SMAE21-

Mathematical Statistics

Mathematical Statistics

Syllabus

Unit I

The probability set function – Random variables – Probability density function – Distribution function – Mathematical expectation – Special mathematical expectations – Chebyshev's inequality.

Unit II

Conditional [robability – Marginal and conditional distribution – Stochastic independence – Some special distributions – The poisson distribution.

Unit III

The gamma and chi – square distributions – The normal distribution – The bivariate normal distribution – Distributions of function of random variables – Sampling theory – Transformations of variables of the discrete type – Transformations of variables of the continuous type.

Unit IV

The β, t and F distributions – Distributions of order statistics – The moment generating function technique – The distribution of $\bar{X}$ and ns^2/σ^2 - Expectations of functions of random variables

Unit V

Limiting distributions – Stochastic convergence – Limiting moment generating function – The central limit theorem – Some theotems on limiting distributions.

Contents

4.5 Expectation of functions of random variables

Unit V

5.1 Limiting Distribution

5.2 Stochastic Convergence

5.3 Limiting Moment Generating Function

5.4 Central Limit Theorem

5.5 Some Theorems on Limiting Distreibutions

Mathematical Statistics

Unit I

DISTRIBUTIONS OF RANDOM VARIABLES

1.1 Introduction:

Many kinds of investigations may be characterized in part by the fact that repeated experimentation, under essentially the same conditions, is more or less standard procedure. For instance, in medical research, interest may center on the effect of a drug that is to be administered; or an economist may be concerned with the prices of three specified commodities at various time intervals; Or an agronomist may wish to study the effect that a chemical fertilizer has on the yield of a cereal grain. The only way in which an investigator can elicit information about any such phenomenon is to perform the experiment. Each experiment terminates with an outcome. But it is characteristic of these experiments that the outcome cannot be predicted with certainty prior to the experiment. Suppose that we have such an experiment, but the experiment is of such a nature that a collection of every possible outcome can be described prior to its performance. If this kind of experiment can be repeated under the same conditions, it is called a **random experiment**, and the collection of every possible outcome is called the experimental space or the **sample space**. We denote the sample space by C.

Definition 1.1.1

An experiment is called a *random experiment* if all possible outcomes of an experiment can be described and determined to the performance of the experiment.

Definition 1.1.2

The collection of all possible outcomes of a random experiment is called *sample space*. It is denoted by C.

Example 1: In the toss of a coin, let the outcome tails be denoted by T and let the outcome heads be denoted by H. If we assume that the coin may be repeatedly tossed under the same conditions, then the toss of this coin is an example of a random experiment in which the outcome is one of the two symbols T or H; that is, the sample space is the collection of these two symbols.
For this example, then C = {H, T}.

Example 2: In the cast of one red die and one white die, let the outcome be the ordered pair (number of spots up on the red die, number of spots up on the white die). If we assume that these two dice may be repeatedly cast under the same conditions, then the cast of this pair of dice is a random experiment. The sample space consists of the 36 ordered pairs:
C = {(1, 1), . . .,(1, 6), (2, 1), . . .,(2, 6), . . .,(6, 6)}.

We generally use small Roman letters for the elements of C such as a, b, or c. Often for an experiment, we are interested in the chances of certain subsets of elements of the sample space occurring. Subsets of C are often called **events** and

are generally denoted by capital Roman letters such as A, B, or C. If the experiment results in an element in an event A, we say the event A has occurred. We are interested in the chances that an event occurs. For instance, in Example 1.1.1 we may be interested in the chances of getting heads; i.e., the chances of the event A = {H} occurring. In the second example, we may be interested in the occurrence of the sum of the up faces of the dice being "7" or "11;" that is, in the occurrence of the event
A = {(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1), (5, 6), (6, 5)}.
Now conceive of our having made N repeated performances of the random experiment.

Then we can count the number of times (the **frequency**) that the event A actually occurred throughout the N performances. The ratio f/N is called the **relative frequency** of the event A in these N experiments. A relative frequency is usually quite erratic for small values of N, as you can discover by tossing a coin.

But as N increases, experience indicates that we associate with the event A a number, say p, that is equal or approximately equal to that number about which the relative frequency seems to stabilize. If we do this, then the number p can be interpreted as that number which, in future performances of the experiment, the relative frequency of the event A will either equal or approximate. Thus, although we cannot predict the outcome of a random experiment, we can, for a large value of N, predict approximately the relative frequency with which the outcome will be in A. The number p associated with the event A is given various names. Sometimes it is called the probability that the outcome of the random experiment is in A; sometimes it is called the probability of the event A; and sometimes it is called the probability measure of A. The context usually suggests an appropriate choice of terminology.

Example 3: Let C denote the sample space of Example 1.1.2 and let B be the collection of every ordered pair of C for which the sum of the pair is equal to seven.

Thus B = {(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1)}.

Suppose that the dice are cast N = 400 times and let f denote the frequency of a sum of seven. Suppose that 400 casts result in f = 60.

Then the relative frequency with which the outcome was in B is f(N)= 60/400 = 0.15. Thus we might associate with B a number p that is close to 0.15, and p would be called the probability of the event B.

EXERCISES

1.1. In each of the following random experiments, describe the sample space s. Use any experience that you may have had (or use your intuition) to assign a value to the probability *p* of the event C in each of the following instances:

(a) The toss of an unbiased coin where the event C is tails.

(b) The cast of an honest die where the event C is a five or a six.

(c) The draw of a card from an ordinary deck of playing cards where the event C occurs if the card is a spade.

(d) The choice of a number on the interval zero to 1 where the event C occurs if the number is less than t.

(e) The choice of a point from the interior of a square with opposite vertices (-1, -1) and (1, 1) where the event C occurs if the sum of the coordinates of the point is less than 1-.

1.2. A point is to be chosen in a haphazard fashion from the interior of a fixed circle. Assign a probability *p* that the point will be inside another circle, which has a radius of one-half the first circle and which lies entirely within the first circle.

1.3. An unbiased coin is to be tossed twice. Assign a probability p_2 to the event that the first toss will be a head and that the second toss will be a tail.

1.2. Algebra of Sets:

The concept of a *set* or a *collection* of objects is usually left undefined.However, a particular set can be described so that there is no misunderstandingas to what collection of objects is under consideration. Forexample, the set of the first 10 positive integers is sufficiently welldescribed to make clear that the numbers -i and 14 are not in the set,while the number 3 is in the set. If an object belongs to a set, it is saidto be an *element* of the set. For example, if *A* denotes the set of real numbers *x* for which $0 \leq x \leq 1$, then $\frac{3}{4}$ is an element of the set *A*. Thefact that $\frac{3}{4}$is an element of the set *A* is indicated by writing $\frac{3}{4} \in A$.

More generally, $a \in A$ means that *a* is an element of the set *A*.
The sets that concern us will frequently be *sets of numbers.*

We say a set C is **countable** if C is finite or has as many elements as there are positive integers.
For example, the sets C1 = {1, 2, . . . , 100} and C2 = {1, 3, 5, 7, . . .}

Definition 1.2.1: If each element of a set A_1 is also an element of setA_2, the set A_1 is called a *subset* of the set A_2. This is indicated by writing $A_1 \subseteq A_2$.

If $A_1 \subseteq A_2$ and $A_2 \subseteq A_1$ then the two sets have the sameelements, and this is indicated by writing $A_1 = A_2$.

***Example* 1**. Let $A_1 = \{x;\ 0 \leq x \leq 1\}$ and $A_2 = \{x;\ -1 \leq x \leq 2\}$. Here the one-dimensional set $A_1$is seen to be a subset of the one-dimensional set

A_2 ; that is $A_1 \subseteq A_2$. Subsequently, when the dimensionality of the set is clear, we shall not make specific reference to it.

***Example* 2**. Let $A_1 = \{(x,y); 0 \leq x = y \leq 1$ and $A_2 = \{(x,y); 0 \leq x \leq 1, 0 \leq y \leq 1\}$. Since the elements of A_1 are the points on one diagonal of the square, then $A_1 \subseteq A_2$.

Definition 1.2.2: If a set A has no elements, A is called the *null set.* This is indicated by writing $A = \emptyset$.

Definition 1.2.3: The set of all elements that belong to at least one of the sets A_1 and A_2 is called the *union* of A_1 and A_2. The union of A_1 and A_2 is indicated by writing $A_1 \cup A_2$. The union of several sets $A_1, A_2, A_3, \ldots$ is the set of all elements that belong to at least one of the several sets. This union is denoted by $A_1 \cup A_2 \cup A_3 \cup \ldots$ or by $A_1 \cup A_2 \cup \ldots \cup A_k$ if a finite number k of sets is involved.

***Example* 3**: Let $A_1 = \{x; x = 0, 1, \ldots, 10\}$ and $A_2 = \{x, x = 8, 9, 10, 11,$ or $11 < x \leq 12\}$. Then $A_1 \cup A_2 = \{x; x = 0, 1, \ldots, 8, 9, 10, 11,$ or $11 < x \leq 12\} = \{x; x = 0, 1, \ldots, 8, 9, 10,$ or $11 \leq x \leq 12\}$.

***Example* 4**. Let A_1 and A_2 be defined as in Example 1. Then $A_1 \cup A_2 = A_2$ •

***Example* 5**: Let $A_2 = \emptyset$. Then $A_1 \cup A_2 = A_1$ for every set A_1.

***Example* 6**: For every set A, $A \cup A = A$.

***Example* 7**: Let $A_k = \{x; 1/(k+1) \leq x \leq 1\}$, $k = 1, 2, 3, \ldots$. Then $A_1 \cup A_2 \cup A_3 \cup \ldots = \{x, 0 < x \leq 1\}$. Note that the number zero is not in this set, since it is not in one of the sets $A_1, A_2, A_3, \ldots$

Definition 1.2.4: The set of all elements that belong to each of the sets A_1 and A_2 is called the *intersection* of A_1 and A_2, The intersection of A_1 and A_2 is indicated by writing $A_1 \cap A_2$. The intersection of several sets $A_1, A_2, A_3, \ldots$ is the set of all elements that belong to each of the sets $A_1, A_2, A_3, \ldots$ " This intersection is denoted by $A_1 \cap A_2 \cap A_3 \cap \ldots$ or by $A_1 \cap A_2 \cap A_3 \cap \ldots \cap A_k$ if a finite number k of sets is involved.

***Example* 8**: Let $A_1 = \{(x, y); (x, y) = (0, 0), (0, 1), (1, 1)\}$ and $A_2 = \{(x, y); (x, y) = (1,1), (1,2), (2, 1)\}$. Then $A_1 \cap A_2 = \{(x, y); (x, y) = (1, 1)\}$.

Definition 1.2.5: The **complement** of an event A is the set of all elements in C which are not in A. We denote the complement of A by A^c. That is, $A^c = \{x \in C / x \notin A\}$

Two events are **disjoint** if they have no elements in common. More formally we define

Definition 1.2.6: Let A and B be events. Then A and B are **disjoint** if $A \cap B = \varphi$
If A and B are disjoint, then we say $A \cup B$ forms a **disjoint union.** The next two examples illustrate these concepts.

Example 9: Suppose we have a spinner with the numbers 1 through 10 on it. The experiment is to spin the spinner and record the number spun. Then $C = \{1, 2, \ldots, 10\}$. Define the events A, B, and C by $A = \{1, 2\}$, $B = \{2, 3, 4\}$, and
$C = \{3, 4, 5, 6\}$, respectively.
$A^c = \{3, 4, \ldots, 10\}$; $A \cup B = \{1, 2, 3, 4\}$; $A \cap B = \{2\}$
$A \cap C = \varphi$; $B \cap C = \{3, 4\}$; $B \cap C \subset B$; $B \cap C \subset C$
$A \cup (B \cap C) = \{1, 2\} \cup \{3, 4\} = \{1, 2, 3, 4\}$
$(A \cup B) \cap (A \cup C) = \{1, 2, 3, 4\} \cap \{1, 2, 3, 4, 5, 6\} = \{1, 2, 3, 4\}$
The reader should verify these results.

Example 10: For this example, suppose the experiment is to select a real number in the open interval (0, 5); hence, the sample space is $C = (0, 5)$. Let $A = (1, 3)$, $B = (2, 4)$, and $C = [3, 4.5)$.
$A \cup B = (1, 4)$; $A \cap B = (2, 3)$; $B \cap C = [3, 4)$
$A \cap (B \cup C) = (1, 3) \cap (2, 4.5) = (2, 3)$ (1.2.3)
$(A \cap B) \cup (A \cap C) = (2, 3) \cup \varphi = (2, 3)$ (1.2.4)
A sketch of the real number line between 0 and 5 helps to verify these results.

Distributive Laws:
For any sets A, B, and C,
$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$
$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$.
These follow directly from set theory.

The next two identities are collectively known as **DeMorgan's Laws**. For any sets A and B,
$(A \cap B)^c = A^c \cup B^c$
$(A \cup B)^c = A^c \cap B^c$.

For instance, in Example 9,
$(A \cup B)^c = \{1, 2, 3, 4\}^c = \{5, 6, \ldots, 10\} = \{3, 4, \ldots, 10\} \cap \{\{1, 5, 6, \ldots, 10\} = A^c \cap B^c$;
while, from Example 10,
$(A \cap B)^c = (2, 3)^c = (0, 2] \cup [3, 5) = [(0, 1] \cup [3, 5)] \cup [(0, 2] \cup [4, 5)] = A^c \cup B^c$.
As the last expression suggests, it is easy to extend unions and intersections to morethan two sets. If $A_1, A_2, \ldots, A_n$ are any sets, we define
$A_1 \cup A_2 \cup \cdots \cup A_n = \{x : x \in A_i$, for some $i = 1, 2, \ldots, n\}$
$A_1 \cap A_2 \cap \cdots \cap A_n = \{x : x \in A_i$, for all $i = 1, 2, \ldots, n\}$.

