## Chapter 1

## PROBABILITY

George P. Wadsworth
Massachusetts Institute of Technology

## Introduction

A common type of phenomenon observed in many areas of human experience is characterized by the production of a series of events which are rather definitely fixed as to their possible distinct forms, but haphazard as to the order in which one form follows upon another; or, alternatively, by the generation of a quantitative variable having a fairly well-defined range of possible values, but no consistent rule of progression. Despite the absence of any discernible law of serial order, there usually exists a striking similarity between the proportionate compositions of any pair of long records of the same phenomenon. The observed regularity of proportion in random series is the empirical basis for the so-called laws of chance, which are the subject matter of the theory of probability.

Controversies over the philosophical foundation of the theory of probability have been going on for many years. Points of view have ranged from the consideration of probability as a mathematical property of an abstract system to the definition of it as a measure of credibility or as an aspect of human psychology. Notwithstanding this diversity of thought as to the philosophical foundation there has been almost universal agreement as to the mathematical superstructure. The essential mathematical properties which probability must possess (and about which there is no argument) suffice to establish the theory of probability on an axiomatic basis as a rigorous branch of pure mathematics. Like the axioms of Euclidean geometry, those of probability represent idealizations of practical experience or extensions of elementary principles of logic.

## The Relative Frequency Theory

The Relative Frequency Theory of probability is founded upon an observational concept. A record is kept of the number of times $\left(\right.$ say $n_{1}$ ) that a certain event E occurs in $n$ trials of an appropriate experiment. The ratio $n_{1} / n$ is called the relative frequency of the event $E$, and the complementary ratio $\left(n-n_{1}\right) / n$, the relative frequency of $E$. Denoting the two relative frequencies by $R(E)$ and $R(\bar{E})$ respectively, we see that each has a mathematically possible range of 0 to $l$ and that $R(E)+R(E) \equiv 1$. We have previously made a point of the regularity of proportion in random series. By this we mean the tendency of relative frequencies to stabilize at definite values as the number of trials increases. This tendency has been verified experimentally on numerous occasions and seems to be inherent in the nature of
random phenomena. Therefore, the existence of a limiting value is postulated, and the probability $\overline{P(E)}$ of the event $\bar{E}$ is defined as the limit approached by the relative frequency as the number of trials increases indefinitely:

$$
\begin{equation*}
P(E) \equiv \operatorname{Lim}_{n \rightarrow \infty} \frac{n_{1}}{n} \equiv \operatorname{Lim}_{n \rightarrow \infty} R(E) \tag{1.1}
\end{equation*}
$$

Similarly,

$$
\begin{equation*}
P(\bar{E}) \equiv \operatorname{Lim}_{n \rightarrow \infty}\left(n-n_{1}\right) / n \equiv \operatorname{Lim}_{n \rightarrow \infty} R(\bar{E}) \text { and } P(E)+P(\bar{E})=1 \tag{1.2}
\end{equation*}
$$

Properties of Relative Frequencies
Property l. For a single event $E$ we have already noted that $R(\bar{E})$ is a real number lying somewhere in the range 0 to 1 and such that $R(E)+R(\bar{E})=1$.

Now consider two events $A$ and $B$, which may exist simultaneously. All possible results of a given experiment may be classified under some one of four mutually exclusive categories: ( $\mathrm{A}, \mathrm{B}),(\mathrm{A}, \overline{\mathrm{B}}),(\overline{\mathrm{A}}, \mathrm{B}),(\overline{\mathrm{A}}, \overline{\mathrm{B}})$--representing the simultaneous occurrence of $A$ and $B$, the simultaneous occurrence of $A$ and $\bar{B}$, etc. Denoting the respective numbers of occurrences by $n_{1}, n_{2}$, $n_{3}$, and $n_{4}$ and the total by $n$, we may summarize the results as shown in Table l. 1.

Table 1. 1

| Category | $(A, B)$ | $(A, \bar{B})$ | $(\bar{A}, B)$ | $(\bar{A}, \bar{B})$ | Total |
| :---: | :---: | :---: | :---: | :---: | :---: |
| Number of <br> Occurrences | $n_{1}$ | $n_{2}$ | $n_{3}$ | $n_{4}$ | $n$ |

The total number of occurrences of $A$ is $n_{1}+n_{2}$, and the total number of occurrences of $B$ is $n_{1}+n_{3}$. Hence, the corresponding relative frequencies are

$$
\begin{equation*}
R(A)=\frac{n_{1}+n_{2}}{n} ; R(B)=\frac{n_{1}+n_{3}}{n} \tag{1.3}
\end{equation*}
$$

It often happens that a sufficient condition for the occurrence of some contingent result is the single or simultaneous occurrence of two events $A, B$. The composite event defined by the occurrence of either A or B alone or both together is denoted by the symbol $A+B$. (As an analogy, a joint meeting of two organizations is open to all who belong to either one, including, of course, all who belong to both.) Since the definition of $A+B$ is satisfied by each of the categories ( $\mathrm{A}, \mathrm{B}$ ) $,(\mathrm{A}, \overline{\mathrm{B}}),(\overline{\mathrm{A}}, \mathrm{B})$, the relative frequency of $A+B$ is

$$
R(A+B)=\frac{n_{1}+n_{2}+n_{3}}{n}=\frac{\left(n_{1}+n_{2}\right)}{n}+\frac{\left(n_{1}+n_{3}\right)}{n}-\frac{n_{1}}{n}
$$

Hence,
Property 2. $R(A+B)=R(A)+R(B)-R(A, B)$
If the separate quantities $n_{l}, n_{2}$, etc. were known, the most direct method of calculating $R(A+B)$ would, of course, be simply that which follows from the definition, namely $\left(n_{1}+n_{2}+n_{3}\right) / n$. The point of introducing the formula stated in Property 2 is that most mathematical problems entail the expression of one variable in terms of others, and the relative frequencies $R(A), R(B)$, $R(A, B)$ are usually the ones most readily available either from direct observation or theoretical considerations.

We now introduce the idea of conditional relative frequencies. In the theory of probability, the word "conditional" signifies restriction to a special class as distinguished from the "unconditional" inclusion of all possibilities. An event E occurring under condition $C$ is denoted by the symbol $E \mid C$, read " $E$ given C. " A conditional relative frequency is calculated in the same manner as an ordinary (unconditional) relative frequency except that the calculation is confined to those events which satisfy the prescribed criterion (condition). The conditional relative frequency of $A$ given $B$, denoted by $R(A \mid B)$, is simply the proportionate number of occurrences of $A$ among all occurrences of $B$, hence the ratio of the number of simultaneous occurrences of $A$ and $B$ to the total number of occurrences of $B$ :

$$
\begin{equation*}
R(A \mid B)=\frac{n_{1}}{n_{1}+n_{3}}=\frac{\frac{n_{1}}{n_{n}}}{\frac{n_{1}+n_{3}}{n}}=\frac{R(A, B)}{R(B)} \tag{1.5}
\end{equation*}
$$

By a similar argument we obtain the conditional relative frequency of $B$ given $A$ as

$$
\begin{equation*}
R(B \mid A)=\frac{n_{1}}{n_{1}+n_{2}}=\frac{R(A, B)}{R(A)} \tag{1.6}
\end{equation*}
$$

Combining these statements into one chain of relationships, we arrive at

$$
\text { Property 3. } R(A, B) \equiv R(A) R(B \mid A) \equiv R(B) R(A \mid B)
$$

Laws of Probability
A system of axioms for a subject can be formulated in more than one way and yet have the same logical consequences. However, the following three properties, suggested by the behavior of relative frequencies, suffice for most purposes as an axiomatic basis of probability. We call them properties rather than axioms because, aside from formality of statement, a system of axioms should be reduced to the fewest possible assertions; whereas we have chosen an extended form of expression of Property 3 in the interest of convenience.

Property 1. (General Character of Probability) The probability $P(E)$ of an event $E$ is a real number in the range of 0 to $l$. The probability of an impossible event is 0 , that of an event certain to occur is 1 , and, in general, $P(E)+P(E)=1$.

Property 2. (Law of Total Probability)

$$
P(A+B)=P(A)+P(B)-P(A, B)
$$

Property 3. (Law of Compound or Joint Probability)
a. If neither $P(A)$ nor $P(B)$ is zero,

$$
P(A, B)=P(A) P(B \mid A)=P(B) P(A \mid B)
$$

b. If either $P(A)$ or $P(B)$ is zero,

$$
P(A, B)=0
$$

In a more advanced treatment of this subject, the counterpart of the number of occurrences of an event is a well-defined mathematical quantity known as measure. In connection with continuous variables, it is usually possible to interpret measure geometrically as area or volume. One of the paradoxes of an infinite number of possibilities, however, is that a probability of zero does not necessarily imply impossibility, nor does a probability of unity necessarily imply certainty. This may be understood by noting that a relative frequency will approach zero if the numerator remains finite while the denominator approaches infinity, or even if the numerator increases at too slow a rate in comparison with the denominator. Again, from the geometrical point of view, a rectangle may be shrunk to a line which has zero area but nevertheless the locus exists.
Fundamental Theorems
The quantity $P(A+B)$ is called the total probability of $A$ and $B$. If two events $A, B$ are mutually exclusive, their simultaneous occurrence is impossible, and $\bar{P}(A, B)=0$. In that case Property 2 yields $P(A+B)=P(A)+P(B)$. This result can be generalized to any finite number of mutually exclusive events. For, let A, B, C be mutually exclusive and set $S=A+B$, then $A+B+C \equiv S+C$ and

$$
P(S+C)=P(S)+P(C)=P(A)+P(B)+P(C)
$$

a result which can be extended by induction to any finite number of mutually exclusive events. Hence, the following theorem:

Theorem 1. (Law of Total Probability for Mutually Exclusive Events) If $A, B, \ldots, N$ are mutually exclusive events, then

$$
\begin{equation*}
\mathrm{P}(\mathrm{~A}+\mathrm{B}+\ldots+\mathrm{N})=\mathrm{P}(\mathrm{~A})+\mathrm{P}(\mathrm{~B})+\ldots+\mathrm{P}(\mathrm{~N}) \tag{1.7}
\end{equation*}
$$

With regard to an infinite number of events the extension is reasonable, but not rigorously demonstrable. Hence, one of the axioms in a technically complete set asserts that the extension does hold for an infinite number of mutually exclusive events. We shall take this extension for granted.

From Property 2 we may also deduce a more general theorem applicable to any finite number of events, whether mutually
exclusive or not. Let $A, B, C$ be any three events, which need not be mutually exclusive, and, as before, set $S=A+B$. Then

$$
\begin{aligned}
P(A+B+C) \equiv P(S+C) & =P(S)+P(C)-P(S, C)=P(A)+P(B) \\
& -P(A, B)+P(C)-P(S, C)
\end{aligned}
$$

Now the symbol ( $\mathrm{S}, \mathrm{C}$ ) means that C occurs in conjunction with A or $B$ or both and has the same logical import as the expression $(A, C)+(B, C)$, since the simultaneous occurrence of $(A, C)$ and ( $B, C$ ) implies nothing more nor less than the simultaneous occurrence of $A, B, C$. Accordingly,

$$
P(S, C) \equiv P[(A, C)+(B, C)]=P(A, C)+P(B, C)-P(A, B, C)
$$

Therefore,

$$
\begin{align*}
P(A+B+C)=P(A)+P(B) & +P(C)-P(A, B)-P(A, C)-P(B, C) \\
& +P(A, B, C) \tag{1.8}
\end{align*}
$$

When this result is extended by induction, the following theorem is obtained.

Theorem 2. (General Law of Total Probability) The probability $P(A+B+\ldots+N)$ equals the algebraic sum of the probabilities of the events in all possible distinct combinations: singles, pairs, triples, ..., N-tuple. The sign is plus for the odd combinations (singles, triples, etc.) and minus for the even combinations (pairs, quadruples, etc.)

