0, s_1, s_2, \ldots \) we can form a sequence of averages of the terms, and that this new sequence converges. In this case we say that the original sequence is summable by means of the averaging process. We will be concerned with only two methods of averaging.

Let \( t_n = (1/n) \sum_{i=0}^{n-1} s_i \) and let \( u_n = \sum_{i=0}^{n} \binom{n}{i} k^{n-i}(1-k)^i s_i \) for some \( k \) such that \( 0 < k < 1 \). Each of these is an average of terms of the sequence, with non-negative coefficients whose sum is 1. If the sequence \( t_1, t_2, \ldots \) converges to a limit \( t \), then we say that the original sequence is Cesaro-summable to \( t \). If the sequence \( u_1, u_2, \ldots \) converges to \( u \), then we say that the original sequence is Euler-summable to \( u \).

For example, consider the sequence 1, 0, 1, 0, 1, 0, \ldots. We find that \( t_n = \frac{1}{2} \) if \( n \) is even, \( \frac{1}{2} + \frac{1}{n} \) if \( n > 1 \) is odd. This sequence converges to \( \frac{1}{2} \), and hence the original sequence is Cesaro-summable to \( \frac{1}{2} \). It is easy to verify that \( \lim_{n \to \infty} u_n = \frac{1}{2} \) and hence the original sequence is also Euler-summable to \( \frac{1}{2} \). But the original sequence diverges.

We will need the following two simple facts concerning summability: (1) If a sequence converges, then it is summable by each method to its limit. (2) If a sequence is summable by both methods, the two sums must be the same.

Summability may also be applied to a series. To say that the series \( \sum_{k=0}^{\infty} a_k \) is summable by a given method means that its sequence of
partial sums \( s_i = \sum_{k=0}^{i} a_k \) is summable by that method. For example if we apply Cesaro-summability to the partial sums, we obtain
\[
t_n = \sum_{k=0}^{n-1} \frac{n-k}{n} a_k.
\]

§ 1.11 Matrices. A matrix is a rectangular array of numbers. An \( r \times s \) matrix has \( r \) rows and \( s \) columns, a total of \( rs \) entries (or components). Three special kinds of matrices will be especially important in this book. A matrix having the same number of rows as columns is called a square matrix. That is, a square matrix is \( r \times r \). If \( r = 1 \), that is, the matrix consists of a single row, then we call it a row vector. If \( s = 1 \), i.e. the matrix has a single column, we call it a column vector. Matrices will be denoted by capitals and vectors by small Greek letters.

Let the \( r \times s \) matrix \( A \) have components \( a_{ij} \), and the \( r' \times s' \) matrix \( B \) have components \( b_{ij} \). Then we define the following operations and relations:

1. The matrix \( kA \) has components \( ka_{ij} \). That is, a multiplication of the matrix by a number means multiplying each component by this number. The matrix \(-A\) is \((-1)A\).
2. If \( r = r' \) and \( s = s' \), then the matrix sum \( A + B \) has components \( a_{ij} + b_{ij} \). That is, addition is carried out componentwise.
3. If \( s = r' \), we define the product \( AB \) to have components \( \sum_{k=1}^{g} a_{ik}b_{kj} \).

Note that the product of an \( r \times s \) and \( s \times t \) matrix is an \( r \times t \) matrix. This definition also applies to the product of a row vector and a matrix, \( \alpha A \), or to a matrix times a column vector, \( A\beta \). In the former case the product of a \( 1 \times r \) and an \( r \times s \) matrix is a \( 1 \times s \) matrix, or a row vector. If the matrix \( A \) is square, the resulting row vector has the same number of components as \( \alpha \). Thus a square matrix may be thought of as a transformation of row vectors. Similarly we can think of it as a transformation of column vectors. This will be our principal use of the product of a vector and a matrix.

4. We say that \( A \geq B \) (or that \( A = B \)) if \( a_{ij} \geq b_{ij} \) (or \( a_{ij} = b_{ij} \)) for all \( i \) and \( j \). That is, matrix relations must hold componentwise—for all corresponding components.

5. Some special matrices play an important role. The \( r \times s \) matrix having all components equal to 0 is denoted by \( O_{rs} \). The subscripts are omitted whenever there is no danger of confusion. The \( r \times r \) matrix having 1's as components \( a_{ii} \) ("on
1.8.11 Definition. Let \(f_1\) and \(f_2\) be two functions with \(M[f_i] = a_i\) and \(\text{sd}[f_i] = b_i\). Then the covariance of \(f_1\) and \(f_2\) is defined by

\[
\text{Cov}[f_1, f_2] = M[(f_1 - a_1)(f_2 - a_2)],
\]

and the correlation of \(f_1\) and \(f_2\) is

\[
\text{Corr}[f_1, f_2] = \frac{\text{Cov}[f_1, f_2]}{b_1 \cdot b_2},
\]

§ 1.9 Stochastic processes. In this section we shall briefly describe the concept of a stochastic process. A more complete treatment may be found in FM Chapter IV or FMS Chapter III.

We wish to give a probability measure to describe an experiment which takes place in stages. The outcome at the \(n\)-th stage is allowed to depend on the outcomes of the previous stages. It is assumed, however, that the probability for each possible outcome at a particular stage is known when the outcomes of all previous stages are known. From this knowledge we shall construct a possibility space and measure for the over-all experiment.

We shall illustrate the construction of the possibility space and measure by a particular example. The general procedure will be clear from this.

1.9.1 Example. We choose at random one of two coins \(A\) or \(B\). Coin \(A\) is a fair coin and coin \(B\) has heads on both sides. The coin chosen is tossed. If a tail comes up a die is rolled. If a head turns up the coin is thrown again. The first stage of the experiment is the choice of a coin. At the second stage, a coin is tossed. At the third stage a coin is tossed or a die is rolled, depending on the outcome of the first two stages.

We indicate the possible outcomes of the experiment by a tree as shown in Figure 1-1.

The possibilities for the experiment are \(t_1 = (A, H, H)\), \(t_2 = (A, H, T)\), \(t_3 = (A, T, 1)\), \(t_4 = (A, T, 2)\), etc. Each possibility may be identified with a path through the trees. Each path is made up of line segments called branches. In the tree we have just given, there are nine paths each having three branches.

We know the probability for each outcome at a given stage when the previous stages are known. For example, if outcome \(A\) occurs on the first stage and \(T\) on the second stage, then the probability of a \(1\) for the third stage is \(1/6\). We assign these known probabilities to the branches and call them branch probabilities.

We next assign weights to the paths equal to the product of the probabilities assigned to the components of the path. For example