1.3: Set Functions.

We havefunctions that can be evaluated, not necessarily at a point, but for anentire set of points. Such functions are naturally called functions of aset or, more simply, *set functions.* We shall give some examples of setfunctions and evaluate them for certain simple sets.

***Example* 1**. Let A be a set in one-dimensional space and let $Q(A)$ be equal to the number of points in A which correspond to positive integers. Then $Q(A)$ is a function of the set A. Thus, if $A = \{x;\ 0 < x < 5\}$, then $Q(A) = 4$; if $A = \{x;\ x = -2, -1\}$, *then* $Q(A) = O$; *if* $A = \{x;\ 0 < x < 6\}$, then $Q(A) = 5$.

***Example* 2**. Let A be a set in two-dimensional space and let $Q(A)$ be the area of A, if A has a finite area; otherwise, let $Q(A)$ be undefined. Thus, if $A = \{(x, y);\ x^2 + y^2 \leq 1\}$, then $Q(A) = \pi$; if $A = \{(x, y);\ (x, y) = (0,0), (1,1), (0, 1)\}$, then $Q(A) = 0$;

***Example* 3**. Let A be a set in three-dimensional space and let $Q(A)$ be the volume of A, if A has a finite volume; otherwise, let $Q(A)$ be undefined. Thus, if $A = \{(x, y, z);\ 0 \leq x \leq 2, 0 \leq y \leq 1, 0 \leq z \leq 3\}$, then $Q(A) = 6$; if $A = \{(x, y, z);\ x^2 + y^2 + Z^2 \geq 1\}$, then $Q(A)$ is undefined.

At this point we introduce the following notations. The symbol $\int_A f(x)$ will mean the ordinary (Riemann) integral of $f(x)$ over a prescribed one-dimensional set A; *the symbol* $\iint_A g(x,y)dxdy$ will mean the Riemann integral of $g(x, y)$ over a prescribed twodimensionalset A; and so on. To be sure, unless these sets A and thesefunctions $f(x)$ and $g(x, y)$ are chosen with care, the integrals willfrequently fail to exist. Similarly, the symbol $\sum_A f(x)$ will mean the sum extended over all $x \in A$; the symbol $\sum_A \sum g(x,y)$ will mean the sum extended over all $(x, y) \in A$; and so on.

1.4 Probability set functions:

Definition 1.4.1: Anysubset of a sample space C is called a event. The event S is called sure event and the event which is empty is called an impossible event.

Definition 1.4.2: Let S be an event sample space. The set P is called the probability set function if it satisfies the following condition.

(i) $P(A) \geq 0$.

(ii) $P(S) = 1$

(iii) $P(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i)$

Where $\{A_n\}$ is any finite sequence of disjoint events. The number P(A) is called the probability of the event a and $P(A) = \frac{n(A)}{n(S)}$.

Definition1.4.3: Let S be the sample space and $A \subseteq S$. Then
(i) $P(\bar{A}) = 1 - P(A)$ where $\bar{A}$ is the complement of A.
(ii) $P(\emptyset) = 0$.
(iii) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ where A and B are events.

Definition1.4.4: Let S be a sample space due to the random experiment. The function X : S →R which assign to each element in S one and only real number is called the random variable.

Definition1.4.5: **Distribution Function**. Let X be a random variable. Then the function F : R →R defined by F(X) = P(X ≤ x) where $-\infty < x < \infty$ is called a distribution function of random variable X.

Definition1.4.6:**Discrete random variable**. If the random variable X takes atmost a countable number of values X_1, X_2, X_3, …, X_n then X is called as a Discrete random variable.

1.5 Probability Density Function.

Definition 1.5.1: Probability Density Function. Let X be a discrete random variable and$P(X = x) = p_i = f(x_i)$. Then the distribution functions of a random variable is defined by $F(X) = P(x_i \leq x) = \sum_{x_i \leq x} f(x_i)$. Here F is called probability density function of the discrete random variable.

Definition1.5.2: Continuous Random Variable. The random variable X is said to be continuous random variable if it can take any value in an interval.

Definition1.5.3: Probability Density Function of continuous random variable. Let X be a continuous random variable taking the values in the interval $(-\infty, \infty)$. Let f(x) be a function such that $f(x) \geq 0\, ; x \in (-\infty, \infty)$ and $\int_{-\infty}^{\infty} f(x)dx = 1$. Then $f(x)$ is called a probability density function.

Definition1.5.4:Distribution functions of the continuous random variable. Let X be a continuous random variable with p.d.f$f(x)$. Define $F(x)$
$F(x)$ is called the distribution function of the continuous random variable of X.

Theorem 1. *For each* C∈ S, *P(C)* = 1 - *P(C*).*
Proof. We have S = C ∪C* and C ∩C* = ∅.

$$P(S) = P(C) + P(C^*)$$

1 = *P(C)* + *P(C*),*
which is the desired result.

Theorem 2. *The probability of the null set* is *zero; that is, P(∅)* = O.
Proof. In Theorem 1, take C = ∅ so that C* =S. Accordingly, wehave
P(∅) = 1 - *P(S)* = 1 - 1 = 0,
and the theorem is proved.

Theorem 3. *If* C_1 *and* C_2 *are subsets of' S such that* $C_1 \subseteq C_2$ *, then* $P(C_1) \leq P(C_2)$ ·

Proof. Now $C_2 = C_1 \cup (C_1^* \cap C_2)$ and $C_1 \cap (C_1^* \cap C_2) = \emptyset$

Hence, $P(C_2) = P(C_1) + P(C_1^* \cap C_2)$

However, $P(C_1^* \cap C_2) \geq 0$ accordingly,

$P(C_2) \geq P(C_1)$ ·

Theorem 4. *For each* $C \in S$ $0 \leq P(C) \leq 1$.

Proof: Since $\emptyset \subseteq C \subseteq S$ using theorem 3

$P(\emptyset) \leq P(C) \leq P(S)$

Hence $0 \leq P(C) \leq 1$

Theorem 5. If C1 and C2 are subsets of S then

$$P(C_1 \cup C_2) = P(C_1) + P(C_2) - P(C_1 \cap C_2)$$

Proof: Each of the sets C_1 U C_2 and C2 can be represented, respectively, as a union of nonintersecting sets as follows:

$C_1 \cup C_2 = C_1 \cup (C_1^* \cap C2)$ and $C_2 = (C_1 \cap C_2) \cup (C_1^* \cap C_2)$

Thus, from (iii) of Definition 1.4.2,

$$P(C_1 \cup C_2) = P(C_1) + P(C_1^* \cap C_2)$$

And$P(C_2) = P(C_1 \cap C_2) + P(C_1^* \cap C_2)$

If the second of these equations is solved $P(C_1^* \cap C_2)$ for and this result substituted in the first equation, we obtain

$$P(C_1 \cup C_2) = P(C_1) + P(C_2) - P(C_1 \cap C_2)$$

This completes the proof.

Example **1**. Let S denote the sample space of Example 2 of Section 1.1.Let the probability set function assign a probability of 1-6 to each of the 36points in S. If $C_1 = \{C;\ C = (1, 1), (2, 1), (3, 1), (4, 1), (5,1)\}$ and $C2 = \{C;\ C = (1,2), (2, 2), (3, 2)\}$, then

$P(C_1) = \frac{4}{36}$, $P(C2) = \frac{3}{36}$, $P(C_1 \cup C_2) = \frac{7}{36}$ and $P(C_1 \cap C_2) = 0$.

Example **2**. Two coins are to be tossed and the outcome is the orderedpair (face on the first coin, face on the second coin). Thus the sample spacemay be represented as $S = \{C;\ C = (H, H), (H, T), (T, H), (T, T)\}$. Let theprobability set function assign a probability of $\frac{1}{4}$ to each element of S.

Let $C_1 = \{C;\ C = (H, H), (H, T)\}$ and $C_2 = \{C;\ C = (H, H), (T, H)\}$.

Then$P(C_1) = P(C_2) = \frac{2}{4}$, $P(C_1 \cap C_2) = \frac{1}{4}$, and, in accordance with Theorem 5,

$$P(C_1 \cup C_2) = \frac{2}{4} + \frac{2}{4}\frac{1}{4} = \frac{3}{4}$$

Example **3**. Let X be a random variable of the discrete type with space

A = *{x; x* = 0, 1,2, 3, 4}. Let$P(A) = \sum_{x\in A} f(x)$ where $f(x) = \frac{4!}{x!(4-x)!}\left(\frac{1}{2}\right)^4$ where $x \in A$
and, as usual, 0! = 1. Then if *A* = *{x; x* = 0, 1}, we have

$$P(A) = \frac{4!}{0!\,(4)!}\left(\frac{1}{2}\right)^4 + \frac{4!}{1!\,(3)!}\left(\frac{1}{2}\right)^4 = \frac{1}{16} + \frac{4}{16} = \frac{5}{16}$$

***Example* 4**. Let X be a random variable of the discrete type with space *A*= *{x; x* = 1, 2, 3, ... }, and let *f(x)* =$\left(\frac{1}{2}\right)^x$, *x*∈ *A*Then Pr (x∈*A)* = $\sum f(x)$
If *A* = *{x; x* = 1, 3, 5, 7, ...}, we have $P(A) = \left(\frac{1}{2}\right)^1 + \left(\frac{1}{2}\right)^3 + \left(\frac{1}{2}\right)^5 + \cdots = \frac{2}{3}$.

(b) *The continuous type of random variable.* Let the one-dimensionalset be such that the Riemann integral $\int f(x)dx = 1$where (1) *f(x)* > 0, $x \in A$, and (2) *f(x)* has at most a finite number of discontinuities in every finite interval that is a subset of *A*. If *S*is the space of the random variable X and if the probability set function *P(A), A* ⊆ Scan be expressed in terms of such an f(x) by
P(A) = Pr (X∈ *A)* = $\int_A f(x)dx$ = 1, then X is said to be a random variable of the *continuous type* and to have a distribution of that type.

***Example* 1**. Let the space *S*= *{x;* 0 <*x* <∞}, and let *f(x)* =$e^{-x}, x \in A$,
If X is a random variable of the continuous type so that Pr (X ∈*A)* =$e^{-x}dx$,
we have, with *A* = *{x;* 0 <*x* < I},
Pr (X ∈*A)* = $\int_0^1 e^{-x}dx = 1 - e^{-1}$.

-

Note that Pr (X ∈*A)* is the area under the graph of *f(x)* = e^{-x}, which lies above the z-axis and between the vertical lines *x* = 0 and *x* = 1.

***Example* 2**. Let X be a random variable of the continuous type with space *S*= *{x;* 0 <*x* <1}. Let the probability set function be *P(A)* =$\int f(x)dx$,
where$f(x) = cx^2$.
Since *P(A)* is a probability set function, *P(S)* = 1. Hence the constant *c* is determined by $\int_0^1 cx^2dx = 1$. c = 3

It is seen that whether the random variable X is of the discrete type or of the continuous type, the probability Pr (X ∈*A)* is completely determined by a f*unctionf(x)*. In either *casef(x)* is called the *probability density function* (hereafter abbreviated p.d.f.) of the random variable X.
If we restrict ourselves to random variables of either the discrete type or the continuous type, we may work exclusively with the p.d.f.*f(x).*
If *f(x)* is the p.d.f. of a continuous type of random variable X and if
A is the set *{x; a* <*x* <*b},* then *P(A)* = Pr (X ∈*A)* can be written as

$$P(a < x < b) = \int_a^b f(x)dx$$

Moreover, if $A = \{x; x = a\}$, then $P(A) = \Pr(X \in A)$ can be written as

$$P(x = a) = \int_a^a f(x)dx = 0$$

The probability of every set consisting of a single point is zero. This fact enables us to write, say, $\Pr(a < X < b) = \Pr(a \leq X \leq b)$.

***Example* 5**. Let the random variable X have the p.d.f $f(x) = \begin{cases} 2x & 0 < x < 1 \\ 0 & otherwise \end{cases}$

Find $\Pr\left(\frac{1}{2} \leq x \leq \frac{3}{4}\right)$ and $\Pr\left(-\frac{1}{2} \leq x \leq \frac{1}{2}\right)$

Solution: First $\Pr\left(\frac{1}{2} \leq x \leq \frac{3}{4}\right) = \int_{\frac{1}{2}}^{\frac{3}{4}} 2xdx = \left(2\frac{x^2}{2}\right)_{\frac{1}{2}}^{\frac{3}{4}}$

$$= \left(\frac{3}{4}\right)^2 - \left(\frac{1}{2}\right)^2 = \frac{9}{16} - \frac{1}{4} = \frac{9-4}{16} = \frac{5}{16}$$

Next, $\Pr\left(-\frac{1}{2} \leq x \leq \frac{1}{2}\right) = \int_{-\frac{1}{2}}^{0} 0dx + \int_0^{\frac{1}{2}} 2xdx$

$$= 0 + \left(\frac{1}{2}\right)^2 = \frac{1}{4}$$

***Example* 6.** Let $f(x,y) = \begin{cases} 6x^2y & 0 \leq x \leq 1, 0 \leq y \leq 1 \\ 0 & otherwise \end{cases}$

Then, $\Pr\left(0 < x < \frac{3}{4}, \frac{1}{3} < y < 2\right) = \int_0^{\frac{3}{4}} \int_{\frac{1}{3}}^{2} f(x,y)dxdy$

$$= \int_{\frac{1}{3}}^{1} \int_0^{\frac{3}{4}} 6x^2ydxdy + \int_1^2 \int_0^{\frac{3}{4}} 0dxdy$$

$$= \frac{3}{8} + 0 = \frac{3}{8}$$

Note that this probability is the volume under the surface $f(x, y) = 6x^2y$ and above the rectangular set $\{(x, y); 0 < x < 3/4, 1/3 < y < 1\}$ in the xy-plane.