The probability of the simultaneous occurrence of two or more events is called the compound probability, or synonymously, the joint probability of the events. The relations stated by Property 3 can be generalized to any finite number of events. Thus, considering the simultaneous occurrence of three events $A, B, C$, let $X$ denote the simultaneous occurrence of $A$ and $B$. Then

$$
(A, B, C)=(X, C)
$$

and

$$
P(A, B, C)=P(X) P(C \mid X)=P(A) P(B \mid A) P(C \mid A, B)
$$

Since all permutations of the letters $A, B, C$ have the same meaning as regards the simultaneous occurrence of the three events, there are six equivalent expressions of the foregoing type. For example, we also have

$$
P(A, B, C)=P(B, C, A)=P(B) P(C \mid B) P(A \mid B, C)
$$

Continuing in the same vein, we may express the joint probability of any finite number $N$ of events $A, B, C, \ldots, M, N$ as the product of N factors, the first of which is the unconditional probability of
any particular one of the events chosen arbitrarily, the second, the conditional probability of any particular one of the remaining events given the occurrence of the one first selected, and the general term, the conditional probability of any particular one of the remaining events given the occurrence of those already chosen. In all there are N! equivalent expressions for this same joint probability. Putting this rule in the form of a theorem, we have:

Theorem 3. (General Law of Compound Probability)

$$
P(A, B, C, \ldots, M, N)=P(A) P(B \mid A) P(C \mid A, B) \ldots P(N \mid A, B, C, \ldots, M)
$$

A special case of great importance arises when the events are independent. In the probability sense, two events $A, B$ are said to be independent when and only when neither one affects the probability of the occurrence of the other. The formal definition of independence is that

$$
P(B \mid A)=P(B) ; P(A \mid B)=P(A)
$$

It turns out that each one of these equations implies the other, for if either holds, the joint probability assumes the symmetrical form

$$
P(A, B)=P(A) P(B)
$$

from which the other equation can be deduced. Consequently, a sufficient definition of independence of two events is that their joint probability equals the product of their respective unconditional probabilities. An advantage of this form of definition is that it does not break down when one of the events has zero probability. The reader's attention is called to the fact that mutually exclusive events are not independent unless the probability of one of them is zero, for, by definition, the probability of the simultaneous occurrence of mutually exclusive events is necessarily zero. In general, $N$ events are independent if the probability of each event is unaffected by the occurrence or nonoccurrence of any of the others, either singly or in combination. A formal definition is as follows:

Definition of Independence: The events $A, B, C, \ldots, N$ are mutually independent as a system if the probability of every N -fold combination which can be formed from these events or their complements in any proportion (as ABC...N; $\bar{A} B C . . . N$; $\mathrm{A} \overline{\mathrm{B}} \mathrm{C} \ldots \mathrm{N} ; \ldots ; \overline{\mathrm{A}} \overline{\mathrm{B}} \mathrm{C} \ldots \mathrm{N} ; \ldots ; \overline{\mathrm{A}} \overline{\mathrm{B}} \overline{\mathrm{C}} . \ldots \mathrm{N})$ factors into the product of the probabilities of the N components of the combination.

This definition is, of course, both the necessary and sufficient condition that the events are independent. If, however, the events are independent, it follows that the joint probability is equal to the product of the individual probabilities. Hence,

Theorem 4. (Law of Compound Probability for Independent Events)

$$
\begin{equation*}
P(A, B, C, \ldots, N)=P(A), P(B), P(C) \ldots P(N) \tag{1.9}
\end{equation*}
$$

Since the $2^{N}$ possible $N$-fold combinations of $A, B, C, \ldots, N$ and $\bar{A}, \bar{B}, \bar{C}, \ldots, \bar{N}$ represent a complete system of mutually exclusive categories, their total probability is necessarily equal to unity. Moreover, the factorization of their probabilities in the case of independence guarantees the analogous probability factorization of all possible combinations of any subset of the initial N events and hence the independence of any subset. (This fact may be established by applying the law of total probability to a typical subset.) On the other hand, the independence of subsets does not suffice for the independence of the entire system.

## General Comments

Since the outcome of a contemplated action is often in doubt prior to its performance, the major use of probability lies in making judicious guesses, for probability represents the summary and generalization of experience. A considered decision to pursue a stated course of action is usually reached by weighing the odds associated with various possible outcomes, insofar as these eventualities can be foreseen. After the fact, the outcome of a particular action is uniquely determined; therefore, with reference to a single situation the probability of any designated event is either one or zero, depending upon whether it does or does not correspond to the actual result. Understood as a ratio, probability has only this trivial interpretation as applied to unique happenings, and we agree not to use the term probability in this sense. Hence, when we say that A has a probability $p$ of succeeding in a proposed enterprise, what we really mean is that according to available records on comparable cases, the relative frequency of successes is $p$. How pertinent the records themselves are may be open to question, but improving the estimate of probability by sharpening the classification is a project for research.

Practical applications of the theory of probability call for an intelligent combination of empirical knowledge and mathematical deduction. The basic probabilities are estimated by computing the appropriate relative frequencies from observational data, or in sufficiently clear-cut situations, they are inferred from a priori considerations. The probabilities of various composite events are then derived by applying the laws of probability to the component events, single or joint. Separate estimates are needed for joint probabilities unless the assumption of independence is justified.
Bayes' Theorem
Given a set of mutually exclusive events $B_{1}, B_{2}, \ldots, B_{n}$. let us assume that the occurrence of one or another of them is a necessary condition for the occurrence of an event A. Depending on the circumstances, the B's may precede A in time or may occur simultaneously with A. From the viewpoint of the logical present, both types of association may be regarded as constituting joint events, and we shall denote the corresponding probabilities by the same symbol $P\left(B_{i}, A\right)$, where $B_{i}$ is a specific one of the B's.

From Property 3 we have

$$
P\left(B_{i}, A\right)=P\left(B_{i}\right) P\left(A \mid B_{i}\right)=P(A) P\left(B_{i} \mid A\right)
$$

Therefore, the conditional probability of $B_{i}$ given $A$ is

$$
P\left(B_{i} \mid A\right)=\frac{P\left(B_{i}\right) P\left(A \mid B_{i}\right)}{P(A)}
$$

This conditional probability is understood to mean (in elementary terms) the proportionate number of times the antecedent (or concomitant) of $A$ is $B_{i}$. In connection with the use of probability in deciding upon a course of action, we have previously noted an idiomatic contraction of the precise formulation of the probability statement, and a similar idiom is used in this context. Thus, the following type of question is often propounded: 'Having observed A on a particular trial, what is the probability that the antecedent was $\mathrm{B}_{\mathrm{i}}$ ?' Taken literally, this question refers to a unique situation and has only the trivial answer of one or zero, as the case may be; but it is really meant to refer (as it were) to the relative frequency of $B_{i}$ among all situations characterized by the occurrence of $A$.

While the probability $P(A)$ could be estimated directly as a relative frequency if records were available, it is often more feasible to depend upon mathematical synthesis. For instance, one might have access to a large amount of data on the relative frequencies of the B's, but comparatively little data on A. This would be the case if A were a new development in the economic world or perhaps a newly discovered symptom in medical research. Under such circumstances it might be possible to deduce the conditional probabilities $P\left(A \mid B_{i}\right)(i=1,2, \ldots, n)$ from theoretical considerations (typically true of kinematic problems) or to design small-scale experiments by which the conditional probabilities could be estimated. When this is possible, $P(A)$ can be computed from the other information. Since the B's are mutually exclusive, the event $A$ is logically equivalent to the following sum of mutually exclusive events:

$$
\left(B_{1}, A\right)+\left(B_{2}, A\right)+\ldots+\left(B_{n}, A\right)
$$

Hence, by Theorem 1,

$$
P(A)=P\left(B_{1}, A\right)+P\left(B_{2}, A\right)+\ldots+P\left(B_{n}, A\right)
$$

which, by Property 3, may be expressed as

$$
P(A)=P\left(B_{1}\right) P\left(A \mid B_{1}\right)+P\left(B_{2}\right) P\left(A \mid B_{2}\right)+\ldots+P\left(B_{n}\right) P\left(A \mid B_{n}\right)
$$

Therefore, substituting this result in the denominator of the expression for $P\left(B_{i} \mid A\right)$, we obtain what is known as Bayes' Theorem:
$P\left(B_{i} \mid A\right)=\frac{P\left(B_{i}\right) P\left(A \mid B_{i}\right)}{P\left(B_{1}\right) P\left(A \mid B_{1}\right)+P\left(B_{2}\right) P\left(A \mid B_{2}\right)+\ldots+P\left(B_{n}\right) P\left(A \mid B_{n}\right)}$

Example An unbiased coin is tossed, and, if it comes up heads, a black ball is placed in an urn, but if tails, a white ball. This is done four times. Another person now samples the urn by drawing out two balls simultaneously, which turn out to be black. What is the probability that there were two black and two white balls in the urn? Because of the method used in filling the urn, there exist five possibilities for the final color distribution of the four balls, and the probability of the occurrence of each color combination can be computed. They are as follows:

Four white $\left(B_{1}\right): P\left(B_{1}\right)=\frac{1}{16}$
Three white and one black $\left(\mathrm{B}_{2}\right): P\left(\mathrm{~B}_{2}\right)=\frac{4}{16}$
Two white and two black $\left(\mathrm{B}_{3}\right): P\left(\mathrm{~B}_{3}\right)=\frac{6}{16}$
One white and three black $\left(\mathrm{B}_{4}\right): P\left(\mathrm{~B}_{4}\right)=\frac{4}{16}$
Four black $\left(B_{5}\right): P\left(B_{5}\right)=\frac{1}{16}$
Since in this case the event (A) cannot occur with antecedents $B_{1}$ and $B_{2}$, the conditional probabilities $P\left(A \mid B_{i}\right)(i=1,2)$ must equal zero. The other three conditional probabilities are computed as

$$
\mathrm{P}\left(\mathrm{~A} \mid \mathrm{B}_{3}\right)=\frac{1}{\mathrm{C}(4,2)}=\frac{1}{6} ; \mathrm{P}\left(\mathrm{~A} \mid \mathrm{B}_{4}\right)=\frac{\mathrm{C}(3,2)}{\mathrm{C}(4,2)}=\frac{1}{2} ; \mathrm{P}\left(\mathrm{~A} \mid \mathrm{B}_{5}\right)=1
$$

and therefore the required probability is given by Bayes' Formula:

$$
P\left(B_{3} \mid A\right)=\frac{\left(\frac{6}{16}\right)\left(\frac{1}{6}\right)}{\left(\frac{6}{16}\right)\left(\frac{1}{6}\right)+\left(\frac{4}{16}\right)\left(\frac{1}{2}\right)+\left(\frac{1}{16}\right)(1)}=\frac{6}{6+12+6}=\frac{1}{4}
$$

Geometrical Probability
Many problems of a probability nature present facets which can be solved geometrically. For this reason. we shall now discuss a simple example of geometrical probability. This subject introduces in a natural way a broader concept of probability which goes beyond the elementary notion of the number of occurrences of discrete events.

In the nature of geometrical probability, all points under consideration lie within prescribed boundaries, and the probability measure is so defined that the total probability of the admissible region is unity, while that of all exterior space is zero. Since any finite portion of space, no matter how small, contains an infinite number of points, probability cannot be defined in terms of the number of points included. Instead, it is defined in terms of the geometrical measure appropriate to the dimensionality of the admissible region: in one dimension, length; in two dimensions, area; in three dimensions, volume. Although it is perfectly possible to define probability as a variable function of position, in geometrical problems it is usually assumed that the probability measure of any subdivision of the admissible region
is directly proportional to the size of the subdivision. Thus, in one dimension, if the admissible region is a line segment of length $R$, the probability measure of any included segment of length $r$ is $r / R$, that of an infinitesimal segment of length $d r$ is $d r / R$, and that of the whole segment is $R / R=1$. The statement "a point is chosen at random" is a conventional expression for the fact that the chosen point can be any point within the region designated, and the probability of its falling within any stated portion of that region is equal to the probability measure of the portion itself.

Example The base $x$ and altitude $y$ of a triangle are obtained by picking points $X$ and $Y$ at random on two line segments of length $a$ and $b$ respectively. (See Fig. 1.1) What is the probability that the area of the triangle with base $x$ and altitude $y$ is less than $a b / 4$ ? Since the area of the triangle is given by $x y / 2$, this quantity is required to be less than $a b / 4$. Now


Fig. 1.1


Fig. 1.2
the hyperbola $x y=a b / 2$ divides the admissible area, which is a rectangle with base $a$ and altitude $b$, into two parts $I$ and II. (See Fig. 1.2) Pairs of values for $\bar{x}$ and $y$ which fall in area I will determine triangles with area less than $a b / 4$, and, of course, those points in II will determine triangles with areas greater than $a b / 4$. Therefore, the desired probability is

$$
P=\frac{\frac{a b}{2}+\int_{\frac{a}{2}}^{a} \frac{a b}{2 x} d x}{a b}=\frac{1}{2}(1+\ln 2)=.85
$$

Distribution Functions and Probability Densities
Up to this point, probability has been discussed from the point of view of individual events. More generally, however, we are dealing with the variables produced by the interplay of a complex system of causes which exhibit irregular variations which are, to all intents and purposes, random. These variables which elude predictability in assuming their different possible values, whether finite or infinite, are called random variables, or, synonomously, variates. Whether we are talking about a finite number of possible outcomes, a set of numerably infinite (but discrete) outcomes, or an infinite set of outcomes, we can speak of a function $F(x)$ which represents the probability that the random
variable will take on this value of x or one which is less. This, of course, is under the assumption that these sets constitute an exhaustive set of possibilities. $F(x)$ is called the distribution function defined over the region of definition of $x$ and, of course, has the properties that $F(-\infty)=0$ and $F(\infty)=1$. If there exists a finite probability that the variate x will take on a specific value, then, of course, the distribution function takes a jump at this particular value, and, if we are dealing with a variate which can take on only discrete values, then the distribution function consists entirely of jumps and that portion of the function between these values consists of horizontal lines. (See Fig. 1.3)

(a) Continuous

(b) Discrete

(c) Combination of Discrete and Continuous

Fig. 1.3
Since $F(x)$ is not defined at one of these jumps, we shall define it such that $F(x)=F(x+0)$. In the particular case where $F(x)$ is continuous, or in the case where it is continuous over some region of its definition, it is possible to differentiate this function over that region such that $\mathrm{dF}(\mathrm{x})=\mathrm{f}(\mathrm{x}) \mathrm{dx}$. In this instance $f(x)$ is called the probability density, and thus the probability that the variate will take on a value between $x$ and $x+d x$ is given by $f(x) d x$. In the case that there exists a finite probability at the point $x_{i}$, the value will be denoted by $f\left(x_{i}\right)$.

This description of distribution functions can, of course, be applied to any number of random variables. For example, $F(x, y)$ would represent the probability that x would take on some value equal to or less than this value of $x$ at the same time that $y$ would take on a value equal to or less than that value of $y$. By our previous definition of independence, if $F(x, y)=F_{1}(x) F_{2}(y)$, where the two functions $F_{1}$ and $F_{2}$ represent the distribution functions of $x$ and $y$ respectively, then $x$ and $y$ are said to be independent. If $F(x, y)$ is differentiable with respect to both variables, then $\frac{\partial^{2} F(x, y)}{\partial x^{\partial} y}=f(x, y)$, where $f(x, y)$ is the joint probability density of the two variables and $f(x, y) d x d y$ represents the probability that $x$ will lie between $x$ and $x+d x$ at the same time that $y$ lies between $y$ and $y+d y$.

Utilizing our previous concepts of probability, it is now possible to define the total probability density of $x$ in terms of the joint probability density of $x$ and $y$ as $\int_{-\infty}^{\infty} f(x, y) d y=g_{1}(x)$ and
the same method would obtain $g_{2}(y)$. Now, if $x$ and $y$ are independent, then $f(x, y)=g_{1}(x) g_{2}(y)$, and this is a necessary and sufficient condition for independence of the variables.

Again, using our concepts of conditional probability, we may now define the probability density of $y$ for a fixed value of $x$. Thus the conditional probability of $x$ would be defined as $h_{1}(y \mid x)$ $=\frac{f(x, y)}{g_{1}(x)}$. The same argument gives us the conditional probability of $x$ for a fixed value of $y$ as $h_{2}(x \mid y)=\frac{f(x, y)}{g_{2}(y)}$. Thus the same fundamental principles which were defined in the section on probability apply to probability functions, and the same arguments can be extended to any number of variables.
Expected Value
Let $F(x)$ be any distribution function of a random variable $x$ and let $\phi(x)$ be any continuous function defined over the region of definition of the random variable $x$. Then, the expected value of $\phi(x)$ is defined as $E[\phi(x)]=\int_{-\infty}^{\infty} \phi(x) d F(x)$. In this case we are using an elementary definition of the Stieltjes integral. In the case of a discrete variable this would be equal to $\sum_{i}^{n} \phi\left(x_{i}\right) f\left(x_{i}\right)$, and, in the case that $F(x)$ is a continuous function, then it would become $\int_{-m}^{\infty} \phi(x) f(x) d x$. In the case that $F(x)$ is a combination of discrete and continuous parts, then, of course, this integral consists of a combination of the two. It is interesting to note that, if $\phi(x)$ is set equal to a new variable $y$, then, since the probability density $g(y)=f\left[\phi^{-1}(y)\right]\left|\frac{d}{d y}\left[\phi^{-1}(y)\right]\right|$, the expected value of $\phi(x)$ equals the expected value of $y$ and equals $\int_{-\infty}^{\infty} y g(y) d y$. Thus the expected value of any function with respect to a given distribution function is the same as the average with respect to its own distribution function. The mean and the variance of any distribution function are defined respectively as follows:

$$
\begin{align*}
\text { Mean } & =E(x)=\int_{-\infty}^{\infty} x d F(x)=\mu  \tag{1.11}\\
\text { Variance } & =E\left[(x-\mu)^{2}\right]=\int_{-\infty}^{\infty}(x-\mu)^{2} d F(x)
\end{align*}
$$

Statistical Models Involving Differential Equations
Example Events occur at random along a line infinite in length. A distance $x$ is chosen at random anywhere along this line. On the basis of the following four assumptions, let us find the probability of exactly $n$ events occurring in this length x : (1) Statistical equilibrium (This hypothesizes that the probability of $n$ observations occurring in length $x$ is exactly the same irrespective of where $x$ originates and thus depends only on the value of x.$)$ (2) The probability that one observation will
fall in an interval $d x$ is proportional to $d x$. (3) The probability of two or more events in an interval dx is of higher order than dx. (4) The probability is differentiable with respect to $x$. In general,

$$
P(n, x+d x)=P(n, x) P(0, d x)+P(n-1, x) P(1, d x)+P(n-2, x) P(2, d x)
$$

$$
+\ldots
$$

and

$$
P(0, d x)=1-P(1, d x)-P(2, d x)-\ldots
$$

Then
$\frac{P(n, x+d x)-P(n, x)}{d x}=\frac{d P(n, x)}{d x}=\frac{P(1, d x)}{d x}[P(n-1, x)-P(n, x)]+\ldots$
Then, under assumptions (1), (2), (3),

$$
\begin{equation*}
\frac{d P(n, x)}{d x}+k P(n, x)=k P(n-1, x) \tag{1.12}
\end{equation*}
$$

This is a first-order differential equation with integrating factor $e^{k x_{j}}$ therefore,

$$
e^{k x} P(n, x)=k \int e^{k x} P(n-1, x) d x+c
$$

For $n=0$, we must have $P(-1, x)$ equal to zero, and, therefore,

$$
P(0, x)=c_{0} e^{-k x}
$$

with $c_{\text {o }}$ equal to unity. Using the integral solution recursively, we have

$$
\begin{equation*}
P(n, x)=\frac{(k x)^{n}}{n!} e^{-k x} \tag{1.13}
\end{equation*}
$$

The assumption of statistical equilibrium permits the origin of $x$ to be chosen arbitrarily without affecting the probability distribution function. This discrete probability function is known as the Poisson distribution.

Example We shall next find the corresponding probability distribution when the density of events is changing with time.

Let $k(t)$ be the density of the events along the previous line.


Fig. 1.4

[^0]Thus the probability that an observation occurs between $t$ and $\mathrm{t}+\mathrm{dt}$ is $\mathrm{k}(\mathrm{t}) \mathrm{dt}$.

Let $P(n, \tau, t)=$ the probability of $n$ calls occurring within a time $\tau$ after a beginning instant $t$. We wish to find this probability. The following relations may be seen to hold.
$P(n, \tau+d \tau, t)=P(n, \tau, t) P(o, d \tau, t+\tau)+P(n-1, \tau, t) P(1, d \tau, t+\tau)+\ldots$
and

$$
P(0, d \tau, t+\tau)=1-P(1, d \tau, t+\tau)-P(2, d \tau, t+\tau)-\ldots
$$

Since $P(l, d \tau, t+\tau)=k(t+\tau) d \tau$, we now have the differential equation

$$
\begin{equation*}
\frac{d P(n, \tau, t)}{d \tau}+k(t+\tau) P(n, \tau, t)=k(t+\tau) P(n-1, \tau, t) \tag{1.14}
\end{equation*}
$$

If we let,

$$
K(t, \tau)=\int_{0}^{\tau} k(t+\tau) d \tau
$$

we can solve the above equation by using an integrating factor $e^{K(t, \tau)}$. The solution gives $P(n, \tau, t)$ in terms of $P(n-1, \tau, t)$. Using the facts that $P(-1, \tau, t)=0$ and that the sum of the $P$ for all n is unity, we can show that:

$$
\begin{equation*}
P(n, \tau, t)=\frac{K^{n_{e}}-K}{n!} \tag{1.15}
\end{equation*}
$$

This is again the Poisson distribution, although for each $\tau$ and $t$ we must find $K$, and indicates great generality for purposes of application of this distribution.

Again, consider the case of a finite population of size $N$, where, on the average, a fraction K of the population are expected to fail, and where failures occur at random. However, once a member fails, it is no longer in the population, so that the population is reduced continually. If $N$ is very large and $K$ is small, we have the Poisson distribution. However, let us consider the situation where this is not the case.