Exercises

1.4.1. For each of the following, find the constant c so that $f(x)$ satisfies the conditions of being a p.d.f. of one random variable X.

(a) $f(x) = c\left(\frac{2}{3}\right)^x x$, $x = 1, 2, 3, \ldots$, zero elsewhere.

(b) $f(x) = cxe^{-x}$, $0 < x <$ (X), zero elsewhere.

1.4.2. Let $f(x) = x/15$, $x = 1,2,3,4,5$, zero elsewhere, be the p.d.f. of X.

Find Pr (X = 1 or 2), Pr (1/2< X <5/2), and Pr (1 ≤*X*≤ 2).

1.6. The Distribution Function

Let the random variable *X* have the probability set function *P(A),*where *A* is a one-dimensional set. Take *x* to be a real number and considerthe set *A* which is an unbounded set from -∞ to *x,* including the point *x* itself. For all such sets *A* we have *P(A)* = Pr*(X* **ϵ***A)* =Pr*(X* ≤ *x).* This probability depends on the point *x;* that is, this probability is a function of the point *x.* This point function is denoted by the symbol *F(x)* = Pr*(X* ≤*x).* The function *F(x)* is called the *distribution function* (sometimes, *cumulative distribution function)* of the random variable *X.* Since *F(x)* = Pr*(X* ≤*x),* then, with *f(x)the p.d.f we have* $F(x) = \sum_{w \le x} f(w)$ for the discrete type of random variable, and $F(x) = \int_{-\infty}^{x} f(w)dw$ for the continuous random variable.

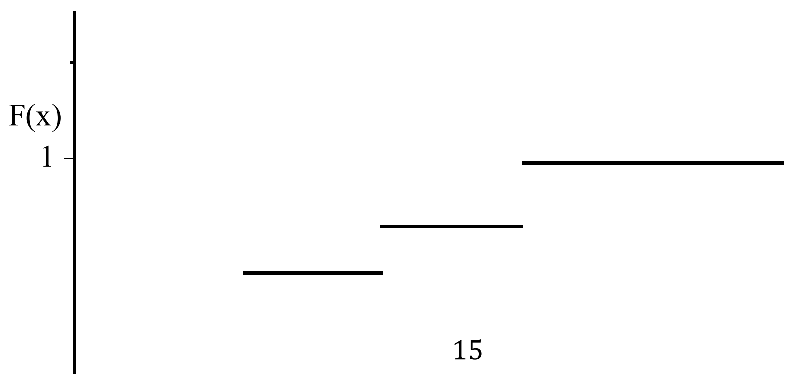
Remark. If X is a random variable of the continuous type, the p.d.f.*f(x)* has at most a finite number of discontinuities in every finite interval. This means (1) that the distribution function *F(x)* is everywhere continuous and (2) that the derivative of *F(x)* with respect to *x* exists and is equal to *f(x)* at each point of continuity *of f(x).* That is, *F'(x)* = *f(x)* at each point of continuity of *f(x).* If the random variable X is of the discrete type, most surely the p.d.f.*f(x)* is *not* the derivative of *F(x)* with respect to *x* (that is, with respect to Lebesgue measure); but f(x) is the(Radon-Nikodym) derivative of *F(x)* with respect to a counting measure. A derivative is often called a *density.* Accordingly, we call these derivatives *probability density functions.*

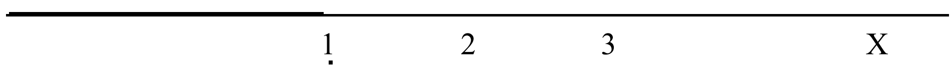
***Example* 1**. Let the random variable X of the discrete type have the p.d.f

f(x) = *x/6, x* = 1,2,3, zero elsewhere.

The distribution function of X is

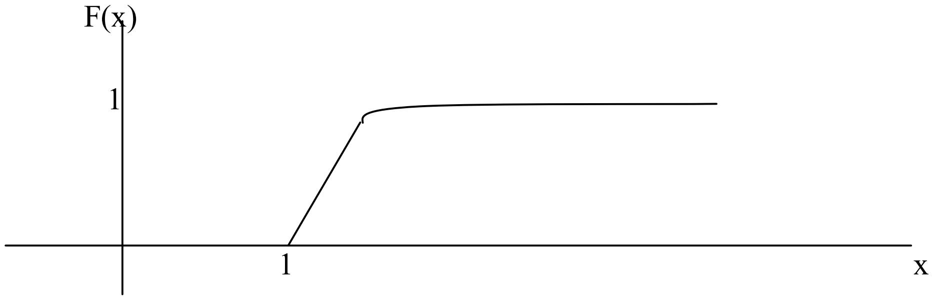
$$F(x) = \begin{cases} 0 & x < 1 \\ \frac{1}{6} & 1 \le x < 2 \\ \frac{3}{6} & 2 \le x < 3 \\ 1 & 3 \le x \end{cases}$$





***Example* 2**. Let the random variable X of the continuous type have thep.d.f. $f(x) = 2/x^3$, $1 < x < \infty$), zero elsewhere. The distribution function F(x) is

$$F(x) = \begin{cases} \int_{-\infty}^{x} 0\,dw = 0 & x < 1 \\ \int_{1}^{x} \frac{2}{w^2}\,dw = 1 - \frac{1}{x^2} & 1 \le x. \end{cases}$$



The graph of this distribution function is depicted in the above figure. Here *F(x)*is a continuous function for all real numbers *x;* in particular, *F(x)* is everywherecontinuous to the right. Moreover, the derivative of *F(x)* with respectto *x* exists at all points except at $x = 1$. Thus the p.d.f. of X is defined by thisderivative except at $x = 1$. Since the set $A = \{x; x = 1\}$ is a set of probabilitymeasure zero [that is, *P(A)* = OJ, we are free to define the p.d.f. at $x = 1$ inany manner we please. One way to do this is to write $f(x) = 2/x^3$, $1 < x < \infty$ and 0 elsewhere. There are several properties of a distribution function *F(x)* thatcan be listed as a consequence of the properties of the probability setfunction. Some of these are the following. In listing these properties, weshall not restrict X to be a random variable of the discrete or continuoustype. We shall use the symbols $F(\infty)$ and $F(-0)$ to mean lim*F(x)*and lim*F(x),* respectively. In like manner, the symbols $\{x; x \le b\}$and
$\{x; x \le -\infty\}$ represent, respectively, the limits of the sets$\{x; x \sim b\}$,$\{x; x \le - b\}$ as $b \to \infty$.

(a) $0 \le F(x) \le 1$ because $0 \le \Pr(X \le x) \le 1$.

(b) *F(x)* is a nondecreasing function of*x*.
For, if $x' < x''$, then$\{x; x \le x''\} = \{x; x \le x'\} \cup \{x; x' < x \le x''\}$

And $\Pr(X \le x'') = \Pr(X \le x') + \Pr(x' < X \le x'')$.

That is $F(x'') - F(x') = \Pr(x' < X \leq x'') \geq 0$.

(c) $F(\infty) = 1$ and $F(-\infty) = 0$ because the set $\{x;\ x \leq \infty\}$ is theentire one-dimensional space and the set $\{x;\ x \leq -\infty\}$ is the null set.

From the proof of (b), it is observed that, if $a < b$, then
$\Pr(a < X \leq b) = F(b) - F(a)$.

Suppose that we want to use $F(x)$ to compute the probability $\Pr(X = b)$.
To do this, consider, with $h > 0$,

$$\lim_{h \to 0} Pr(b - h < X \leq b) = \lim_{h \to 0}[F(b) - F(b - h)]$$

Intuitively, it seems that $\lim_{h \to 0} \Pr(b - h < X \leq b)$ should exist and be equal to $\Pr(X = b)$ because, as h tends to zero, the limit of the set $\{x;\ b - h < x \leq b\}$ is the set that contains the single point $x = b$. Thefact that this limit is $\Pr(X = b)$ is a theorem that we accept withoutproof. Accordingly, we have$\Pr(X = b) = F(b) - F(b-)$,where F(b-) is the left limit of F(x) at b.
That is, the probability that $X = b$ is the height of the step that $F(x)$ has at $x = b$.
Hence, if the distribution function $F(x)$ is continuous at $x = b$, then$\Pr(X = b) = 0$.
There is a fourth property of $F(x)$ that is now listed.

(d) $F(x)$ is continuous to the right at each point x.

To prove this property, consider, with $h > 0$,

$$\lim_{h \to 0} Pr(a < X \leq a + h) = \lim_{h \to 0}[F(a + h) - F(a)]$$

We accept without proof a theorem which states, with $h > 0$, that

$$\lim_{h \to 0} Pr(a < X \leq a + h) = P(0) = 0$$

Here also, the theorem is intuitively appealing because, as h tends , the limit of the set
$\{x;\ a < x \leq a + h\}$is the null set. Accordingly, we write $0 = F(a+) - F(a)$,where $F(a+)$ is the right-hand limit of $F(x)$ at $x = a$. Hence $F(x)$ is continuous to the right at every point $x = a$.

The preceding discussion may be summarized in the following manner:

A distribution function $F(x)$ is a non decreasing function of x, which is everywhere continuous to the right and has $F(-0) = 0$,$F(\infty) = 1$. The probability $\Pr(a < X \leq b)$ is equal to the difference $F(b) - F(a)$. If x is a discontinuity point of $F(x)$, then the probability $\Pr(X = x)$ is equal to the jump which the distribution function has at the point x. If x is a continuity point of $F(x)$, then

Pr (X = *x)* = 0.Let X be a random variable of the continuous type that has p.d.f f*(x)* , and let *A* be a set of probability measure zero; that is, *P(A)* =Pr (X E *A)* = O. It has been observed that we may change the definition of *j(x)* at any point in *A* without in any way altering the distribution
of probability. The freedom to do this with the p.d.f *j(x),* of a continuous type of random variable does not extend to the *F(x);* for, if *F(x)* is changed at so much as one point *x,* the probability Pr (X ~ *x)* = *F(x)* is changed, and we have a different distribution of probability. That is, the distribution function *F(x),* not the p.d.f f *(x),* is really the fundamental concept.

***Example* 3**. Let a distribution function be given by

$$F(x) = \begin{cases} 0 & x < 0 \\ \frac{x+1}{2} & 0 \le x < 1 \\ 1 & 1 \le x \end{cases}$$

Then, for instance,
Pr (-3 < X $\le \frac{1}{2}$) = F($\frac{1}{2}$) - *F(* -3) = 3/4 - 0 = 3/4
And Pr*(X* = 0) = *F(0)* - *F(0-)* = 1/2 - 0 = l/2

We see that *F(x)* is not always continuous, nor is a step function. Accordingly, the corresponding distribution is neither of the continuous type nor of the discrete type. It may be described as a mixture of those types. We shall now point out an important fact about a function of a random variable. Let *X* denote a random variable with space d. Consider the function Y = *u(X)* of the random variable X. Since *X* is a function defined on a sample space S, then Y = *u(X)* is a composite function defined on S. That is, Y = *u(X)* is itself a random variable which, has its own space $B = \{y; y = u(x), x \epsilon d\}$ and its own probability set function. If $y \epsilon B$the event Y = *u(X)* ≤*y* occurs when, and only when, the event $X \epsilon A \subset d$ occurs, where $A = \{x; u(x) \le y\}$.That is, the distribution function of Y is

$$G(y) = \Pr(Y \le y) = \Pr[u(X) \le y] = P(A)$$

The following example illustrates a method of finding the distribution function and the p.d.f. of a function of a random variable.

***Example* 4**. Let $f(x) = \frac{1}{2}, -1 < x < 1$, zero elsewhere, be the p.d.f. of the random variable X. Define the random variable Y by Y = X^2. We wish to find the p.d.f. of Y. If $y \ge 0$, the probability $\Pr(Y \le y)$ is equivalent to

$$\Pr(X^2 \le y) = \Pr(-\sqrt{y} \le X \le \sqrt{y}).$$

Accordingly, the distribution function of Y, $G(y) = \Pr(Y \le y)$*,* is given by

$$G(y) = \begin{cases} 0, & y < 0, \\ \int_{-\sqrt{y}}^{\sqrt{y}} \frac{1}{2} dx = \sqrt{y}, & 0 \le y < 1, \\ 1, & 1 \le y. \end{cases}$$

Since Y is a random variable of the continuous type, the p.d.f. of Y is *g(y) = G'(y)* at all points of continuity of *g(y).* Thus we may write

$$g(y) = \begin{cases} \frac{1}{2\sqrt{y}}, & 0 < y < 1, \\ 0 & elsewhere. \end{cases}$$

Let the random variables *X* and Y have the probability set function *P(A),* where *A* is a two-dimensional set. If *A* is the unbounded set $\{(u, v); u \le x, v \le y\}$, where *X* and y are real numbers, we have

$$P(A) = \Pr[(X, Y) \, \epsilon A] = \Pr(X \le x, Y \le y).$$

This function of the point *(x, y)* is called the *distribution function* of *X* and Y and is denoted by

$$F(x, y) = \Pr(X \le x, Y \le y).$$

If X and Yare random variables of the continuous type that havep.d.f.$f(x, y)$, then

$$F(x, y) = \int_{-\infty}^{y} \int_{-\infty}^{x} f(u, v) \, du \, dv.$$

Accordingly, at points of continuity of $f(x, y)$, we have

$$\frac{\partial^2 F(x, y)}{\partial x \, \partial y} = f(x, y).$$

It is left as an exercise to show, in every case, that

$$\Pr(a < X \le b, c < Y \le d) = F(b, d) - F(b, c) - F(a, d) + F(a, c).$$

for all real constants $a < b, c < d$.