Let $P\left(n^{\prime}, \tau, t\right)=$ probability that $n^{\prime}$ items fail in time $\tau$ following time $t$.
$P\left(n^{\prime}, \tau, t \mid n\right)=$ probability that $n^{\prime}$ items fail in time $\tau$, after $t$ given that $n$ items failed in time $t$.
$\mathrm{N}=$ size of population at time 0 .
$\mathrm{K}=$ average fraction of failures per unit time.
We assume (1) The probability of one failure in time $d \boldsymbol{\tau}$ is proportional to $d \tau$ and to the size of the population $=K(N-n) d \tau$.
(2) The probability of more than one failure is of higher order than $d \tau$. (3) The function $P$ is differentiable with respect to $\tau$. (4) Statistical equilibrium exists. Then

$$
\begin{aligned}
P\left(n^{\prime}, \tau+d \tau, 0\right) & =P\left(n^{\prime}, \tau, 0\right)\left[1-\left(N-n^{\prime}\right) \operatorname{Kd} \tau-\ldots\right] \\
& +P\left(n^{\prime}-1, \tau, o\right)\left(N-n^{\prime}+1\right) k d \tau+\ldots
\end{aligned}
$$

which yields

$$
\frac{d P\left(n^{\prime}, \tau, o\right)}{d \tau}+\left(N-n^{\prime}\right) K P\left(n^{\prime}, \tau, 0\right)=\left(N-n^{\prime}+1\right) k P\left(n^{\prime}-1, \tau, 0\right)
$$

For $n^{\prime}=0, P(-1, \tau, 0)=0$, and we now have as the solution

$$
\begin{equation*}
P\left(n^{\prime}, \tau, 0\right)=\frac{N!}{\left(N-n^{\prime}\right)!n^{\prime}!} e^{-N K \tau}\left(e^{K \tau}-1\right)^{n^{\prime}} \tag{1.16}
\end{equation*}
$$

To generalize the analysis, we may use the $P\left(n^{\prime}, \tau, o\right)$ to obtain $P\left(n^{\prime}, \tau, t \mid n\right)$ by replacing $N$ by $N-n$, and zero by $t$. Thus,

$$
P\left(n^{\prime}, \tau, t \mid n\right)=\frac{(N-n)!}{n^{\prime!}\left(N-n-n^{\prime}\right)!} e^{-(N-n) K \tau}\left(e^{K \tau}-1\right)^{n^{\prime}}
$$

Since

$$
P\left(n^{\prime}, \tau, t\right)=\sum_{n=0}^{n^{\prime}} P\left(n^{\prime}, \tau, t \mid n\right) P(n, t, 0)
$$

$=\sum_{n=0}^{n^{\prime}} \frac{(N-n)!}{n^{\prime}!\left(N-n-n^{\prime}\right)!} e^{-(N-n) K \tau}\left(e^{K \tau}-1\right)^{n^{\prime}} \frac{N!}{(N-n)!n!} e^{-N K t}\left(e^{K t}-1\right)^{n}$
which we can rewrite as
$=\frac{N!}{n^{\prime}!\left(N-n^{\prime}\right)!} e^{-N K(t+\tau)}\left(e^{K \tau}-1\right)^{n^{\prime}} \sum_{n=0}^{n^{\prime}} \frac{\left(N-n^{\prime}\right)!}{n!\left(N-n^{\prime}-n\right)!}\left(e^{K(t+\tau)}-e^{K \tau}\right)^{n}$
If we recognize the sum as the binomial expansion of $\left(1-\mathrm{e}^{\mathrm{K} \tau}+\mathrm{e}^{\mathrm{K}(\mathrm{t}+\tau)}\right)^{\mathrm{N}-\mathrm{n}^{\prime} \text {, then* }}$
$P(n, \tau, t)=\frac{N!}{n^{\prime}!\left(N-n^{\prime}\right)!}\left(e^{K \tau}-1\right)^{n^{\prime}} e^{-N K(t+\tau)}\left(1-e^{K \tau}+e^{K(t+\tau)}\right)^{N-n^{\prime}}$

Mathematical Models Involving Difference Equations
The probability of obtaining exactly $x$ success in $t$ inde pendent trials, assuming the probability of success on the $k^{\text {th }}$ trial is $p_{k}$, can be found as the solution of the difference equation

$$
p_{x t}=p_{t} p_{x-1, t-1}+q_{t} p_{x, t-1}
$$

where $p_{x t}$ is the probability that is required and the following boundary conditions are specified.

$$
p_{x, o}=0 \quad x>0, \quad p_{o \circ}=1, \quad p_{o t}=q_{1} q_{2} \cdots q_{t} \quad t>0
$$

[^1]Assume a solution of the form

$$
a_{t}(\xi)=p_{o, t}+p_{1, t} \xi+\ldots=\sum_{x=0}^{\infty} p_{x, t} x^{x}
$$

then

$$
\begin{aligned}
& q_{t} a_{t-1}(\xi)=q_{t} p_{o, t-1}+\sum_{x=1}^{\infty} q_{t} p_{x, t-1 \xi}{ }^{x} \\
& \xi p_{t} a_{t-1}(\xi)=\sum_{x=1}^{\infty} p_{t} p_{x-1, t-1} \xi^{x}
\end{aligned}
$$

Where, by the addition of these two expressions and with the aid of the boundary condition, we have

$$
a_{t}(\xi)=\left(p_{t} \xi+q_{t}\right) a_{t-1}(\xi) \quad \text { and } \quad a_{o}(\xi)=0
$$

Therefore,

$$
a_{t}(\xi)=\left(p_{1} \xi+q_{1}\right)\left(p_{2} \xi+q_{2}\right) \ldots\left(p_{t} \xi+q_{t}\right)
$$

If all the $p_{k}$ are equal, we, of course, obtain the binomial distribution from the coefficient of $\xi^{x}$ in $(p \xi+q)^{t}$ or

$$
\begin{equation*}
p_{x, t}=\frac{(t)(t-1) \ldots(t-x+1)}{x!} p^{x} q^{t-x}=C(t, x) p^{x} q-x \tag{1.18}
\end{equation*}
$$

Example Two companies A and B, at the present time, have an equal chance of obtaining any new customer who comes into the market. By increasing the advertising budget, A can increase its probability to $\mathrm{p}=.75$. The executives do not wish to proceed spending this money unless they are reasonably sure that they can obtain 100 new customers before the competition can obtain 50. We may compute this probability from the difference equation

$$
p_{x, t}=.75 p_{x-1, t}+.25 p_{x, t-1}
$$

where $p_{x, t}$ is the probability that A will obtain his 100 customers before $B$ his 50 , where A has $x$ to go to reach 100 and $B$ has $t$ remaining to reach 50. The boundary conditions are

$$
\begin{array}{cc}
p_{x, 0}=0 & x>0 \\
p_{0, t}=1 & t>0 \quad p_{00}=0 \\
\text { If } a_{x}(\xi)=p_{x, 0}+p_{x, 1} \xi+p_{x, 2} \xi_{2}+\ldots=\sum_{t=0}^{\infty} p_{x, t} \xi t
\end{array}
$$

then, as before,

$$
a_{x}(\xi)=\frac{.75}{1-.25 \xi} a_{x-1}(\xi) \quad \text { and } \quad a_{0}(\xi)=\frac{\xi}{1-\xi}
$$

whence

$$
a_{x}(\xi)=\frac{\xi(.75)^{x}}{(1-\xi)(1-.25 \xi)^{x}}
$$

and thus

$$
P=p_{100,50}=(.75)^{100}\left[1+\frac{100}{1}(.25)+\frac{(100)(101)(.75)^{2}}{2!}+\ldots\right]
$$

Example Let us consider a sequence of trials which are the result of correlated events and where the outcome for any trial can result in the occurrence of any one of three categories. The next trial is related to the present outcome in a probability sense.

Let $p_{i}$ equal the unconditional probability of the occurrence of category $i(i=1,2,3)$ such that $p_{1}+p_{2}+p_{3}=1$. Also, $a_{i j}$ is the conditional probability that category $i$ will occur in the next trial given that category $j$ occurred in the present trial ( $i, j=1,2,3$ ). The $a_{i j}$ 's are not all independent, since, of course,

$$
p_{i}=p_{i} a_{i 1}+p_{2} a_{i 2}+p_{3} a_{i 3} \quad i=1,2,3
$$

Let $W\left(i: n_{;} ; n_{1}, n_{2}, n_{3}\right)$ be the probability of exactly $n_{1}$ instances of category $1, n_{2}$ instances of category $2, n_{3}$ instances of category 3 out of $n$ trials, but where the $n^{\text {th }}$ trial belonged to category $i$ ( $i=1,2,3$ ). Then the following difference equations hold with the necessary boundary conditions, provided the notation is simplified as $W\left(i: n_{i} ; n_{1}, n_{2}, n_{3}\right)=\bar{U}(i)$ and $\sum_{i=1}^{3} \bar{U}(i)=\bar{U}$

$$
i=1
$$

(1) $W\left(1: n+1 ; n_{2}, n_{3}\right)=a_{11} U(1)+a_{12} \bar{U}(2)+a_{13} \bar{U}(3)$
(2) $W\left(2: n+1 ; n_{1}, n_{2}+1, n_{3}\right)=a_{21} \bar{U}(1)+a_{22} \bar{U}(2)+a_{23} \bar{U}(3)$
(3) $W\left(3: n+1 ; n_{1}, n_{2}, n_{3}+1\right)=a_{31} U(1)+a_{32} \bar{U}(2)+a_{33} \bar{U}(3)$
where

$$
\begin{aligned}
& W\left(1: n ; 0, n_{2}, n_{3}\right)=W\left(2: n ; n_{1}, 0, n_{3}\right)=W\left(3: n_{1} ; n_{1}, n_{2}, 0\right)=0 \\
& W(1: n ; n, 0,0)=p_{1} a_{11}^{n-1} ; W(2: n ; 0, n, 0)=p_{2} a_{22^{n-1}}^{n} \\
& W(3: n ; 0,0, n)=p_{3} a_{33^{n-1}}
\end{aligned}
$$

Let us now consider this probabilistic situation by considering the following generating functions and remembering that one of the categories ( 3 in this case) is dependent:

20

$$
f(i)=\sum_{n=1}^{\infty} \sum_{n_{2}=0}^{n-n_{1}} \sum_{n_{1}=0}^{n} \bar{U}(i) x_{1}^{n}{ }_{1}^{n} x_{2}^{n} x_{3}^{n}
$$

where $\mathrm{i}=1,2,3$ are the corresponding generating functions for the above situations so that

$$
f\left(x_{1}, x_{2}, x_{3}\right)=f(1)+f(2)+f(3)=\sum_{n=1}^{\infty} \sum_{n_{2}=0}^{n-n_{1}} \sum_{n_{1}=0}^{n} \bar{U}_{1}^{n}{ }_{1}^{n} x_{2}^{n_{2}} x_{3}^{n}
$$

Multiply the first equation where $i=1$ by $x_{1} l^{+1} x_{2}^{n_{2}} x_{3}^{n+1}$ and sum over all indices and use the initial conditions to obtain

$$
\left(a_{11} x_{1} x_{3}-1\right) f(1)+a_{12} x_{1} x_{3} f(2)+a_{13} x_{1} x_{3} f(3)=-p_{1} x_{1} x_{3}
$$

Two more equations for the $f(i)$ 's may be obtained by multiplying the second equation by $x_{1}{ }^{n} x_{2}{ }_{2} 2^{+1} x_{3}^{n+1}$ and the last equation, where $i=3$, by $x_{1}^{n} l_{x_{2}}^{n} x_{x_{3}}^{n+1}$ and summing again. So, we have

$$
a_{21} x_{2} x_{3} f(1)+\left(a_{22} x_{2} x_{3}-1\right) f(2)+a_{23} x_{1} x_{3} f(3)=-p_{2} x_{2} x_{3}
$$

and

$$
a_{31} x_{3} f(1)+a_{32} x_{3} f(2)+\left(a_{33} x_{3}-1\right) f(3)=-p_{3} x_{3}
$$

These equations can be solved for $f(1), f(2)$, and $f(3)$, and the determinant of the coefficients is

$$
A=\left|\begin{array}{lll}
a_{11} x_{1} x_{3}-1 & a_{12} x_{1} x_{3} & a_{13} x_{1} x_{3} \\
a_{21} x_{2} x_{3} & a_{22} x_{2} x_{3}-1 & a_{23} x_{2} x_{3} \\
a_{13} x_{3} & a_{32} x_{3} & a_{33} x_{3}-1
\end{array}\right|
$$

which is different from zero in the neighborhood of $x_{1}=x_{2}=x_{3}=0$.
In fact, it is equal to

$$
\begin{aligned}
A=\Delta x_{1} x_{2} x_{3}^{3}-x_{3}^{2}\left[A_{11} x_{2}\right. & \left.+A_{22} x_{1}+A_{33} x_{1} x_{2}\right] \\
& +x_{3}\left(a_{11} x_{1}+a_{22} x_{2}+a_{33}\right)-1
\end{aligned}
$$

where

$$
\Delta=\left|\begin{array}{ccc}
a_{11} & a_{12} & a_{13} \\
a_{21} & a_{22} & a_{23} \\
a_{31} & a_{32} & a_{33}
\end{array}\right|
$$

and $A_{i j}$ are the cofactors.
Therefore,

$$
f\left(x_{1}, x_{2}, x_{3}\right)=f(1)+f(2)+f(3)=\frac{B}{A}
$$

where

$$
\begin{aligned}
B=-x_{1} x_{2} x_{3}^{3}+x_{3}^{2}\left[x_{1} x_{2}\left(A_{33}-p_{3} \Delta\right)\right. & \left.+x_{1}\left(A_{22}-p_{2} \Delta\right)+x_{2}\left(A_{11}-p_{1} \Delta\right)\right] \\
& -\left(p_{1} x_{1}+p_{2} x_{2}+p_{3}\right) x_{3}
\end{aligned}
$$

If the events themselves are independent, this reduces to the generating function for the multinomial, and the coefficient of $x^{n}$ is

$$
\left(p_{1} x_{1}+p_{2} x_{2}+p_{3}\right)^{n}
$$

In the case where the number of categories is reduced to two, we have

$$
\begin{equation*}
f\left(x_{1}, x_{2}\right)=\frac{\Delta x_{1} x_{2}^{2}+\left(p_{1} x_{1}+p_{2}\right) x_{2}}{\Delta x_{1} x_{2}^{2}-\left(a_{11} x_{1}+a_{22}\right) x_{2}+1} \tag{1.19}
\end{equation*}
$$

When all of the events are independent and equal to the $p_{i}{ }^{\prime} s$, we have the generating function for the binomial so that the coefficient of $x$ in $\left(p_{1} x_{1}+p_{2}\right)^{n}$

Thus the case of dependence between the occurrence of the successes in the various trials can be determined. B. O. Koopman* obtained the equivalent to the Poisson distribution when a correlation $r$ existed between subsequent trials. The limiting value of the binomial with this correlation $r$ is

$$
f_{\infty}(x)=e^{\frac{-c(1-r)(1-x)}{1-r x}}
$$

where $c=n p$ and is the limit of this product as $n \rightarrow \infty$ and $p \rightarrow 0$. Applications of Probability Theory

Example A seasonal article which must be ordered in advance and stocked by a department store sells for $\$ 100$ per unit and costs the store $\$ 50$ per unit irrespective of disposal; however,

[^2]any article not sold during the season must be sold at a sacrifice to a special dealer for $\$ 35$ per unit. Given that the distribution of customer orders for the item is $f(z)=e^{-9} 9 z / z$ ! ( $z=0,1,2, \ldots, \infty$ ) and that the number of orders during any season is a random draw from this distribution, how many units should be stocked in order to maximize the expected value of the profit?

Letting $x$ denote an arbitrary number of units stocked, we shall express the expected value of the profit as a function of $x$, say $\bar{P}(x)$, and then determine $x$ so as to maximize this function. The profit function is

$$
\left.\begin{array}{rlrl}
P_{A}=100 z+35(x-z) & 50 x & =65 z & 15 x
\end{array} \quad 0<z<x\right)
$$

The overall expected value of the profit $\overline{\mathrm{P}}(\mathrm{x})$, for any value of x , is given by

$$
\begin{aligned}
\bar{P}(x) & =\sum_{z=0}^{\infty} P(x) f(z)=\sum_{z=0}^{x} P_{A} f(z)+\sum_{z=x+1}^{\infty} P_{B} f(z) \\
& =\sum_{0}^{x}(65 z-15 x) f(z)+\sum_{x+1}^{\infty} 50 x f(z)=50 x+65 \sum_{z=0}^{\infty}(z-x) f(z)
\end{aligned}
$$

An obvious method of arriving at the maximum profit would be to evaluate this function at successive points until the maximum is located. For illustrative purposes we have carried out this calculation, with the results shown in Fig. 1. 5.


Fig. 1.5

To determine the optimum value of $x$, let us investigate whether we would increase or decrease the expected profit by increasing x by one unit. By substitution,

$$
\bar{P}(x+1)=50(x+1)+65 \sum_{0}^{x+1}(z-x-1) f(z)=50(x+1)+65 \sum_{0}^{x}(z-x-1) f(z)
$$

since $(z-x-1)=0$, when $z=x+1$. Subtracting $\bar{P}(x)$ from $\bar{P}(x+1)$, we find that the difference is

$$
\begin{aligned}
\Delta \bar{P}(x)=\bar{P}(x+1)-\bar{P}(x) & =50+65 \sum_{o}^{x}[(z-x-1)-(z-x)] f(z) \\
& =50-65 \sum_{0}^{x} f(z)=50-65 F(x)
\end{aligned}
$$

where $F(x)$ is the value of the distribution function of $z$ at $z=x$. From the latter equation, it is clear that $\Delta \bar{P}(x)$ will be positive as long as $F(x)<50 / 65$ and will be negative when $F(x)>50 / 65$, while, if $F(x)=50 / 65, \bar{P}(x)=\bar{P}(x+1)$. Therefore, the maximum value of the profit will be obtained if we choose $x$ to be the smallest value for which $\Delta \bar{P}(x)$ is zero or negative. For, in case $F(x)=50 / 65$ for an integral value of $x$, we would get the same expected profit with $x$ as with $x+1$, and the smaller number has the advantage of smaller investment.

Example A company has a sole purchaser for its product. If this purchaser on a particular day does not obtain the number of items which he requests, he will merely buy them on the open market, but his failure to obtain the necessary number from this company does not affect any future course of action. The margin of profit on the item is $m$ dollars, and, since the item is perishable, the loss is $n$ dollars for any items not sold to this single purchaser. This purchaser will buy either 100, 200, or 300 items on a given day, with probabilities $p_{1}, p_{2}, 1-p_{1}-p_{2}$ respectively. What should be the strategy of the company in order to maximize its profits?

Let x be the number of items the company produces. It is clear that $x$ will not be greater than 300, since this is the maximum number which can be sold. The profit will be given by different mathematical expressions, depending on whether x is less than 100 , between 100 and 200, and between 200 and 300. The following is the expected value of the profit (P) for each of these ranges of x .

$$
\begin{aligned}
& 0 \leq x \leq 100: \quad E(P)=m x \\
& 100<x \leq 200: \quad E(P)=p_{1}[100 m-n(x-100)]+\left(1-p_{1}\right) m x \\
&=100 m+(x-100)\left[m-p_{1}(m+n)\right]
\end{aligned}
$$

$$
\begin{aligned}
200<\mathrm{x} \leq 300: \quad \mathrm{E}(\mathrm{P}) & =\mathrm{p}_{1}[100 \mathrm{~m}-\mathrm{n}(\mathrm{x}-100)]+\mathrm{p}_{2}[200 \mathrm{~m}-\mathrm{n}(\mathrm{x}-200] \\
& +\left(1-\mathrm{p}_{1}-\mathrm{p}_{2}\right) \mathrm{mx}=200 \mathrm{~m}-100(\mathrm{~m}+\mathrm{n}) \mathrm{p}_{1} \\
& +(x-200)\left[m-(m+n)\left(p_{1}+p_{2}\right)\right]
\end{aligned}
$$



Fig. 1.6
In a simple case like this, one would guess without any analysis that there are only three strategies capable of maximizing the profit. The company would produce 100,200 , or 300 items. This is clearly shown in Fig. 1.6, where the expected value of the profit is plotted against $x$. In examining the equation for the expected value of the profit in the range from $x=100$ to $x=200$, we see that the slope is positive if $p_{1}<m / m+n$, which means that, if this inequality holds, 200 items should be produced as the profit increases continuously from $x=100$ to $x=200$. If, however, the slope is negative, the company would, of course, make only the 100 items. If we now examine the equation for the expected profit in the final range, that is from 200 to 300 , we find that the slope is positive for this segment if $p_{1}+p_{2}<m / m+n$, which again gives the condition that the company should move to the 300 -item level. It is obviously impossible for the slope to be negative in the middle range and positive in the final range. Therefore, the best strategy out of a theoretically infinite number of possible strategies is determined by comparing the probabilities $p_{1}$ and $p_{1}+p_{2}$ with the ratio $m /(m+n)$ of marginal profit per item to the sum of marginal profit per item and loss per item. The same analysis holds for a continuous demand which has the probability density function $f(v) \circ<v<\infty$, since in this case the expected value of the profit is

$$
\bar{P}(x)=\int_{0}^{x}[(m+n) v-n x] f(v) d v+\int_{x}^{V} m x f(v) d v
$$

or, rewriting $\int_{x}^{V} m x f(v) d v$ as $m x\left[1-\int_{0}^{x} f(v) d v\right]$,
the expected value becomes

$$
\bar{P}(x)=m x+(m+n) \int_{0}^{x}(v-x) f(v) d v
$$

so the maximum average profit is reached if x is so chosen that

$$
F(x)=\frac{m}{m+n}=\frac{\text { unit profit }}{\text { unit profit }+ \text { unit loss }}
$$

If $m=\$ 1$ and $n=\$ 2$, then $x$ must be chosen so that the area under $f(v)$ to its left will equal $1 / 3$.

It might be interesting to ask the question, at this point, as to what would happen if the loss of sales due to lack of the item might also produce a loss to the company because of a good-will factor. This loss of future profit could be taken into account during the current year by assigning a net loss to the company for each item for which there is a demand that cannot be supplied. From a practical viewpoint, the loss per item might be proportionately greater in case there were many items which could not be supplied. However, for illustrative purposes, let us consider that the loss to the company is a constant for each item that cannot be supplied. The integral for the expected value of the profit would then be

$$
\bar{P}(x)=\int_{0}^{x}[(m+n) v-n x] f(v) d v+\int_{x}^{V}[m x-a(v-x)] f(v) d v
$$

where $a$ is the loss to the company for each item not available to a customer. Differentiating this function with respect to x we obtain

$$
F(x)=\frac{m+a}{m+a+n}
$$

Therefore, the effect of the good-will factor in determining the value of $x$ for the best strategy is to add the loss per item due to good will to the margin of profit per item in order to determine an effective margin of profit.

In the above example without considering good will let us suppose that the probability density of sales can be represented by

$$
f(v)=\frac{1}{2 b} \quad(s-b \leq v \leq s+b)
$$

This is a situation in which a sales forecast indicates that the amount to be sold is $s$, but previous history indicates that there is an error in estimating the sales and that this error is equally likely to assume any value from -b to $b$. What is the additional profit that would be realized by stocking the number which maximizes the profit rather than stocking the amount $s$ which is the estimated sale of the item?