The distribution function of the *n* random variables X_1,X_2,….,$X_n$is the point function

$$F(x_1, x_2, \dots, x_n) = \Pr(X_1 \le \ x_1, X_2 \le x_2, \dots, X_n \le x_n)$$

An illustrative example follows.

***Example* 5**. Let $f(x, y, z) = e^{-(x+y+z)}, 0 < x, y, z < \infty$, zero elsewhere,be the p.d.f. of the random variables X, Y, and *Z.* Then the distribution function of X, Y, and *Z* is given by

$$F(x,y,z) = \begin{cases} \Pr(X \leq x, Y \leq y, Z \leq z) \\ \int_0^z \int_0^y \int_0^x e^{-u-v-w}\, du\, dv\, dw \\ (1-e^{-x})(1-e^{-y})(1-e^{-z}), \qquad 0 \leq x, y, z < \infty \end{cases}$$

and is equal to zero elsewhere. Incidentally, except for a set of probability measure zero, we have

$$\frac{\partial^3 F(x,y,z)}{\partial x\, \partial y\, \partial z} = f(x,y,z).$$

Certain Probability Models

The probability model described in the following:

***Example* 1**. Let a card be drawn at random from an ordinary deck of 52 playing cards. The sample space S is the union of $k = 52$ outcomes, and it is reasonable to assume that each of these outcomes has the same probability $\frac{1}{52}$.Accordingly, if E_1 is the set of outcomes that are spades, $P(E_1) = \frac{13}{52} = \frac{1}{4}$ because there are $r_1 = 13$ spades in the deck; that is, $\frac{1}{4}$ is the probability of drawing a card that is a spade. If $E_2$is the set of outcomes that are kings, $P(E_2) = \frac{4}{52} = \frac{1}{13}$ because there are $r_2 = 4$ kings in the deck; that is, $\frac{1}{13}$ is the probability of drawing a card that is a king. These computations are very easy because there are no difficulties in the determination of the appropriate values of r and *k.* However, instead of drawing only one card, suppose that five cards are taken, at random and without replacement, from this deck. We can think of each five-card hand as being an outcome in a sample space. It is reasonable to assume that each of these outcomes has the same probability. Now if E_1 is the set of outcomes in which each card of the hand is a spade, $P(E_1)$ is equal to the number $r_1$of all spade hands divided by the total number, say *k,* of five-card hands. It is shown in many books on algebra that

$$r_1 = \binom{13}{5} = \frac{13!}{5!\,8!} \qquad and \quad k = \binom{52}{5} = \frac{52!}{5!\,47!}$$

In general, if *n* is a positive integer and if *x* is a nonnegative integer with$x \leq n$, then the binomial coefficient

$$\binom{n}{x} = \frac{n!}{x!\,(n-x)!}$$
is equal to the number of combinations of *n* things taken *x* at a time. Thus, here,

$$P(E_1) = \frac{\binom{13}{5}}{\binom{52}{5}} = \frac{(13)(12)(11)(10)(9)}{(52)(51)(50)(49)(48)} = 0.0005,$$

approximately. Next, let E_2 be the set of outcomes in which at least one card is a spade. Then E_2^* is the set of outcomes in which no card is a spade. There are $r_2^* = \binom{39}{5}$such outcomes. Hence

$$P(E_2^*) = \frac{\binom{13}{5}}{\binom{52}{5}} \qquad and \;\; P(E_2) = 1 - P(E_2^*).$$

Now suppose that E_3 is the set of outcomes in which exactly three cards are kings and exactly two cards are queens. We can select the three kings in any one of $\binom{4}{3}$ ways and the two queens in anyone of $\binom{4}{2}$ ways. By a well-known counting principle, the number of outcomes in E_3 is

$$r_3 = \binom{4}{3}\binom{4}{2}.$$

Thus, $P(E_3) = \binom{4}{3}\binom{4}{2} \Big/ \binom{52}{5}$. Finally, let E_4be the set of outcomes in which there are exactly two kings, two queens, and one jack. Then

$$P(E_4) = \frac{\binom{4}{2}\binom{4}{2}\binom{4}{1}}{\binom{52}{5}},$$

because the numerator of this fraction is the number of outcomes in E_4.

***Example* 2**. A lot, consisting of 100 fuses, is inspected by the following procedure. Five of these fuses are chosen at random and tested; if all 5"blow" at the correct amperage, the lot is accepted. If, in fact, there are 20 defective fuses in the lot, the probability of accepting the lot is, under appropriate assumptions,

$$\frac{\binom{80}{5}}{\binom{100}{5}} = 0.32,$$

approximately. More generally, let the random variable X be the number of defective fuses among the 5 that are inspected. The space of X is d =*{x; x* = 0, 1, 2, 3, 4, 5} and the p.d.f. of X is given by

$$f(x) = \Pr(X = x) = \begin{cases} \frac{\binom{20}{x}\binom{80}{5-x}}{\binom{100}{5}}, & x = 0,1,2,3,4,5, \\ 0 & elsewhere. \end{cases}$$

This is an example of a discrete type of distribution called a *hyper geometric distribution.*

1.7. Mathematical Expectation

Let X be a random variable having a p.d.f. f(x), and let *u(X)* be a function of X such that $\int_{-\infty}^{\infty} u(x)f(x)\,dx$ exists, if X is a continuous type of random variable, or such that

$$\boldsymbol{E}[\boldsymbol{u}(\boldsymbol{x})] = \sum_{\boldsymbol{x}} \boldsymbol{u}(\boldsymbol{x})\boldsymbol{f}(\boldsymbol{x})$$

exists, if X is a discrete type of random variable. The integral, or the sum, as the case may be, is called the *mathematical expectation.*

Remarks.

The usual definition of *E[u(X)]* requires that the integral(or sum) converge absolutely.

We may observe that *u(X)* is a random variable Y with its own distribution of probability. Suppose the p.d.f. of Y is g(y). Then E(y) is given by

$$\int_{-\infty}^{\infty} yg(y)\,dy \quad or \sum_{y} yg(y),$$

According as Y of the continuous type or of the discrete type.

Results:

(a) If k is a constant, then E(k) = k.

(b) If k is a constant and v is a function, then E(kv) = kE(v).

(c) If k_1 and k_2 are constants and v_1 and v_2 are functions, then $E(k_1v_1+k_2v_2) = k_1E(v_1) + k_2E(v_2)$.

***Example* 1**. Let X have the p.d.f.

$$f(x) = \begin{cases} 2(1-x), & 0 < x < 1, \\ 0 & elsewhere. \end{cases}$$

Then

$$E(X) = \int_{-\infty}^{\infty} xf(x)dx = \int_{-\infty}^{\infty} (x)2(1-x)dx = \frac{1}{3},$$

$$E(X^2) = \int_{-\infty}^{\infty} x^2 f(x)dx = \int_{-\infty}^{\infty} (x^2)2(1-x) = \frac{1}{6},$$

And, of course,

$$E(6X+3X^2) = 6(1/3)+3(1/6) = 5/2$$

***Example* 2**. Let X have the p.d.f.

$$f(x) = \begin{cases} \frac{x}{6}, & x = 1,2,3, \\ 0, & elsewhere \end{cases}.$$

Then,

$$E(X^3) = \sum_x x^3 f(x) = \sum_{x=1}^{3} x^3 \frac{x}{6} = \frac{1}{6} + \frac{16}{6} + \frac{81}{6} = \frac{98}{6}.$$

***Example* 3**. Let X and Y have the p.d.f.

$$f(x,y) = \begin{cases} x + y, & 0 < x < 1.\ 0 < y < 1, \\ 0 & elsewhere. \end{cases}$$

Then,

$$\begin{aligned} E(XY^2) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy^2 f(x,y) dx\, dy \\ &= \int_0^1 \int_0^1 xy^2(x+y)\, dx\, dy = \frac{17}{72} \end{aligned}$$

***Example* 4**. Let us divide, at random, a horizontal line segment of length 5 into two parts. If X is the length of the left-hand part, it is reasonable to assume that X has the p.d.f.

$$f(x) = \begin{cases} \frac{1}{5}, & 0 < x < 5, \\ 0 & elsewhere. \end{cases}$$

The expected value of the length X is E(X) = 5/2 and the expected value of the length 5-x is E(5-x) = 5/2. But the expected value of the product of the two lengths is equal to

$$E[X(5-x)] = \int_0^5 x(5-x)\left(\frac{1}{5}\right) dx = \frac{25}{6} \neq \left(\frac{5}{2}\right)^2.$$

That is, in general, the expected value of a product is not equal to the product of the expected values.

***Example* 5**. A bowl contains five chips, which cannot be distinguished by a sense of touch alone. Three of the chips are marked \$1 each and the remaining two are marked \$4 each. A player is blindfolded and draws, at random and without replacement, two chips from the bowl. The player is paid an amount equal to the sum of the values of the two chips that he draws and the game is over. If it costs \$4.75 cents to play this game, would we care to participate for any protracted period of time? Because we are unable to distinguish the chips by sense of touch, we assume that each of the 10 pairs that can be drawn has the same probability of being drawn. Let the random variable X be the number of chips, of the two to be chosen, that are marked \$1. Then, under our assumption, X has the hyper geometric p.d.f.

$$f(x) = \begin{cases} \dfrac{\binom{3}{x}\binom{2}{2-x}}{\binom{5}{2}}, & x = 0,1,2, \\ 0 & elsewhere. \end{cases}$$

If X = x, the player receives $u(x) = x + 4(2 - x) = 8 - 3x$ dollars. Hence his mathematical expectation is equal to

$$E[8 - 3X] = \sum_{x=0}^{2} (8 - 3x) f(x) = \frac{44}{10},$$

or \$4.40.

1.8. Some Special Mathematical Expectations
let *u(X) = X,* where *X* is a random variable of the discrete type having a p.d.f. *f(x).* Then

$$E(x) = \sum_{x} x f(x).$$

If the discrete points of the space of positive probability density are $a_1, a_2, a_3, \ldots$, then

$$E(X) = a_1 f(a_1) + a_2 f(a_2) + a_3 f(a_3) + \cdots.$$

This sum of products is seen to be a "weighted average" of the values $a_1, a_2, a_3, \ldots$, the "weight" associated with each a_i, being $f(a_i)$. This suggests that we call *E(X)* the arithmetic mean of the values of *X,* or, more simply, the *mean value* of X (or the mean value of the distribution).

The mean value μ of a random variable X is defined, when it exists, to be $\mu = E(X)$, where *X* is a random variable of the discrete or of the continuous type.

The variance of *X* will be denoted by a_2 , and we define a_2, if it exists, by $a_2 = E[(X - \mu)^2]$, whether X is a discrete or a continuous type of random variable.

It is worthwhile to observe that

$$\sigma^2 = E[(X - \mu)^2] = E(X^2 - 2\mu X + \mu^2);$$

and since E is a linear operator,

$$\sigma^2 = E(X^2) - 2\mu E(X) + \mu^2$$

$$= E(X^2) - 2\mu^2 + \mu^2$$

$$\sigma^2 = E(X^2) - \mu^2.$$

***Example* 1**. Let X have the p.d.f.

$$f(x) = \begin{cases} \frac{1}{2}(x+1), & -1 < x < 1, \\ 0 & elsewhere \end{cases}.$$

Then the mean value of X is

$$\mu = \int_{-\infty}^{\infty} x f(x) dx = \int_{-1}^{1} x \frac{x+1}{2} dx = \frac{1}{3}$$

While the variance of X is

$$\sigma^2 = \int_{-\infty}^{\infty} x^2 f(x) dx - \mu^2 = \int_{-1}^{1} x^2 \frac{x+1}{2} dx - \left(\frac{1}{3}\right)^2 = \frac{2}{9}.$$

***Example* 2**. If X has the p.d.f.

$$f(x) = \begin{cases} \frac{1}{x^2}, & 1 < x < \infty, \\ 0 & elsewhere. \end{cases}$$

Then the mean value of X does not exist, since

$$\int_{1}^{\infty} x \frac{1}{x^2} dx = \lim_{b \to \infty} \int_{1}^{b} \frac{1}{x} dx$$

$$= \lim_{b \to \infty} (\ln b - \ln 1)$$

does not exist.

***Example* 3**. Given that the series

$$\frac{1}{1^2}+\frac{1}{2^2}+\frac{1}{3^2}+\cdots.$$

converges to $\pi^2/6$. Then

$$f(x)=\begin{cases}\frac{6}{\pi^2x^2}, & x=1,2,3,\dots,\\ 0 & elsewhere.\end{cases}$$

is the p.d.f. of a discrete type of random variable X. The moment-generating function of this distribution, if it exists, is given by

$$M(t)=E(e^{tX})=\sum_x e^{tx}f(x)$$

$$=\sum_{x=1}^{\infty}\frac{6e^{tx}}{\pi^2x^2}.$$

***Example* 4**. Let X have the moment-generating function $M(t)=e^{t^2/2}, -\infty < t < \infty$. We can differentiate *M(t)* any number of times to find the moments of X. However it is instructive to consider this alternative method. The function *M(t)* is represented by the following MacLaurin's series.

$$e^{t^2/2}=1+\frac{1}{1!}\left(\frac{t^2}{2}\right)+\frac{1}{2!}\left(\frac{t^2}{2}\right)^2+\cdots+\frac{1}{k!}\left(\frac{t^2}{2}\right)^k+\cdots$$

$$=1+\frac{1}{2!}t^2+\frac{(3)(1)}{4!}t^4+\cdots+\frac{(2k-1)\dots(3)(1)}{(2k)!}t^{2k}+\cdots$$

In general, the MacLaurin's series for M(t) is

$$M(t)=M(0)+\frac{M'(0)}{1!}t+\frac{M''(0)}{2!}t^2+\cdots+\frac{M^{(m)}(0)}{m!}t^m+\cdots$$

$$=1+\frac{E(X)}{1!}t+\frac{E(X^2)}{2!}t^2+\cdots+\frac{M(X^m)}{m!}t^m+\cdots$$

Thus the coefficient of $(t^m/m!)$ in the MacLaurin's series representation of M(t) is $E(X^m)$. So, for our particular M(t), we have

$$E(X^{2k})=(2k-1)(2k-3)\dots(3)(1)=\frac{(2k)!}{2^kk!}$$

k = 1,2,3,…, and $E(X^{2k-1}) = 0$, k = 1,2,3,….