The expected value of the profit ( P ), if we stock $s$ is given by
$E\left(P_{s}\right)=\frac{1}{2 b} \int_{s-b}^{s}[(m+n) x-n s] d x+\frac{l}{2 b} \int_{s}^{s+b} m s d x=s m-\frac{b(m+n)}{4}$

By choosing $n$ sufficiently large in this formula, it is seen that the expected profit may become negative. The possibility of this situation is enhanced when $b$ is almost as large as $s$. The expression for the maximum expected value of the profit is given by

$$
E\left(P_{\max }\right)=\frac{1}{2 b} \int_{s-b}^{y}[m x-n(s-x)] d x+\frac{1}{2 b} \int_{y}^{s+b} \operatorname{msdx}
$$

where

$$
y=\frac{m(s+b)+n(s-b)}{m+n}
$$

Evaluating the maximum expected value of the profit by substituting this value of $y$ in the integral for $E\left(P_{\text {max }}\right)$, we obtain

$$
E\left(P_{\max }\right)=m s-\frac{b m n}{m+n}
$$

Even for $b=s$, this expression still remains positive for any $n$. The net gain in profit by stocking the value $y$ rather than the value $s$ is now obtained by subtracting the two expected profits as

$$
E\left(P_{\max }\right)-E\left(P_{s}\right)=\frac{b}{4} \frac{(m-n)^{2}}{(m+n)}
$$

This expression is equal to zero when $m=n$, but for $m \neq n$ the gain is always positive. The numerical unit of this calculated difference is necessarily the total gain in dollars since $b$ has the unit of number of items and both $m$ and $n$ are in dollars. This calculated difference is independent of $s$, but, of course, in most practical cases the range of error of forecasting is a function of the value forecasted. If we assume a constant percentage error of estimate in the sales, then $b=k s$ (where $k$, in this case, must lie between 0 and 1) and the percentage gain in profit would be given by

$$
\frac{100 \frac{\mathrm{ks}}{4} \frac{(m-n)^{2}}{m+n}}{s m-k s\left(\frac{m+n}{4}\right)}=\frac{100 k(m-n)^{2}}{(m+n)[4 m-k(m+n)]} \approx 25 k\left[\frac{(1-K)^{2}}{1+K}\right]
$$

The last approximate value is obtained under the assumption that $\mathrm{k}(\mathrm{m}+\mathrm{n})$ is small compared to 4 m and with the notation that K is the ratio of $n$ to $m$. As an illustration, if we lose $\$ 2$ for every $\$ 1$ (the margin of profit) so that $K=2$ and the range of error of the estimate of the sales forecast is $\pm 25 \%$, then the percentage gain in profit is approximately $2 \% .{ }^{-}$If, however, $\mathrm{k}=3$ in this instance, the gain is about $6 \%$.

Example Raw material must be ordered in the spring for sale in a finished product in the fall. The cost of raw material is $\$ \mathrm{c}_{1}$ per item. If this raw material is not used, it has a scrap value which is $\$ c_{2}$, where $c_{2}<c_{1}$. The raw material may be processed
during the summer at a cost of $g_{1}(x)$ dollars per item (that is, the unit cost), where $x$ is the number to be manufactured. The loss that is incurred on those completed products which are not sold can be again expressed as the fact that the scrap value is $\$ c_{3}$ per item. It is here assumed that the scrap value is less than the cost incurred to produce the item. On the other hand, it is possible to produce $W$ items during the fall season (thus at the time that the demand is known) at a higher rate of $g_{2}(W)$ dollars per item so that one never has to suffer the loss on manufactured goods not sold, but merely on the raw material. There is, however, a maximum total number (A) which can be produced during the season of sale itself because of limited plant capacity. With a selling price of $\$$ s per item, what is the maximum amount of raw material to be ordered and how much should be processed during the summer in order to maximize the profit, assuming the probability of sales can be represented by a density function $f(v) \quad 0<v<\infty$ ?

Assuming that the total number of pieces of raw material is made up of the $x$ pieces which are to be manufactured during the summer and $y$ pieces which are to be produced during the season if the demand is demonstrated, we have for the expected value of the profit the following expression, where it must be remembered that the function representing the profit differs in the three ranges of the volume of sales, namely, $0<v<x$, $\mathrm{x}<\mathrm{v}<\mathrm{x}+\mathrm{y}$, and $\mathrm{x}+\mathrm{y}<\mathrm{v}<\infty$.
Expected profit $=E(P)=\int_{0}^{x}\left[s v-c_{1}(x+y)-x g_{1}(x)+c_{2} y+c_{3}(x-v)\right] f(v) d v$

$$
\begin{aligned}
& +\int_{x}^{x+y}\left[s v-c_{1}(x+y)-x g_{1}(x)-(v-x) g_{2}(v-x)+c_{2}(x+y-v)\right] f(v) d v \\
& +\int_{x+y}^{\infty}\left[s(x+y)-c_{1}(x+y)-x g_{1}(x)-y g_{2}(y)\right] f(v) d v
\end{aligned}
$$

It is tentatively assumed at this point in the analysis that there is no restriction on the amount $y$ which can be manufactured during the season, and the integral for this expected profit can be differentiated first with respect to $x$ and then with respect to $y$.

For this situation the derivatives with respect to the limits on the integrals cancel each other in both equations because of the continuity conditions that exist from one range of $v$ to the next. The only time that differentiation with respect to the limits produces non-vanishing terms is when a setup charge or its equivalent is introduced at a particular volume of sales so that it occurs in one integral and not the other, which is not true here. Therefore, we have for the two derivatives

$$
\begin{aligned}
\frac{\partial E(P)}{\partial x} & =-c_{1}+c_{3}-\frac{d}{d x}\left[x_{1}(x)\right]+\left(c_{2}-c_{3}\right) \int_{x}^{x+y} f(v) d v \\
& -\int_{x}^{x+y} \frac{d}{d x}\left[(v-x) g_{2}(v-x)\right] f(v) d v+\left(s-c_{3}\right) \int_{x+y}^{\infty} f(v) d v
\end{aligned}
$$

$$
\frac{\partial E(P)}{\partial y}=-c_{1}+c_{2}+\left[s-c_{2}-\frac{d}{d y}\left[y g_{2}(y)\right]\right] \int_{x+y}^{\infty} f(v) d v
$$

The integral involving $\frac{d}{d x}\left[(v-x) g_{2}(v-x)\right]$ must be kept intact, since this term involves $v$, which is the variable of integration; whereas the other coefficients of $f(v)$ in the several integrals are constants.

Following the usual technique, one would set these two derivatives equal to zero in order to obtain a stationary point as the minimum. This would generate two equations in the two unknowns $x$ and $y$, which could then be solved by any of several methods of successive approximations.

It must be noted that in this problem there are very definite restrictions which have to be satisfied, and a stationary point may not occur within these limitations. The total area under the probability density $f(v)$ must equal unity, $x$ and $y$ must each be equal to or greater than zero, and $y$ cannot be greater than $A$. Such minimization with bilateral restrictions is, of course, often found in practical applications. In order to analyze the difficulties and to illustrate the solution of a problem in this category, we now simplify the above expressions by assuming $g_{1}(x)$ and $g_{2}(W)$ to be constants $c_{4}$ and $c_{5}$ respectively. Then

$$
\begin{gathered}
\frac{\partial E(P)}{\partial x}=\left(-c_{1}+c_{3}+c_{4}\right)+\left(c_{2}-c_{3}+c_{5}\right) \int_{x}^{x+y} f(v) d v+\left(s-c_{3}\right) \int_{x+y}^{\infty} f(v) d v \\
\frac{\partial E(P)}{\partial y}=-c_{1}+c_{2}+\left[s-c_{2}-c_{5}\right] \int_{x+y}^{\infty} f(v) d v
\end{gathered}
$$

If the expression for $\frac{\partial E(P)}{\partial y}$ is set equal to zero, we have

$$
\int_{x+y}^{\infty} f(v) d v=\frac{c_{1}-c_{2}}{s-c_{2}-c_{5}}
$$

Since $c_{1}>c_{2}$ and $s>c_{1}+c_{5}$ by the conditions imposed upon the problem, this integral is always positive and less than unity. It therefore always uniquely determines the value of $x+y$ if the function $f(v)$ is given, and thus the stationary point, as far as $\frac{\partial E(P)}{\partial y}$ is concerned, is also uniquely determined.

Substituting this value for the integral in $\frac{\partial E(P)}{\partial y}$, we have

$$
\frac{\partial E(P)}{\partial x}=\left(-c_{1}+c_{3}+c_{4}\right)+\left(c_{2}-c_{3}+c_{5}\right) \int_{x}^{x+y} f(v) d v+\frac{\left(s-c_{3}\right)\left(c_{1}-c_{2}\right)}{s-c_{2}-c_{5}}
$$

which gives the slope of the $E(P)$ surface in the $x$ direction.
For a stationary point to exist, it is necessary that $\frac{\partial E(P)}{\partial x}$ change sign. This is not always the case, as one can see by
letting $c_{3}=c_{5}$ and $c_{2}=0$, which are possible values of these constants. The sign of the derivative is then always positive, which indicates that the profit is increasing with $x$, so that we should not plan to produce any of the product after the season starts. Since the value of $x+y$ is known from the integral and $y=0$, we have determined the amount of $x$ to order. The exact opposite conclusion would be the case if for some values of the constants, the derivative were always negative with, therefore, $x=0$ and thus $y=x+y$. This, of course, neglects the limitation on the total amount of $y$ possible. More generally, the fact that $\int_{x}^{x+y} f(v) d v$ is restricted to lie between 0 and $1-\int_{\partial E(P)}^{\infty} f(v) d v$ means that there are many situations where $\frac{\partial E(P)}{\partial x}$ will not change sign in the possible region of the integral, and the above argument will again necessarily follow. For some ranges of the constants, the partial derivative $\frac{\partial E(P)}{\partial x}$ can be set equal to zero and a value for $\int_{x}^{x+y} f(v) d v$ determined which is within its range of possibility. In this case, we have a unique stationary point and can calculate the values of $x$ and $y$.

After $x$ and $y$ are determined, we must now consider the restriction that $y \leq A$. If $y$ comes out less than or equal to $A$, we have the final solution to the problem. If, however, $y$ is greater than $A$, it indicates that $y$ should be as large as possible, namely, A. We cannot, however, assign the excess of y over A to $x$, since this may not produce a maximum profit, but $E(P)$ must be again set up as a function of $x$ alone as

$$
\begin{aligned}
E(P) & =\int_{0}^{x}\left[s v-c_{1}(x+A)-x g_{1}(x)+c_{2} A+c_{3}(x-v)\right] f(v) d v \\
& +\int_{x}^{x+A}\left[s v-c_{1}(x+A)-x g_{1}(x)-(v-x) g_{2}(v-x)+c_{2}(x+A-v)\right] f(v) d v \\
& +\int_{x+A}^{\infty}\left[s(x+A)-c_{1}(x+A)-x g_{1}(x)-A g_{2}(A)\right] f(v) d v
\end{aligned}
$$

This function can now be differentiated with respect to $x$ and the derivative set equal to zero. Since the conditions of the problem were such that a profit is possible, and the range of x is between 0 and $\infty$, this process will yield a stationary point and therefore constitute a solution to the problem.

Example In 1920 Rutherford, Chadwick, and Ellis studied the emission of a-particles from a radioactive substance. It was found empirically that the distribution of the number of particles emitted during a time interval of 7.5 seconds was adequately represented by a Poisson distribution with $\mu=3.87$, or more generally that the number of particles emitted in $t$ seconds was a Poisson variate with $\mu=\rho t$, where $\rho=3.87 / 7.5=.516$. Instrumentally, it is customary to clock the length of time taken to register a fixed number of particles rather than to count the number of particles emitted during a fixed
interval of time. Let us find the distribution of the length of time required to register 40 particles. In general, let $G(t)$ denote the probability that it will take more than $t$ seconds for the emission of $n+1$ particles. Clearly, this can happen in $n+1$ mutually exclusive ways, namely by the emission of any smaller number of particles $0,1,2, \ldots, n$ in $t$ seconds. Hence, the corresponding total probability, obtained by summing the separate probabilities of each number of particles as given by the Poisson distribution, is

$$
G(t)=\sum_{y=0}^{n} e^{-\mu} \frac{\mu^{y}}{y!}
$$

and the latter sum can be expressed as an integral, which yields

$$
G(t)=\frac{1}{\Gamma(n+1)} \int_{\mu}^{\infty} x^{n} e^{-x} d x
$$

Now the distribution function $F(t)$ is the complement of $G(t)$, whence

$$
F(t)=1-G(t)=\frac{1}{\Gamma(n+1)} \int_{0}^{\mu} x^{n} e^{-x} d x
$$

and the density function $f(t)$ is given by

$$
f(t)=\frac{d F(t)}{d t}=\frac{d F(\mu)}{d \mu} \frac{d \mu}{d t}=\frac{\rho^{n+1}}{\Gamma(n+1)} t^{n} e^{-\rho t}
$$

Therefore, the distribution of waiting time $t$ for the emission of $\mathrm{n}+\mathrm{l}$ particles is

$$
f(t)=\frac{\rho^{n+1}}{\Gamma(n+1)} t^{n} e^{-\rho t} \quad(0, \infty)
$$

Distribution of Waiting Time for Emission of $n+1$ a-Particles

In particular, for the problem at hand, $n=39, \rho=.516$, and the distribution is

$$
f(t)=\frac{(.516)^{40}}{39!} t^{39} e^{-(.516) t} \quad(0 \leq t<\infty)
$$

Further interesting deductions follow readily from this result. Setting $n=0$, we find that the distribution of waiting time for one particle is given by the exponential distribution

$$
\begin{equation*}
f(t)=\rho e^{-\rho t} \quad(0 \leq t<\infty) \tag{1.20}
\end{equation*}
$$

Example The mail arrives at Company A some time between 8 a.m. and 9 a.m., and for practical purposes it is equally likely to be any time within that hour. The office boy must have the president's mail delivered to him by 9 a. m. The boy apparently delivers the mail any time between its arrival at the company
and $9 \mathrm{a} . \mathrm{m}$., and again, the time of this delivery is equally likely within the limits of the possible time interval. What is the probability that the mail will reach the president's desk between $t$ and $t+d t$ ?

The probability that the mail will be delivered to the company between $T$ and $T+d T$ is $d T / 1$ (since the unit of time is an hour) while the conditional probability that it will be delivered to the president's desk between $t$ and $t+d t$, knowing that it arrived at the company in the neighborhood of $T$, is $d t /(9-T)$. By definition the joint probability that the mail will arrive at the company between $T$ and $T+d T$ and be delivered to the president between $t$ and $t+d t$ is

$$
P(T, t) d T d t=\frac{d T d t}{9-T} \quad(8<T<9 ; T<t<9)
$$

Having obtained the joint distribution, the two marginals can be evaluated by means of integration. However, the marginal or total probability of $T$ is already known and the answer to our question involves the marginal distribution of $t$, which is

$$
P_{1}(t) d t=d t \int_{8}^{t} \frac{d T}{9-T}=-\ln (9-t) d t \quad(8<t<9)
$$

where again the unit of tirne is an hour. The area under the $P_{1}(t)$ curve is, of course, unity.

Example The average number of cars per day that take a country-road connection between two main roads is 15 . The only building located on this road is an antique shop, and the percentage of cars which go through this road that stop at the shop is . 12 . What is the chance that exactly $x$ cars stop at the antique shop on a given day?

Since the average number of cars per day is given and no further information regarding their destination or reason for traveling the road is contained in the statement, the only reasonable assumption one can make is that the number of cars is Poisson distributed on any given day. This would be equivalent to assuming that the cars go through this road individually and collectively at random. However, if it is known that $n$ cars travel through this road on a given day with a constant probability of 12 of stopping, then the conditional probability of exactly $x$ stops out of $n$ cars must be binomially distributed. These two distributions are:

$$
P_{1}(n)=\frac{e^{-15}(15)^{n}}{n!} \quad(n=0,1,2, \ldots, \infty)
$$

and

$$
P_{2}(x \mid n)=C(n, x)(.88)^{n-x}(.12)^{x} \quad(x=0,1, \ldots, n)
$$

The joint probability of $n$ cars and $x$ stops will thus be the product of these two probabilities. The answer to our problem is, of course, the marginal distribution of $x$, and this is obtained
by summing over $n$. It is to be observed that $n$ must be at least as large as x or otherwise it would be impossible to have available the $x$ cars to stop at the antique shop. Therefore,

$$
P_{3}(x)=\sum_{n=x}^{\infty} \frac{e^{-15}(15)^{n}}{n!} C(n, x)(.88)^{n-x}(.12)^{x}
$$

By setting $n-x=y$ and observing that the resulting summation is merely the expansion for $e^{z}$, one obtains the result

$$
P_{3}(x)=\frac{e^{-15}(.12)^{x}(15)^{x}}{x!} \sum_{y=0}^{\infty} \frac{(15)^{y}(.88)^{y}}{y!}=\frac{e^{-(15)(.12)}[(15)(.12)]^{x}}{x!}
$$

This result is very general, and, if the average number of cars were $\lambda$ and the probability of a car stopping $p$, then the probability of exactly $x$ stops on any one day would be given by a Poisson distribution with the parameter $\mu$ equal to $\lambda$ p.

Example In a department store customers arrive in the time interval from $9 \mathrm{a} . \mathrm{m}$. to $5 \mathrm{p} . \mathrm{m}$. according to some distribution, which may be interpreted as a density function of time at arrival. Thus the total number of customers entering the store between $\tau$ and $\tau+\mathrm{d} \tau$ is $\mathrm{Nf}(\tau) \mathrm{d} \tau$, where N is the total number of customers in any one day. The probability that a person coming in at time $\tau$ will leave between $t$ and $t+d t$ will, of course, be denoted by $g(t \mid \tau) d t$. What is the distribution of the number of people in the store between the hours of $9 \mathrm{a} . \mathrm{m}$. and $5 \mathrm{p} . \mathrm{m}$. ?

If we change the scale so that 0 corresponds to $9 \mathrm{a} . \mathrm{m}$. and 1 to $5 \mathrm{p} . \mathrm{m}$. and redefine our notation accordingly, we have the condition that the integral of $N f(\tau) d \tau$ from 0 to 1 is equal to $N$. We also have the condition that the integral of the conditional distribution $g(t \mid \tau) d t$ from $\tau$ to 1 must equal unity. The average number of people actually in the store at any time $T$ is given by $\mathrm{Nw}(\mathrm{T})$, where $w(T)$ is the function to be determined. Now

$$
\begin{aligned}
& \mathrm{Nf}(\tau) \mathrm{d} \tau=\text { number who enter between } \tau \text { and } \tau+\mathrm{d} \tau \\
& \int_{\tau}^{\mathrm{T}} \mathrm{~g}(\mathrm{t} \mid \tau) \mathrm{dt}=\begin{array}{l}
\text { fraction of those people who come in at } \tau \\
\text { leave between } \tau \\
\\
1-\int_{\tau}^{\mathrm{T}} \mathrm{~g}(\mathrm{t} \mid \tau) \mathrm{d} d
\end{array} \\
&=\begin{array}{l}
\text { fraction who enter at } \tau \\
\text { at time } T .
\end{array}
\end{aligned}
$$

Then

$$
N w(T)=N \int_{0}^{T} f(\tau) d \tau\left[1-\int_{\tau}^{T} g(t \mid \tau) d t\right]
$$

which now permits $w(T)$ to be evaluated if $f(\tau)$ and $g(t \mid \tau)$ are known.
Example In the inspection of light bulbs, vacuum tubes, condensers, and many other items, where a destructive test is necessary in order to determine their life or other characteristics, it is
desirable often to simplify the testing techniques. For example, if one is testing electric light bulbs in order to determine the length of life, out of $n$ bulbs tested, it is possible that some of these will continue to burn for a long while and that a test rack or shipment will be held up for this period. One technique to overcome this difficulty is to test a large number of items, but on the basis of the time to failure of the first few, determine the characteristics of the entire lot. Assuming a very simple life curve for electric light bulbs of the form $p(t)=\frac{1}{\bar{t}} e^{-t / \sqrt{t}}$, where $\bar{t}$ is the mean life of a large lot, the probability that an individual bulb will burn out between $t$ and $t+d t$ is $p(t) d t$. For a fixad value of $\bar{t}$, what is the probability of obtaining a pair of values for $t_{1}$ and $t_{2}$ which would be less probable than a given observed pair?

If $n$ bulbs are placed on the test rack and the time to failure observed in all cases, we would produce a series of times $t_{1}, t_{2}$, $\ldots, t_{n}$ which we might consider in a sequence of order in which $t_{l}$ is the smallest and $t_{n}$ the longest. The joint distribution of $t_{1}$ and $t_{2}$ may be obtained in the usual way as

$$
\begin{aligned}
p\left(t_{1}, t_{2}\right) d t_{1} d t_{2}= & \frac{n!}{1!1!(n-2)!}\left[\frac{1}{\frac{1}{t}} e^{-t / \sqrt{t}} d t_{1}\right]\left[\frac{1}{\frac{1}{t}} e^{-t_{2} / \tilde{t}} d t_{2}\right] . \\
& {\left[\frac{1}{\frac{t}{t}} \int_{t_{2}}^{\infty} e^{-t / t} d t\right]^{n-2} \quad\left[\begin{array}{l}
0<t_{1}<\infty \\
t_{2}>t_{1}
\end{array}\right.} \\
= & (n)(n-1) e^{\left[t_{1}+(n-1) t_{2}\right]}
\end{aligned}
$$

It must be true, of course, that

$$
\int_{0}^{\infty} \int_{t_{1}}^{\infty} p\left(t_{1}, t_{2}\right) d t_{1} d t_{2}=1
$$

If the experiment is conducted as outlined, it means that the time of the failure of the first bulb is noted, and we will call this value $a$, and the time to the second failure will be $b$. Thus, in each case the experiment would result in an actual pair of values ( $t_{1}, t_{2}$ ) denoted as ( $a, b$ ). Since the probability density function is a constant, if the exponent of $e$ is constant, all pairs of values which satisfy the condition $t_{1}+(n-1) t_{2}=a+(n-1) b=c$ have the same probability of occurrence for a given value of $\bar{t}$ as the observed point ( $\mathrm{a}, \mathrm{b}$ ). In reference to Fig. 1. 7, it is now possible to find the probability that $\left(t_{1}, t_{2}\right)$ will fall in the shaded area having observed the fixed point ( $a, b$ ) as

$$
\begin{aligned}
p & =\frac{n(n-1)}{t^{2}} \int_{0}^{c / n} d t_{1} e^{-t_{1} / \bar{t}} \int_{t_{1}}^{\frac{c-t}{n-1}} e^{-\frac{t_{2}}{\bar{t}}(n-1)} d t_{2} \\
& =1-e^{-c / \bar{t}}\left(1+\frac{c}{t}\right)
\end{aligned}
$$



Fig. 1.7
Since $c$ is fixed by the observed value of the first and second bulb to burn out, and if we know the value of $\overline{\mathrm{t}}$, then the area as found from the above expression will give the probability that we would observe by chance a sample which might be considered "worse" than that which was observed.

Example The selling price of a specific article to the public will vary somewhere between $\$ .25$ and $\$ .35$ per item, depending upon the final cost of production as determined from the initial run. It is assumed that any price within this range is equally likely to occur. However, the volume of the sales as a probability density function is dependent upon this final retail figure and, expressed in millions of pieces, is of the form $f(v \mid c)=\left(c^{2} v\right) e^{-c v}$ $0<v<\infty$, where $c$ is the cost per piece in dollars. The effect of substituting values of $c$ ranging from .25 to .35 in this equation is gradually to move the mode of the curve toward the $y$ axis; thus increasing the price reduces the volume of sales. What is the probability that the sales will exceed 9 million pieces?