1.9. Chebyshev's Inequality

Theorem 6. Let u(X) be a nonnegative function of the random variable X. If E[u(X)] exists, then, for every positive constant c,

Proof.
The proof is given when the random variable X is of the continuous type; but the proof can be adapted to the discrete case if we replace integrals by sums. Let $A = \{x;\ u(x) \geq c\}$ and let *f(x)* denote the p.d.f. of X. Then

$$E[u(X)] = \int_{-\infty}^{\infty} u(x)f(x)dx = \int_{A} u(x)f(x)dx + \int_{A^*} u(x)f(x)dx.$$

Since each of the integrals in the extreme right-hand member of the preceding equation is nonnegative, the left-hand member is greater than or equal to either of them. In particular,

$$E[u(X)] \geq \int_{A} u(x)f(x)dx.$$

However, if $x \epsilon A$, then $u(x) \geq c$; accordingly, the right-hand member of the preceding inequality is not in creed if we replace *u(x)* by *c.* Thus

$$E[u(X)] \geq c\int_{A} f(x)dx.$$

Since

$$\int_{A} f(x)dx = \Pr(X \in A) = \Pr[u(X) \geq c],$$

It follows that

$$E[u(x)] \geq cPr[u(X) \geq c],$$

which is the desired result.

Theorem 7. Chebyshev's Inequality. Let the random variable X have a distribution of probability about which we assume only that there is a finite variance σ^2. This, of course, implies that there is a mean μ. Then for every k > 0,

$$\Pr(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2},$$

or, equivalently,

$$\Pr\ (|X-\mu|<ka)\geq 1-\frac{1}{k^2}.$$

Proof.
In the above Theorem 6 take $u(X)=(X-\mu)^2$ and $c=k^2\sigma^2$. Then we have

$$\Pr[(X-\mu)^2\geq k^2\sigma^2]\leq\frac{E[(X-\mu)^2]}{k^2\sigma^2}.$$

Since the numerator of the right-hand member of the preceding inequality is σ^2, the inequality may be written

$$\Pr(|X-\mu|\geq k\sigma)\leq\frac{1}{k^2},$$

which is the desired result. Naturally, we would take the positive number k to be greater than 1 to have an inequality of interest.

It is seen that the number $1/k^2$ is an upper bound for the probability$\Pr(|X-\mu|\geq k\sigma)$.In the following example this upper bound and the exact value of the probability are compared in special instances.

***Example* 1**. Let X have the p.d.f.

$$f(x)=\begin{cases}\frac{1}{2\sqrt{3}}, & -\sqrt{3}<x<\sqrt{3},\\ 0 & elsewhere.\end{cases}$$

Here $\mu = 0$ and $\sigma^2 = 1$. If $k=\frac{3}{2}$, we have the exact probability

$$\Pr(|X-\mu|\geq k\sigma)=\Pr\left(|X|\geq\frac{3}{2}\right)=1-\int_{-3/2}^{3/2}\frac{1}{2\sqrt{3}}dx=1-\frac{\sqrt{3}}{2}.$$

By Chebyshev's inequality, the preceding probability has the upper bound $1/k^2=\frac{4}{9}$. Since $1-\frac{\sqrt{3}}{2}=0.134$, approximately, the exact probability in this case is considerably less than the upper bound $\frac{4}{9}$. If we take k = 2, we have the exact probability

$$\Pr(|X-\mu|\geq 2\sigma)=\Pr(|X|\geq 2)=0.$$

This again is considerably less than the upper bound $\frac{1}{k^2}=\frac{1}{4}$ provided by Chebyshev's inequality.

In each instance in the preceding example, the probability $\Pr\ (|X-\mu|\geq k\sigma)$ and its upper bound $1/k^2$ differ considerably. This suggests that this inequality might be made sharper. However, if we want an inequality that holds for k>0 and holds for all random variables having finite variance, such an improvement is impossible as is shown by the following example.

***Example* 2.** Let the random variable X of the discrete type have probabilities $\frac{1}{8}, \frac{6}{8}, \frac{1}{8}$ at the points x = -1, 0, 1, respectively. Here $\mu = 0$ and σ^2= t. If $k = 2$, then $1/k^2$= t and

$$\Pr(|x-\mu| \geq k\sigma) = \Pr(|X| \geq 1) = \frac{1}{4}.$$

That is, the probability $\Pr(|X-\mu| \geq k\sigma)$here attains the upper bound $1/k^2 = \frac{1}{4}$. Hence the inequality cannot be improved without further assumptions about the distribution of X.

Exercise:

1. A point is to be chosen in a haphazard fashion from the interior of a fixed circle. Assign a probability p that the point will be inside another circle, which has a radius of one-half the first circle and which lies entirely within the first circle.
2. An unbiased coin is to be tossed twice. Assign a probability P_1 to the event that the first toss will be held and that the second toss will be a tail. Assign a probability P_2 to the event that there will be one head and one tail in the two tosses.

3. Find the union $A_1 \cup A_2$ and the intersection $A_1 \cap A_2$ of the two sets A_1 and A_2, where
a) $A_1 = \{x; x = 0,1,2\}$, $A_2 = \{x; x = 2,3,4\}$
b) $A_1 = \{x, 0 < x < 2\}$, $A_2 = \{x; 1 \leq x < 3\}$
c) $A_1 = \{(x, y); 0<x<2, 0<y<2\}$, $A_2 = \{(X,Y); 1<X<3, 1<Y<3\}$.

4. If the sample space is $C = C_1 \cup C_2$ and if $P(C_1) = 0.8$ and $P(C_2) = 0.5$, find $P(C_1 \cap C_2)$.

5. Let the space of the random variable X be $A = \{x; 0<x<10\}$ if $A_1 = \{x; 0<x<1/2\}$ and $A_2 = \{x; ½<x<1\}$,find $P(A_2)$ if $P(A_1)$.

6. Let the subsets $A_1 = \{x; ¼<x<1/2\}$ and $A_2 = \{x; ½<x<1\}$ of the space $A = \{x; 0<x<1\}$ of the random variable X be such that $P(A_1) = 1/8$ and $P(A_2) = ½$. Find $P(A_1 \cup A_2)$. $P(A_1^*)$ and $P(A_1^* \cap A_2^*)$.

7. Give $\int_A \left[\frac{1}{\pi}(1+x^2)\right] dx$, where $A \subset A = \{x; -\infty<x<\infty\}$ show that the integral could serve as a probability set function of a random variable X whose space is A.

8. What is the value of $\int_0^\infty xne^{-x}\, dx$, where n is a nonnegative integer?

9. Let $f(x) = \begin{cases} \frac{x}{15}, & x = 1,2,3,4,5, \\ 0 & elsewhere. \end{cases}$ Be the p.d.f. of X. Find Pr(X=1 or 2). Pr(1/2<X<5/2) and Pr(1≤X≤2).

10. Compute the probability of being dealt at random and without replacements a 13-card bridge hand consisting of (a) 6 spades, 4 hearts, 2 diamonds and 1 club; (b) 13 cards of the same suit.

11. Let X have the uniform distribution given by the p.d.f. $f(x) = \begin{cases} \frac{1}{5}, & x = -2,-1,0,1,2, \\ 0 & elsewhere. \end{cases}$ Find the p.d.f of Y = X^2.

12. Let the p.d.f. of X and Y be $f(x,y) = \begin{cases} e^{-x-y}, & 0 < x < \infty, 0 < y < \infty, \\ 0 & elsewhere \end{cases}$. Let u(X,Y) =X, v(X,Y) = Y and w(X,Y) = XY. Show that E[u(X,Y)].E[v(X,Y)] = E[w(X,Y)].

13. Let X have a p.d.f. f(x) is positive at x = -1,0,1 and is zero elsewhere

a) If f(0) = ½, find E(X^2).
b) If f(0) = ½, and if E(X) = 1/6, determine F(-1) and f(1).

14. Let $f(x) = \begin{cases} (\frac{1}{2})^3, & x = 1,2,3, \ldots, \\ 0 & elsewhere. \end{cases}$ Be the p.d.f. of the random variable X. Find the moment-generating function, the mean and the variance of X.

15. Let the random variable X have the p.d.f.
$f(x) = \begin{cases} p, & x = -1,1, \\ 1-2p, & x = 0 \\ 0 & elsewhere \end{cases}$, Where 0<p<1/2. Find the measure of kurtosis as a function of p. Determine its value when p = 1/3, p = 1/5, p = 1/10 and p = 1/100. Note that the kurtosis increases as p decreases.

Unit II
CONDITIONAL PROBABILITY AND STOCHASTIC INDEPENDENCE

2.1 Conditional Probability

Let the probability set function P(C) be defined on the sample space and let C_1 be a subset of such that $P(C_1)>0$. The *conditional probability* of the event C_2, relative to the event C_1 or, more briefly, the conditional probability of C_2, given C_1 is denoted by $P(C_2|C_1)$ and is defined by

$$P(C_1|C_1) = 1 \qquad and\ \ P(C_2|C_1) = P(C_1 \cap C_2|C_1)$$

Hence

$$P(C_2|C_1) = \frac{P(C_1 \cap\ C_2)}{P(C_1)}.$$

In a suitable definition of the conditional probability of the event C_2, given the event C_1, provided $P(C_1) > 0$.

Let *P* denote the probability set function of the induced probability on A. If A_1 and A_2 are subsets of A, the conditional probability of the event A_2 , given the event A_1 , is

$$P(A_2|A_1) = \frac{P(A_1 \cap A_2)}{P(A_1)}$$

provided $P(A_1) >0$.

***Example* 1**. A hand of 5 cards is to be dealt at random and without replacement from an ordinary deck of 52 playing cards. The conditional probability of an all-spade hand (C_2) , relative to the hypothesis that there are at least 4 spades in the hand (C_1) , is, since $C_1 \cap C_2 = C_2$.

$$P(C_2|C_1) = \frac{P(C_2)}{P(C_1)} = \frac{\binom{13}{5}/\binom{52}{5}}{[\binom{13}{4}\binom{39}{1} + \binom{13}{5}]/\binom{52}{5}}.$$

***Example* 2.** A bowl contains eight chips. Three of the chips are red and the remaining five are blue. Two chips are to be drawn successively, at random and without replacement. We want to compute the probability that the first draw results in a red chip (C_1) and that the second draw results in a blue chip (C_2) . It is reasonable to assign the following probabilities:

$$P(C_1) = \frac{3}{8} \qquad \text{and} \quad P(C_2|C_1) = \frac{5}{7}.$$

Thus, under these assignments, we have

$$P(C_1 \cap C_2) = \left(\frac{3}{8}\right)\left(\frac{5}{7}\right) = \frac{15}{56}.$$

***Example* 3**. From an ordinary deck of playing cards, cards are to be drawn successively, at random and without replacement. The probability that the third spade appears on the sixth draw is computed as follows. Let C_1 be the event of two spades in the first five draws and let C_2 be the event of a spade on the sixth draw. Thus the probability that we wish to compute is $P(C_1 \cap C_2)$. It is reasonable to take

$$P(C_1) = \frac{\binom{13}{2}\binom{39}{3}}{\binom{52}{5}} \qquad and \;\; P(C_2|C_1) = \frac{11}{47}.$$

The desired probability $P(C_1 \cap C_2)$ is then the product of these two numbers.More generally, if $X + 3$ is the number of draws necessary to produce exactly three spades, a reasonable probability model for the random variable X is given by the p.d.f.

$$f(x) = \begin{cases} \left[\frac{\binom{13}{2}\binom{39}{x}}{\binom{52}{2+x}}\right]\left(\frac{11}{50-x}\right), & x = 0,1,2,\ldots,39, \\ 0 & elsewhere. \end{cases}$$

Then the particular probability which we computed is $P(C_1 \cap C_2) = \Pr(X = 3) = f(3)$.

***Example* 4**. Four cards are to be dealt successively, at random and without replacement, from an ordinary deck of playing cards. The probability of receiving a spade, a heart, a diamond, and a club, in that order, is $\left(\frac{13}{52}\right)\left(\frac{13}{81}\right)\left(\frac{13}{50}\right)\left(\frac{13}{49}\right)$. This follows from the extension of the multiplication rule.In this computation, the assumptions that are involved seem clear.

2.2 Marginal and Conditional Distributions

Let $f(x_1, x_2)$be the p.d.f. of two random variables $X_1$and X_2.From this point on, for emphasis and clarity, we shall call a p.d.f. or a distribution function a *joint* p.d.f. or a *joint* distribution function when more than one random variable is involved. Thus$f(x_1, x_2)$ is the joint p.d.f. of the random variables X_1 and X_2. Consider the event $a < X_1 < b$, $a < b$. This event can occur when and only when the event $a < X_1 < b$, $-\infty < X_2 < \infty$ occurs; that is, the two events are equivalent, so that they have the same probability. But the probability of the latter event has been defined and is given by

$$\Pr(a < X_1 < b, -\infty < X_2 < \infty) = \int_a^b \int_{-\infty}^{\infty} f(x_1, x_2) dx_2 dx_1$$

for the continuous case, and by

$$\Pr(a < X_1 < b, -\infty < X_2 < \infty) = \sum_{a<x_2<b} \sum_{x_2} f(x_1, x_2)$$

for the discrete case. Now each of

$$\int_{-\infty}^{\infty} f(x_1, x_2) dx_2 \qquad and \qquad \sum_{x_2} f(x_1, x_2)$$

Is a function of x_1 alone, say $f_1(x_1)$. Thus, for every a<b, we have

$$\Pr(a < X_1 < b) = \begin{cases} \int_a^b f_1(x_1) dx_1, & (continuous\ case), \\ \sum_{a<x_1<b} f_1(x_1), & (discrete\ case). \end{cases}$$

so that $f_1(x_1)$is the p.d.f. of X_1 alone. Since $f_1(x_1)$ is found by summing (or integrating) the joint p.d.f.$f(x_1, x_2)$ over all x_2 for a fixed x_1 we can think of recording this sum in the" margin" of the x_1x_2- plane. Accordingly,$f_1(x_1)$is called the marginal p.d.f. of X_1 In like manner

$$f_2(x_2) = \begin{cases} \int_{-\infty}^{\infty} f(x_1, x_2) dx_1, & (continuous\ case), \\ \sum_{x_1} f(x_1, x_2), & (discrete\ case). \end{cases}$$

is called the marginal p.d.f. of X_2.

***Example* 1**. Let the joint p.d.f. of X_1 and X_2be

$$f(x_1, x_2) = \begin{cases} \frac{x_1 + x_2}{21}, & x_1 = 1,2,3, x_2 = 1,2, \\ 0 & elsewhere. \end{cases}$$

Then, for instance,

$$\Pr(X_1 = 3) = f(3,1) + f(3,2) = \frac{3}{7}$$

And

$$\Pr(X_2 = 2) = f(1,2) + f(2,2) + f(3,2) = \frac{4}{7}.$$

On the other hand, the marginal p.d.f. of X_1 is

$$f_1(x_1) = \sum_{x_2=1}^{2} \frac{x_1 + x_2}{21} = \frac{2x_1 + 3}{21}, \qquad x_1 = 1,2,3$$

Zero elsewhere, and the marginal p.d.f. of X_2

$$f_2(x_2) = \sum_{x_1=1}^{3} \frac{x_1 + x_2}{21} = \frac{6 + 3x_2}{21}, \quad x_2 = 1,2,$$

Zero elsewhere. Thus preceding probabilities may be computed as $\Pr(X_1 = 3) = f_1(3) = \frac{3}{7}$
and $\Pr(X_2 = 2) = f_2(2) = \frac{4}{7}$.

***Example* 2**. Let the joint p.d.f. of X_1 and X_2be

$$f(x_1, x_2) = \begin{cases} 2, & 0 < x_1 < x_2 < 1, \\ 0 & elsewhere \end{cases}$$

Then the marginal probability density functions are, respectively,

$$f_1(x_1) = \begin{cases} \int_{x_1}^{1} 2dx_2 = 2(1 - x_1), & 0 < x_1 < 1, \\ 0 & elsewhere. \end{cases}$$

And

$$f_2(x_2) = \begin{cases} \int_{0}^{x_2} 2dx_1 = 2(x_2), & 0 < x_2 < 1, \\ 0 & elsewhere. \end{cases}$$

The conditional p.d.f. of X_1, given $X_2 = x_2$, is

$$f(x_1|x_2) = \begin{cases} \frac{2}{2x_2} = \frac{1}{x_2}, & 0 < x_1 < x_2, 0 < x_2 < 1, \\ 0 & elsewhere. \end{cases}$$

Here the conditional mean and conditional variance of X_1, given $X_2 = x_2$, are respectively,

$$E(X_1|x_2) = \int_{-\infty}^{\infty} x_1 f(x_1|x_2)\, dx_1$$

$$= \int_0^{x_2} x_1 \left(\frac{1}{x_2}\right) dx_1$$

$$= \frac{x_2}{2}, \quad 0 < x_2 < 1$$

and

$$E\{[X_1 - E(X_1|x_2)]^2|x_2\} = \int_0^{x_2} (x_1 - \frac{x_2}{2})^2 \left(\frac{1}{x_2}\right) dx_1$$

$$= \frac{x_2^2}{12}, \qquad 0 < x_2 < 1$$

Finally, we shall compare the values of Pr $(0 < X_1 < \frac{1}{2} |X_2 = \frac{3}{4})$ andPr $(0 < X_1 < \frac{1}{2})$.
We have

$$\Pr\left(0 < X_1 < \frac{1}{2}\middle|X_2 = \frac{3}{4}\right) = \int_0^{\frac{1}{2}} f(x_1|\frac{3}{4}) dx_1 = \int_0^{\frac{1}{2}} \left(\frac{4}{3}\right) dx_1 = \frac{2}{3},$$

But

$$\Pr\left(0 < X_1 < \frac{1}{2}\right) = \int_0^{\frac{1}{2}} f_1(x_1) dx_1 = \int_0^{\frac{1}{2}} 2(1 - x_1) dx_1 = \frac{3}{4}.$$

Let the random variables $X_1, X_2, \ldots X_n$ have the joint p.d.f. f $(x_1, x_2, \ldots, x_n)$.If the random variables are of the continuous type, then by an argument similar to the two-variable case, we have for every $a < b$,

$$\Pr(a < X_1 < b) = \int_a^b f_1(x_1) dx_1,$$

where $f_1(x_1)$ is defined by the (n-1) fold integral

$$f_1(x_1) = \int_{-\infty}^{\infty} \ldots \ldots \int_{-\infty}^{\infty} \mathrm{f}\,(\mathrm{x}_1, \mathrm{x}_2, \ldots, \mathrm{x}_n) dx_2 \ldots\ldots dx_n$$

Accordingly,$f_1(x_1)$ is the p.d.f. of the one random variableX_1 and $f_1(x_1)$is called the marginal p.d.f. of X_1.The marginal probability density functions, $f_2(x_2), \ldots, f_n(x_n)$ of $x_2, \ldots, x_n$ respectively, are similar (n- l)-fold integrals. Each marginal p.d.f. has been a p.d.f. of one random variable. It is convenient to extend this terminology to joint probability density functions. Let f $(x_1, x_2, \ldots, x_n)$ be the joint p.d.f. of the n random variables $X_1, X_2, \ldots X_n$. Take any group of $k < n$ of these random variables and let us find the joint p.d.f. of them. This joint p.d.f. is called the marginal p.d.f. of this particular group of *k* variables. The marginal p.d.f. of X_2, X_4, X_5, is the joint p.d.f. of this particular group of three variables, namely

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x_1, x_2, \ldots, x_6)\, dx_1 dx_3 dx_6$$

if the random variables are of the continuous type.

If $f_1(x_1) > 0$, the symbol

$$f(x_2, x_3, \ldots, x_n|x_1) = \frac{f(x_1, x_2, \ldots, x_n)}{f_1(x_1)}$$

and $f(x_2, x_3, \ldots, x_n|x_1)$ is called the joint conditional p.d.f. of $X_2, \ldots X_n$ given by $X_1 = x_1$. The joint conditional p.d.f of any *n* - 1 random variables, say $X_1, \ldots\ldots, X_{i-1}, X_{i+1}, \ldots\ldots, X_n$ given $X_i = x_i$ is defined as the joint p.d.f. of $X_1, X_2, \ldots X_n$ divided by marginal p.d.f.$f_i(x_i)$ provided $f_i(x_i) > 0$. More generally, the joint conditional p.d.f. of *n* - *k* of the random variables, for given values of the remaining k variables, is defined as the joint p.d.f. of the *n* variables divided by the marginal p.d.f. of the particular group of *k* variables, provided the latter p.d.f. is positive. the conditional expectation of $u(X_2, \ldots X_n)$given $X_1 = x_1$is, for random variables of the continuous type, given by

$$E[u(X_2, \ldots\ldots, X_n|x_1) = \int_{-\infty}^{\infty}\ldots\ldots\int_{-\infty}^{\infty} u(x_2, \ldots, x_n) f(x_1, x_2, \ldots, x_n|x_1) dx_2 \ldots\ldots dx_n,$$

provided $f_1(x_1) > 0$ and the integral converges (absolutely).

2.3 The Correlation Coefficient

Let *X,* Y, and *Z* denote random variables that have joint p.d.f.f(x, y, z). The means of X, Y, and *Z,* say μ_1, μ_2 and $\mu_3$are obtained by taking u(x, y, z)to be *x, y,* and *z,* respectively; and the variances of *X,* Y, and *Z,* say σ_1^2, σ_2^2 and σ_3^2 are obtained by setting the function *u(x, y, z)*equal to $(x - \mu_1)^2$, $(y - \mu_2)^2$ and $(z - \mu_3)^2$, respectively,

$$E[(X-\mu_1)(Y-\mu_2)] = E(XY - \mu_2 X - \mu_1 Y + \mu_1\mu_2).$$

$$= E(XY) - \mu_2 E(X) - \mu_1 E(Y) + \mu_1\mu_2$$

$$= E(XY) - \mu_1\mu_2.$$

This number is called the *covariance* of *X* and Y. The covariance of *X* and *Z* is given by $E[(X - \mu_1)(Z - \mu_3)]$, and the covariance of Y and Z is $E[(Y - \mu_2)(Z - \mu_3)]$. If each of σ_1 and σ_2 positive, the number

$$\rho_{12} = \frac{E[(X-\mu_1)(Y-\mu_2)]}{\sigma_1\sigma_2}$$

is called the *correlation coefficient* of X and Y.

***Example* 1.** Let the random variables X and *Y* have the joint p.d.f.

$$f(x,y) = \begin{cases} x+y, & 0<x<1, 0<y<1, \\ 0 & elsewhere. \end{cases}$$

Compute the correlation coefficient of X and *Y*. When only two variables are under consideration, we shall denote the correlation coefficient by ρ. Now

$$\mu_1 = E(X) = \int_0^1\int_0^1 x(x+y)dx\,dy = \frac{7}{12}$$

and

$$\sigma_1{}^2 = E(X^2) - \mu_1{}^2 = \int_0^1\int_0^1 x^2(x+y)dx\,dy - \left(\frac{7}{12}\right)^2 = \frac{11}{144}.$$

Similarly

$$\mu_2 = E(Y) = \frac{7}{12} \quad and \quad \sigma_2{}^2 = E(Y^2) - \mu_2{}^2 = \frac{11}{144}.$$

The covariance of X and Y is

$$E(XY) - \mu_1\mu_2 = \int_0^1\int_0^1 xy(x+y)dx\,dy - \left(\frac{7}{12}\right)^2 = -\frac{1}{144}.$$

Accordingly, the correlation coefficient of X and Y is

$$\rho = \frac{-\frac{1}{144}}{\sqrt{\left(\frac{11}{144}\right)\left(\frac{11}{144}\right)}} = -\frac{1}{11}.$$

***Example* 2.** Let the continuous-type random variables *X* and Y have the joint p.d.f.

$$f(x,y) = \begin{cases} e^{-y}, & 0<x<y<\infty \\ 0 & elsewhere. \end{cases}$$

The moment-generating function of this joint distribution is

$$M(t_1, t_2) = \int_0^{\infty} \int_0^{\infty} \exp(t_1 x + t_2 y - y)\, dy\, dx$$

$$= \frac{1}{(1 - t_1 - t_2)(1 - t_2)}$$

Provided $t_1 + t_2 < 1$ and $t_2 < 1$. For this distribution Equations

$$\sigma_1{}^2 = E(X^2) - \mu_1{}^2 = \frac{\partial^2 M(0,0)}{\partial t_1^2} = \mu_1^2$$

Becomes,

$$\therefore \mu_1 = 1, \mu_2 = 2, \sigma_1{}^2 = 1, \sigma_2{}^2 = 2 \; and \; E[(X - \mu_1)(Y - \mu_2)] = 1.$$

Furthermore, the moment-generating functions of the marginal distributions of X and Y are, respectively,

$$M(t_1, 0) = \frac{1}{1 - t_1}, \quad t_1 < 1,$$

$$M(0, t_2) = \frac{1}{(1 - t_2)^2}, \quad t_2 < 1.$$

These moment-generating functions are, of course, respectively, those of the marginal probability density functions

$$f_1(x) = \int_x^{\infty} e^{-y} dy = e^{-x}, \quad 0 < x < \infty$$

Zero elsewhere, and

$$f_2(y) = e^{-y} \int_x^{\infty} dy = ye^{-y}, \quad 0 < y < \infty$$

Zero elsewhere.

2.4 Stochastic Independence

Let X_1 and X_2denote random variables of either the continuous or the discrete type which have the joint p.d.f. f (x_1,x_2) and marginal probability density functions.$f_1(x_1)$ and .$f_2(x_2)$respectively. The joint p.d.f. f (x_1,x_2) as

$$f(x_1,x_2) = f(x_2|x_1)f_1(x_1).$$

Definition: Let the random variables X_1 and X_2have the joint p.d.f.$f(x_1,x_2)$ and the marginal probability density functions $f_1(x_1)$ and$f_2(x_2)$,respectively. The random variables X_1 and $X_2$are said to be stochastically independent if, and only if, $f(x_1,x_2) = f_1(x_1)f_2(x_2)$.Random variables that are not stochastically independent are said to be stochastically dependent.