Since the probability that the price will lie between $c$ and $c+d c$ is 10 dc and the conditional probability of $\mathrm{v} \mid \mathrm{c}$ is as indicated above, the joint probability may be directly computed as

$$
f(v, c) d v d c=[10 \mathrm{dc}]\left[\mathrm{c}^{2} \mathrm{ve}^{-c v^{2}} \mathrm{dv}\right]
$$

From this expression the marginal distribution of $v$ can now be obtained by integrating over the variable c. Thus

$$
\left.g(v) d v=10 v d v \int_{.25}^{.35} c^{2} e^{-c v} d c=-10 v d v\left[\frac{e^{-c v}}{v^{3}}\left(v^{2} c^{2}+2 c v+2\right)\right]\right]_{.25}^{.35}
$$

The probability that the sales will exceed 9 million is thus given by the area underneath the marginal distribution curve from 9 to infinity as

$$
\begin{array}{r}
P(v>9)=10 \int_{9}^{\infty}\left[\frac{e^{-.25 v}}{v^{2}}\left[(.25)^{2} v^{2}+2(.25) v+2\right]-\frac{e^{-.35 v}}{v^{2}}\left[(.35)^{2} v^{2}\right.\right. \\
\\
+2(.35) v+2]] d v
\end{array}
$$

Example Particles are emanating from a radioactive source at an average rate of a impulses per unit of time. Since these pulses occur at random, they can be described by a Poisson distribution which, for a given time period, $t$, gives the following probability for exactly $x$ emanations:

$$
P(x \mid t)=\frac{(a t)^{x} e^{-a t}}{x!} \quad(x=0,1,2, \ldots, \infty)
$$

Let us assume that time is an equally likely variable over any range and therefore its density function can be represented by $g(t)=1 / T \quad 0<t<T$. Find the conditional distribution of $t$ for $a$ fixed value of $x$.

The joint probability that x emanations will be observed and the corresponding time $t^{\prime}$ falls between $t$ and $t+d t$ is

$$
f_{1}(x, t) d t=\frac{d t}{T} \quad \frac{(a t)^{x} e^{-a t}}{x!}
$$

Since this joint distribution is a function of $x$ and $t$, it is now possible to integrate $t$ over the appropriate limits and obtain the marginal distribution of $x$ (the marginal distribution of $t$, of course, is known) as

$$
f_{2}(x)=\frac{1}{T} \int_{0}^{T} \frac{(a t)^{x} e^{-a t} d t}{x!}
$$

We know, however, that the required conditional probability $w(t \mid x) d t$ can be obtained by dividing the joint probability $f_{1}(x, t) d t$ by the marginal distribution $f_{2}(x)$. Thus

$$
w(t \mid x) d t=\frac{(a t)^{x} e^{-a t} d t}{\int_{0}^{T}(a t)^{x} e^{-a t} d t}
$$

It is now possible to let $T \rightarrow \infty$, thus permitting the denominator of the above expression to be evaluated as a Gamma function. Thus the final probability that $t^{\prime}$ will lie between $t$ and $t+d t$ at the same time that $x$ (the number of emanations) assumes a given fixed value is given by

$$
w(t \mid x) d t=\frac{a(a t)^{x} e^{-a t} d t}{x!} \quad(0<t<\infty)
$$

which is the same answer as previously derived.
Example The number of orders $n$ per day for a certain chemical is Poisson distributed (parameter $\mu$ ), and the weight in lbs, per order is exponentially distributed (parameter $\beta$ ). Assuming that the weights from order to order are independent, find the distribution of total weight of orders per day.

Denote the weight of the $i^{\text {th }}$ order in a day by $y_{i}$ and the total weight of $n$ orders by $w$. Since $w \equiv 0$ if $n=0$, and the latter
event has a finite probability of occurrence $\left(e^{-\mu}\right)$, the distribution function of $w$ is discontinuous at zero. Thus, for $w=0$, we have,

$$
P(w=0)=P(n=0)=e^{-\mu}
$$

However, a density function exists for $w>0$.
Consider a fixed number of orders $n>0$. Then, if $\mathrm{w}=\mathrm{y}_{1}+\mathrm{y}_{2}+\ldots+\mathrm{y}_{\mathrm{n}}$, the conditional distribution of w becomes

$$
\phi(w \mid n)=\frac{1}{\beta^{n} \Gamma(n)} w^{n-1} e^{-w / \beta} \quad(0, \infty)
$$

Since $n$ is an integer, we may replace $\Gamma(n)$ by ( $n-1$ )! and write the joint distribution of $n$ and $w$ for $n>0$ as follows:

$$
\begin{aligned}
& f(n, w)=e^{-\mu} \frac{\mu^{n}}{n!} \frac{1}{\beta^{n}(n-1)!} w^{n-1} e^{-w / \beta} \\
&=\frac{\mu}{\beta} e^{-\mu} e^{-w / \beta} \frac{\left(\frac{\mu w}{\beta}\right)^{n-1}}{n!(n-1)!} \quad(n=1,2, \ldots, \infty ; \\
&0<w<\infty)
\end{aligned}
$$

Therefore, the marginal distribution of $w$ in the range $w>0$ becomes

$$
h(w)=\sum_{n=1}^{\infty} f(n, w)=\frac{\mu}{\beta} e^{-\mu} e^{-w / \beta} \sum_{k=0}^{\infty} \frac{\left(\frac{\mu w}{\beta}\right)^{k}}{k!(k+1)!} \quad(0<w<\infty)
$$

where $\mathrm{k}=\mathrm{n}$ - 1 .
We recognize this series as the modified Bessel function of the first kind of order unity. The latter function (symbol $\left.I_{1}(u)\right)$ may be expressed as follows:

$$
I_{1}(u)=\frac{u}{2} \sum_{k=0}^{\infty} \frac{\left(\frac{u}{2}\right)^{2 k}}{k!(k+1)!}
$$

Substituting $\frac{\mu w}{\beta}=\left(\frac{u}{2}\right)^{2}$, we thus obtain

$$
h(w)=e^{-\mu} \frac{\sqrt{\mu}}{\sqrt{\beta}} w^{-1 / 2} e^{-w / \beta} I_{1}\left(2 \sqrt{\frac{\mu w}{\beta}}\right) \quad(0<w<\infty)
$$

and this density function, together with the equation $P(w=0)=e^{-\mu}$ defines the distribution of $w$.

In the more general case in which the weight per order is a Gamma variate (parameters $a, \beta$ ) we would still have $P(w=0)=e^{-\mu}$, while for $n>0$ the conditional distribution of $w$ would be given by

$$
\phi(w \mid n)=\frac{1}{\beta^{n a+n} \Gamma(n a+n)} w^{n a+n-1} e^{-w / \beta} \quad(0<w<\infty)
$$

and the marginal density function would be

$$
h(w)=e^{-\mu} e^{-w / \beta} \sum_{n=1}^{\infty} \frac{\mu^{n} w^{n a+n-1}}{n!\beta^{n a+n} \Gamma(n a+n)} \quad(0<w<\infty)
$$

In this case, however, the series expansion for the marginal density does not reduce to a Bessel function.

## PROBLEMS FOR CHAPTER 1

1. The random variables x and y are jointly distributed as follows:

$$
f(x, y)=120 x(y-x)(1-y), \quad 0 \leq x \leq y ; \quad 0 \leq y \leq 1
$$

(a) Find the conditional distribution of x for a fixed value of y .
(b) Derive the distribution of the quotient $u=x / y$.
2. A noise source has an amplitude of $v$ volts which has a probability density $f(v)=e^{-v}, v \geq 0$. The noise is fed into a circuit which subtracts one volt and squares the remainder. What is the distribution of the output?
3. A manufacturer wishes to estimate the mean $\mu$ of a certain attribute x of his product by computing the sample mean $\overline{\mathrm{x}}$ of a set of $n$ independent observations of $x$. The cost $C$ of the sampling process is composed of one part proportional to the sample size and one part proportional to the magnitude of the error of the estimate:

$$
c=\sqrt{\frac{2}{\pi}} n+16|\bar{x}-\mu|
$$

If $x$ is assumed to be normally distributed with unit variance, find the value of $n$ which minimizes the expected value of the cost.
4. In the joint distribution of two independent unit normal variates $\mathrm{x}, \mathrm{y}$ show the probability that a random point will fall within the square enclosed by the lines $x=-a, x=a, y=-a, y=a$ is less than the probability that it will fall within the circle of equal area defined by $x^{2}+y^{2}=4 a^{2} / \pi$. With this fact in mind, prove the inequality

$$
\frac{1}{\sqrt{2 \pi}} \int_{-a}^{a} e^{-x^{2} / 2} d x \leq \sqrt{1-e^{-2 a^{2} / \pi}}
$$

5. An owner of five overnight cabins is considering having TV sets to rent to cabin occupants. He finds that about a half of his customers would be willing to rent a set, on the average. If he buys three sets, what fraction of the evenings (when all cabins are occupied) will there be more requests than TV sets? What rental must he charge to make it worth while to have three sets? A count of 100 evenings shows that the number of evenings when $m$ cabins were occupied were:

$$
\begin{aligned}
N(m)=\frac{\text { Number of evenings (in }}{} 100 \text { ) when } m \text { cabins were occupied } \\
\hline m=\begin{array}{llllllll}
0 & 1 & 2 & 3 & 4 & 5 \\
N(m) & =1 & 3 & 8 & 14 & 18 & 56
\end{array}
\end{aligned}
$$

What fraction of all evenings will three TV sets satisfy requests?
How often will the third set be needed? Decide first whether occupant arrivals are random and what the "average demand" for cabins is at this location. Estimate the fraction of time a sixth cabin would be occupied, if a sixth were built.
6. In a building with $n$ lamps lighted the same time per day, a bulb is replaced immediately when it burns out. The costs for replacing one bulb consis of purchasing cost $a_{1}$, the cost $a_{2}$ of changing the bulb (i.e. taking out the old and putting a new one in the lamp) and all the costs $a_{3}$ for bringing the bulb, ladder, and other necessary things to the required place. Another alternative is to change all the bulbs at the same time periodically with time period $T$, but still change a bulb when it burns out. When all the bulbs are to be changed together, the costs $a_{1}$ and $a_{2}$ are the same per bulb, but instead of $a_{3}$, we now have a total cost $a_{4}$ for bringing all the bulbs and other materials to the place. Determine $T$ such that the average cost per unit of time the bulbs are lit becomes a minimum. The lifetime of a bulb is unknown, but we shall assume that its probability density is

$$
f(t)=\frac{1}{K}(0<t<K)
$$

7. A man has a string of $m$ Christmas tree lights on a circular cord. They have the property that if one bulb burns out, the entire string fails. In his search for the faulty bulb, the man is twice as likely to test the bulbs successively around the circle as he is to try them at random, forgetting after each trial which bulb he had tested. Given that he finds the defective bulb on the second trial, what is the probability that he in fact used the systematic (cyclic) procedure?
8. Three separate cages are arranged in a circle with connections between them. A mouse is free to move from one cage to another in either direction, but the average time that he stays in any one cage is different from the others. At time $t=0$ he is observed in a particular cage and obviously as $t$ increases the probability that he is in this cage approaches the steady-state probability. What is the relationship that exists between the various rates of leaving and arriving from one cage to another, so that the probability at time $t$ that he is in the same cage is on the borderline between a pure exponential decay and a damped oscillation as the type of approach to this steady-state probability?

[^0]:    * cf. Fry, Probability and its Engineering Uses, Van Nostrand, 1928, p. 233.

[^1]:    * For a consideration of the case of the continuous density function, see A. J. Lotka, Théorie Analytique des Associations Biologiques, Vol. 2, Actualités Scientifiques et Industrielles, No. 780 (1939), Hermann et Cie, Paris.

[^2]:    * Proc. Nat. Acad. Sci. U.S. A., 36, 202-207 (1950)