***Example* 1**. Let the joint p.d.f. of X_1 and X_2 be

$$f(x_1,x_2) = \begin{cases} x_1 + x_2, & 0 < x_1 < 1, 0 < x_2 < 1 \\ 0 & elsewhere. \end{cases}$$

It will be shown that X_1 and X_2 are stochastically dependent. Here the marginal probability density functions are

$$f_1(x_1) = \int_{-\infty}^{\infty} f(x_1,x_2)dx_2 = \int_0^1 (x_1 + x_2)dx_2 = x_1 + \frac{1}{2}, \quad 0 < x_1 < 1$$
$$= 0 \qquad elsewhere,$$

and

$$f_2(x_2) = \int_{-\infty}^{\infty} f(x_1,x_2)dx_2 = \int_0^1 (x_1 + x_2)dx_1 = \frac{1}{2} + x_2, \quad 0 < x_1 < 1,$$

$$= 0 \qquad elsewhere.$$

Since $f(x_1,x_2) \not\equiv f_1(x_1)f_2(x_2)$, the random variables X_1 and X_2 are stochastically dependent.

The following theorem makes it possible to assert, without computing the marginal probability density functions, that the random variables X_1 and X_2 of Example 1 are stochastically dependent.

Theorem 1. Let the random variablesX_1 and X_2 have the joint p.d.f.$f(x_1,x_2)$.Then X_1 and $X_2$are stochastically independent if and only if$f(x_1,x_2)$can be written as a product of a nonnegative function of X_1 alone and a nonnegative function of X_2 alone. That is,

$$f(x_1,x_2) \equiv g(x_1)h(x_2),$$

where $g(x_1) > 0$, $x_1 \epsilon A_1$, zero elsewhere, and $h(x_2) > 0, x_2 \epsilon A_2$, zero elsewhere.

Proof.

If X_1 and X_2 are stochastically independent, then$f(x_1,x_2) \equiv f_1(x_1)f_2(x_2)$, where $f_1(x_1)$ *and* $f_2(x_2)$are the marginal probability densityfunctions X_1 and X_2 of respectively. Thus, the condition is $f(x_1,x_2) \equiv g(x_1)h(x_2)$,fulfilled.

Conversely, if $f(x_1, x_2) \equiv g(x_1)h(x_2)$ then, for random variables of the continuous type, we have

$$f_1(x_1) = \int_{-\infty}^{\infty} g(x_1)h(x_2)dx_2 = g(x_1)\int_{-\infty}^{\infty} h(x_2)dx_2 = c_1 g(x_1)$$

And

$$f_2(x_2) = \int_{-\infty}^{\infty} g(x_1)h(x_2)dx_1 = h(x_2)\int_{-\infty}^{\infty} g(x_1)dx_1 = c_2 h(x_2)$$

Where c_1 and c_2 are constants, not functions of x_1 *and* x_2. Moreover, $c_1c_2 = 1$ because

$$1 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x_1)h(x_2)dx_1dx_2 = \left[\int_{-\infty}^{\infty} g(x_1)dx_1\right]\left[\int_{-\infty}^{\infty} h(x_2)dx_2\right] = c_2c_1.$$

The results imply that

$$f(x_1, x_2) \equiv g(x_1)h(x_2) \equiv c_1 g(x_1)c_2 h(x_2) \equiv f_1(x_1)f_2(x_2).$$

Accordingly, X_1 and X_2 are stochastically independent.

From the above example 1 we see that the joint p.d.f.

$$f(x_1, x_2) = \begin{cases} x_1 + x_2, & 0 < x_1 < 1, 0 < x_2 < 1 \\ 0 & elsewhere. \end{cases}$$

cannot be written as the product of a nonnegative function of x_1 alone and a nonnegative function of x_2 alone. Accordingly, X_1 and X_2 are stochastically dependent.

Theorem 2. If X_1 and X_2 are stochastically independent random variables with marginal probability density functions $f_1(x_1)$ *and* $f_2(x_2)$ respectively, then

$$\Pr(a < X_1 < b, c < X_2 < d) = \Pr(a < X_1 < b)\Pr(c < X_2 < d)$$

for every a < b and c < d, where a, b, c and d are constants.

Proof.
From the stochastic independence of X_1 and X_2, the joint p.d.f. of X_1 and X_2 is $f_1(x_1)f_2(x_2)$. Accordingly, in the continuous case,

$$\Pr(a < X_1 < b, c < X_2 < d) = \int_a^b \int_c^d f_1(x_1)f_2(x_2)\, dx_2 dx_1$$

$$= \left[\int_a^b f_1(x_1)dx_1\right]\left[\int_c^d f_2(x_2)dx_2\right]$$

$$= \Pr(a < X_1 < b)\Pr(c < X_2 < d);$$

or, in discrete case,

$$\Pr(a < X_1 < b, c < X_2 < d) = \sum_{a<X_1<b}\sum_{c<X_2<d} f_1(x_1)f_2(x_2)$$

$$= \left[\sum_{a<X_1<b} f_1(x_1)\right]\left[\sum_{c<X_2<d} f_2(x_2)\right]$$

$$= \Pr(a < X_1 < b)\Pr(c < X_2 < d),$$

Example 3. In Example 1, X_1 and X_2were found to be stochastically dependent. There, in general,

$$\Pr(a < X_1 < b, c < X_2 < d) \neq \Pr(a < X_1 < b)\Pr(c < X_2 < d).$$

For instance,

$$\Pr\left(0 < X_1 < \frac{1}{2}, 0 < X_2 < \frac{1}{2}\right) = \int_0^{\frac{1}{2}}\int_0^{\frac{1}{2}}(x_1 + x_2)\,dx_1dx_2 = \frac{1}{8},$$

whereas

$$\Pr\left(0 < X_1 < \frac{1}{2}\right) = \int_0^{\frac{1}{2}}\left(x_1 + \frac{1}{2}\right)dx_1 = \frac{3}{8}$$

and

$$\Pr\left(0 < X_2 < \frac{1}{2}\right) = \int_0^{\frac{1}{2}}\left(\frac{1}{2} + x_2\right)dx_2 = \frac{3}{8}.$$

Theorem 3. Let $X_1$and X_2 denote random variables have the joint p.d.f.$f(x_1, x_2)$ marginal probability density functions$f_1(x_1)$ *and* $f_2(x_2)$,respectively. Furthermore, let $M(t_1, t_2)$ denote the moment-generating function of the distribution. Then $X_1$and X_2 are stochastically independent if and only if $M(t_1, t_2) = M(t_1, 0)M(0, t_2)$.

Proof.

If $X_1$and X_2 are stochastically independent, then

$$M(t_1, t_2) = E(e^{t_1x_1+t_2x_2})$$

$$= E(e^{t_1x_1}e^{t_2x_2})$$

$$= E(e^{t_1x_1})E(e^{t_2x_2})$$

$$= M(t_1, 0)M(0, t_2).$$

Thus the stochastic independence of X_1 and X_2 implies that the moment-generating function of the joint distribution factors into the product of the moment-generating functions of the two marginal distributions.

Suppose next that the moment-generating function of the joint distribution of X_1 and $X_2$is given by$M(t_1, t_2) = M(t_1, 0)M(0, t_2)$.Now$X_1$ has the unique moment-generating function which, in the continuous case, is given by

$$M(t_1, 0) = \int_{-\infty}^{\infty} e^{t_1x_1} f_1(x_1)dx_1.$$

Similarly, the unique moment-generating function of X_2, in the continuous case, is given by

$$M(0, t_2) = \int_{-\infty}^{\infty} e^{t_2x_2} f_2(x_2)dx_2.$$

Thus we have

$$M(t_1, 0)M(0, t_2) = \left[\int_{-\infty}^{\infty} e^{t_1x_1} f_1(x_1)dx_1\right]\left[\int_{-\infty}^{\infty} e^{t_2x_2} f_2(x_2)dx_2\right]$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{t_1x_1+t_2x_2} f_1(x_1)f_2(x_2)dx_1dx_2 .$$

We are given that $M(t_1, t_2) = M(t_1, 0)M(0, t_2)$; so

$$M(t_1, t_2) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{t_1x_1+t_2x_2} f_1(x_1)f_2(x_2)dx_1dx_2 .$$

But $M(t_1, t_2)$ is the moment-generating function of X_1 and X_2. Thus also

$$M(t_1, t_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{t_1 x_1 + t_2 x_2} f_1(x_1) f_2(x_2) dx_1 dx_2 .$$

The uniqueness of the moment-generating function implies that the two distributions of probability that are described by $f_1(x_1)f_2(x_2)$ and $f(x_1, x_2)$ are the same. Thus

$$f(x_1, x_2) = f_1(x_1) f_2(x_2).$$

That is $M(t_1, t_2) = M(t_1, 0)M(0, t_2)$, then X_1 and X_2 are stochastically independent.

Some Special Distribution

2.5 The Binomial, Trinomial and Multinomial Distribution

If n is a positive integer, that

$$(a + b)^n = \sum_{x=0}^{n} \binom{n}{x} b^x a^{n-x}.$$

Consider the function defined by

$$f(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & x = 0,1,2, \dots, n, \\ 0 & elsewhere \end{cases},$$

where n is a positive integer $0 < p < 1$. Under these conditions it is clear that $f(x) \geq 0$ and that

$$\sum_x f(x) = \sum_{x=0}^{n} \binom{n}{x} p^x (1-p)^{n-x}$$

$$= [(1-p) + p]^n = 1.$$

That is, *f(x)* satisfies the conditions of being a p.d.f. of a random variable *X* of the discrete type. A random variable *X* that has a p.d.f. of the form of *f(x)* is said to have a *binomial distribution,* and any such *f(x)* is called a *binomial p.d.f.* A binomial distribution will be denoted by the symbol b(n,p).

If we say that X is $b\left(5,\frac{1}{3}\right)$, we mean that X has the binomial p.d.f.

$$f(x) = \begin{cases} \binom{5}{x}\left(\frac{1}{3}\right)^x\left(\frac{2}{3}\right)^{5-x'}, & x = 0,1,\dots,5, \\ 0 & elsewhere. \end{cases}$$

***Example* 1**. The binomial distribution with p.d.f.

$$f(x) = \begin{cases} \binom{7}{x}\left(\frac{1}{2}\right)^x\left(1-\frac{1}{2}\right)^{7-x}, & x = 0,1,2,\dots,7, \\ 0 & elsewhere. \end{cases}$$

has the moment generation function

$$M(t) = \left(\frac{1}{2}+\frac{1}{2}e^t\right)^7,$$

Has mean $\mu = np = \frac{7}{2}$, and has variance $\sigma^2 = np(1-p) = \frac{7}{4}$. Furthermore, if X is the random variable with this distribution, we have

$$\Pr(0 \le X \le 1) = \sum_{x=0}^{1} f(x) = \frac{1}{128} + \frac{7}{128} = \frac{8}{128}$$

and

$$\Pr(X = 5) = f(5) = \frac{7!}{5!\,2!}\left(\frac{1}{2}\right)^5\left(\frac{1}{2}\right)^2 = \frac{21}{128}.$$

***Example* 2**. If the moment-generating function of a random variable X is

$$M(t) = \left(\frac{2}{3}+\frac{1}{3}e^t\right)^5,$$

Then X has a binomial distribution with n = 5 and p = $\frac{1}{3}$; that is, the p.d.f. of X is

$$f(x) = \begin{cases} \binom{5}{x}\left(\frac{1}{3}\right)^x\left(\frac{2}{3}\right)^{5-x}, & x = 0,1,2,\dots,5, \\ 0 & elsewhere. \end{cases}$$

here $\mu = np = \frac{5}{3}$ and $\sigma^2 = np(1-p) = \frac{10}{9}$.

***Example* 3**. Consider a sequence of independent repetitions of a random experiment with constant probability p of success. Let the random variable Y denote the total number of failures in this sequence before the r^{th} success; that is, Y + r is equal to the number of trials necessary to produce exactly r successes. Here r is a fixed positive integer. To determine the p.d.f. of Y, let y be an element of $\{y; y = 0,1,2,\ldots,\}$. Then, by the multiplication rule of probabilities, Pr (Y = *y) = g(y)* is equal to the product of the probability

$$\binom{y+r-1}{r-1} p^{r-1}(1-p)^y$$

of obtaining exactly r - 1 successes in the first y + r - 1 trials and the probability p of a success on the $(y+r)^{th}$ trial. Thus the p.d.f. *g(y)* of Y is given by

$$g(y) = \begin{cases} \binom{y+r-1}{r-1} p^{r-1}(1-p)^y, & y = 0,1,2,.., \\ 0 & elsewhere. \end{cases}$$

A distribution with a p.d.f. of the form g(y) is called a negative binomial distribution; and any such g(y) is called a negative binomial p.d.f. The distribution derives its name from the fact that g(y) is a general term in the expansion of Pr[l - (1 - P)]$^{-r}$. It is left as an exercise to show that the moment-generating function of this distribution is $M(t) = p^r[1-(1-p)e^t]^{-r}$,for $t < -\ln(1-p)$.If r = 1, then Y has the p.d.f.

$$g(y) = p(1-p)^y, \;\; y = 0,1,2, \ldots,$$

Zero elsewhere, and the moment-generating function $M(t) = p^r[1-(1-p)e^t]^{-1}$. In the special case r = 1, we say that Y has a geometric distribution.

2.6 The Poisson Distribution

The series

$$1 + m + \frac{m^2}{2!} + \frac{m^3}{3!} + \cdots = \sum_{x=0}^{\infty} \frac{m^x}{x!}$$

converges, for all values of m, to e^m. Consider the function *f(x)* defined by

$$f(x) = \begin{cases} \frac{m^x e^{-m}}{x!}, & x = 0,1,2, \ldots, \\ 0 & elsewhere. \end{cases}$$

where m > 0. Since m > 0, then $f(x) \geq 0$ and

$$\sum_{x} f(x) = \sum_{x=0}^{\infty} \frac{m^x e^{-m}}{x!} = e^{-m} \sum_{x=0}^{\infty} \frac{m^x}{x!} = e^{-m} e^{m} = 1;$$

that is,f(x) satisfies the conditions of being a p.d.f. of a discrete type of random variable. A random variable that has a p.d.f. of the form f(x)is said to have a Poisson distribution, and any such f(x) is called a Poisson p.d.f.

***Example* 1**. Suppose that *X* has a Poisson distribution with μ = 2. Then the p.d.f. of *X* is

$$f(x) = \begin{cases} \dfrac{2^x e^{-2}}{x!}, & x = 0,1,2, \dots, \\ 0 & elsewhere. \end{cases}$$

The variance of this distribution is $\sigma^2 = \mu = 2$. If we wish to compute Pr$(1 \leq X)$, we have

$$\Pr(1 \leq X) = 1 - \Pr\,(X = 0)$$

$$= 1 - f(0) = 1 - e^{-2} = 0.865,$$

approximately.

***Example* 2**. If the moment-generating function of a random variable *X* is

$$M(t) = e^{4(e^t - 1)},$$

Then X has a poisson distribution μ = 4. Accordingly, by way of example,

$$\Pr(X = 3) = \frac{4^3 e^{-4}}{3!} = \frac{32}{3} e^{-4}$$

Or

$$\Pr(X = 3) = \Pr(X \leq 3) - \Pr(X \leq 2) = 0.433 - 0.238 = 0.195.$$

***Example* 3**. Let the probability of exactly one blemish in 1 foot of wire be about $\frac{1}{1000}$ and let the probability of two or more blemishes in that length be, for all practical purposes, zero. Let the random variable *X* be the number of blemishes in 3000 feet of wire. If we assume the stochastic
independence of the numbers of blemishes in non overlapping intervals, then the postulates of the Poisson process are approximated, with $\lambda = \frac{1}{1000}$and *w* = 3000. Thus *X* has an approximate Poisson distribution with mean$3000\left(\frac{1}{1000}\right) = 3$. For

example, the probability that there are exactly five blemishes in 3000 feet of wire is

$$\Pr(X = 5) = \frac{3^5 e^{-3}}{5!}$$

and

$$\Pr(X = 5) = \Pr(X \leq 5) - \Pr(x \leq 4) = 0.101,$$

approximately.

Unit III

3.1 The Gamma and Chi-Square Distributions

The Gamma function of X is

$$\Gamma(\alpha) = \int_0^\infty y^{\alpha-1} e^{-y} dy$$

If α = 1, Clearly

$$\Gamma(1) = \int_0^\infty e^{-y} dy = 1.$$

If α > 1, an integration by parts show that

$$\Gamma(\alpha) = (\alpha-1)\int_0^\infty y^{\alpha-2} e^{-y} dy = (\alpha-1)\Gamma(\alpha-1).$$

Accordingly, if α is a positive integer greater than 1,

$$\Gamma(\alpha) = (\alpha-1)(\alpha-2)\ldots(3)(2)(1)\Gamma(1) = (\alpha-1)!.$$

Since $\Gamma(1) = 1$, this suggests that we take 0! = 1 , as we have done.
In the integral that defines $\Gamma(\alpha)$. Let us introduce a new variable x by writing $y = \frac{x}{\beta}$, where $\beta > 0$. Then

$$\Gamma(\alpha) = \int_0^\infty \left(\frac{x}{\beta}\right)^{\alpha-1} e^{-x/\beta} \left(\frac{1}{\beta}\right) dx,$$

Or equivalently,

$$1 = \int_0^\infty \frac{(x)^{\alpha-1}}{\Gamma(\alpha)\beta^\alpha} e^{-x/\beta} dx.$$

Since α > 0, β > 0 and $\Gamma(\alpha) > 0$, we see that

$$f(x) = \begin{cases} \frac{(x)^{\alpha-1}}{\Gamma(\alpha)\beta^\alpha} e^{-x/\beta}, & 0 < x < \infty, \\ 0 & elsewhere. \end{cases}$$

is a p.d.f. of a random variable of the continuous type.

***Example* 1**. Let *X* be a random variable such that

$$E(X^m) = \frac{m+3}{3!}3^m, \qquad m = 1,2,3, ...$$

Then the moment generating function of X is given by the series

$$M(t) = 1 + \frac{4!\,3}{3!\,1!}t + \frac{5!\,3^2}{3!\,2!}t^2 + \frac{6!\,3^3}{3!\,3!}t^3 + \cdots.$$

This however. is the Maclaurin's series for$(1-3t)^{-4}$provided that $-1 < 3t < 1$. Accordingly,X has a gamma distribution with $a = 4$ and $\beta = 3$.

***Example* 2**. If X has the moment-generating function $M(t) = (1-2t)^{-8}, t > \frac{1}{2}$. Then X is $X^2(16)$.

If the random variable X is$X^2(r)$, then, with $c_1 < c_2$, we have

$$\Pr(c_1 \le X \le c_2) = \Pr(X \le c_2) - \Pr(X \le c_1),$$

Since $\Pr(X = c_1) = 0$. To compute such a probability, we need the value of an integral like

$$\Pr(X \le x) = \int_0^x \frac{1}{\Gamma(r/2)2^{r/2}} w^{r/2-1} e^{-w/2} dw.$$

***Example* 3**. Let X have a gamma distribution with $\alpha = r/2$, where r is a positive integer, and $\beta > 0$. Define the random variable $Y = \frac{2X}{\beta}$. We seek the p.d.f. of Y. Now the distribution function of Y is

$$G(y) = \Pr(Y \le y) = \Pr\left(x \le \frac{\beta y}{2}\right).$$

If $y \le 0$, then G(y) = 0; but if y > 0, then

$$G(y) = \int_0^{\beta y/2} \frac{1}{\Gamma(r/2)\beta^{r/2}} x^{r/2-1} e^{-x/\beta} dx.$$

Accordingly, the p.d.f. of Y is

$$g(y) = G'(y) = \frac{1}{\Gamma(r/2)\beta^{r/2}} (\beta y/2)^{r/2-1}\, e^{-y/2}$$

$$= \frac{1}{\Gamma(r/2)2^{r/2}} (y)^{r/2-1} e^{-y/2}$$

If y > 0. That is, Y is $X^2(r)$.

3.2 The Normal Distribution

Consider the integral

$$I = \int_{-\infty}^{\infty} exp(y^2/2)\, dy.$$

This integral exists because the integrand is a positive continuous function which is bounded by an integrable function; that is,

$$0 < exp(y^2/2) < \exp(-|y| + 1), \quad -\infty < y < \infty,$$

and

$$\int_{-\infty}^{\infty} \exp(-|y| + 1)\, dy = 2e.$$

To evaluate the integral I, we note that I > 0 and that I^2 may be written

$$I^2 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \exp\left(\frac{-y^2 + z^2}{2}\right) dydz.$$

***Example* 1**. If *X* has the moment-generating function

$$M(t) = e^{2t+32t^2},$$

Then X has a normal distribution $\mu = 2, \sigma^2 = 64.$

Thus, if we say that the random variable X is n(0,1), we mean that X has a normal distribution
with mean $\mu = 0$and variance$\sigma^2 = 1$, so that the p.d.f. of X is

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \quad -\infty < x < \infty$$

Moreover, if

$$M(t) = e^{t^2/2}.$$

then X is n(0,1).

The graph of

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right], \qquad -\infty < x < \infty,$$

is seen (1) to be symmetric about a vertical axis through $x = \mu$,(2) to have its maximum of $\frac{1}{\sigma\sqrt{2\pi}}$at $x = \mu$and (3) to have the x-axis as a horizontal asymptote. It should be verified that (4) there are points of inflection at $x = \mu \pm \sigma$.

Theorem 1. If the random variable X is n (μ, σ^2), $\sigma^2 > 0$, then the random variable
$W = \frac{X-\mu}{\sigma}$is n(0,1).

Proof.

The distribution function *G(w)* of *W* is, since σ> 0,

$$G(w) = \Pr\left(\frac{X-\mu}{\sigma} \leq w\right) = \Pr(X \leq w\sigma + \mu).$$

That is,

$$G(w) = \int_{-\infty}^{w\sigma+\mu} \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] dx.$$

If we change the variable of integration by writing$y = \frac{(x-\mu)}{\sigma}$, then

$$G(w) = \int_{-\infty}^{w} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy.$$

Accordingly, the p.d.f. g(w) = g'(w) of the continuous – type random variable W is

$$g(w) = \frac{1}{\sqrt{2\pi}} e^{-w^2/2}, \quad -\infty < w < \infty.$$

Thus W is n(0,1), which is the desired result.

Theorem 2. If the random variable X is n (μ, σ^2), $\sigma^2 > 0$,then therandom variable $V = \frac{(X-\mu)^2}{\sigma^2}$
is $X^2(1)$.

Proof.

Because V = W^2, where$W = \frac{X-\mu}{\sigma}$is n(0,1) the distribution function *G(v)* of *V* is, for$v \geq 0$,

$$G(v) = \Pr(W^2 \leq v) = \Pr(-\sqrt{v} \leq W \leq \sqrt{v}).$$

If we change the variable of integration by writing $w = \sqrt{y}$, then

$$G(v) = \int_0^v \frac{1}{\sqrt{2\pi}\sqrt{y}} e^{-y/2} dy, \quad 0 \leq v.$$

Hence the p.d.f.$g(v) = G'(v)$ of the continuous – type random variable V is

$$g(v) = \begin{cases} \dfrac{1}{\sqrt{2\pi}} v^{1/2-1e^{-v/2}}, & 0 < v < \infty, \\ 0 & elsewhere. \end{cases}$$

Since g(v) is a p.d.f. and hence

$$\int_0^\infty g(v) dv = 1$$

it must be that $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$ and thus V is $X^2(1)$.

3.3 The Bivariate Normal Distribution

Let us investigate the function

$$f(x,y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{-q/2}, \qquad -\infty < x < \infty, -\infty < y < \infty,$$

where, with $\sigma_1 > 0, \sigma_2 > 0$, and $-1 < y < 1$,

$$q = \frac{1}{1-\rho^2}\left[\left(\frac{x-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x-\mu_1}{\sigma_1}\right)\left(\frac{y-\mu_2}{\sigma_2}\right) + \left(\frac{y-\mu_2}{\sigma_2}\right)^2\right]$$

At this point we do not know that the constants$\mu_1, \mu_2, {\sigma_1}^2, {\sigma_2}^2 and\ \rho$ represent parameters of a distribution. As a matter of fact, we do notknow that$f(x, y)$has the properties of a joint p.d.f. It will now be shownthat:

(a) f(x, y) is a joint p.d.f.
(b) X is n($\mu_1, {\sigma_1}^2$) and Y is n($\mu_2, {\sigma_2}^2$).
(c) ρ is the correlation coefficient of X and Y.

A joint p.d.f. of this form is called a bivariate normal p.d.f., and therandom variables X and Yare said to have a bivariate normal distribution.

***Example* 1**. Let us assume that in a certain population of married couples the height X_1 of the husband and the height X_2 of the wife have a bivariatenormal distribution with parameters
μ_1 = 5.8 feet, μ_2= 5.3 feet, σ_1 =σ_2 =0.2 foot, andρ = 0.6. The conditional p.d.f. of X_2 , given X_1 = 6.3, is normal with mean5.3 +(0.6)(6.3 -5.8) =5.6 and standard deviation (0.2) $\sqrt{(1-0.36)} = 0.16$. Accordingly, given that the height of the husband is 6.3 feet, theprobability that his wife has a height between 5.28 and 5.92 feet is

$$\Pr(5.28 < X_2 < 5.92|x_1 = 6.3) = N(2) - N(-2) = 0.954.$$

The moment-generating function of a bivariate normal distributioncan be determined as follows. We have

$$M(t_1, t_2) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{t_1x+t_2y} f(x, y)dxdy$$

$$= \int_{-\infty}^{\infty} e^{t_1x} f_1(x)\,[\int_{-\infty}^{\infty} e^{t_2y} f(y|x)dy]dx$$

for all real values of $t_1$and t_2 . The integral within the brackets is the moment generating function of the conditional p.d.f.$f(y|x)$. Since $f(y|x)$is a normal p.d.f. with mean $\mu_2 + \rho\left(\frac{\sigma_2}{\sigma_1}\right)(x-\mu_1)$ and variance ${\sigma_2}^2(1-\rho^2)$, then

$$\int_{-\infty}^{\infty} e^{t_2y} f(y|x)dy = \exp\left\{t_2\left[\mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x-\mu_1)\right] + \frac{{t_2}^2{\sigma_2}^2(1-\rho^2)}{2}\right\}.$$

Accordingly, M(t_1, t_2) can be written in the form

$$\exp\left\{t_2\mu_2 - t_2\,\rho\frac{\sigma_2}{\sigma_1}(\mu_1) + \frac{{t_2}^2{\sigma_2}^2(1-\rho^2)}{2}\right\}\int_{-\infty}^{\infty} \exp\,[(t_1 + t_2\,\rho\left(\frac{\sigma_2}{\sigma_1}\right)(x)] f_1(x)dx.$$

But $E(e^{tX}) = \exp\,[\mu_1 t + \frac{\sigma_1{}^2 t}{2}]$ for all the real values of t. Accordingly, if we set $t = t_1 + t_2\rho\left(\frac{\sigma_2}{\sigma_1}\right)$, we see that $M(t_1, t_2)$ is given by

$$\exp\left\{t_2\mu_2 - t_2\,\rho\frac{\sigma_2}{\sigma_1}(\mu_1) + \frac{t_2{}^2\sigma_2{}^2(1-\rho^2)}{2} + \mu_1\left(t_1 + t_2\rho\left(\frac{\sigma_2}{\sigma_1}\right)\right) + \sigma_1{}^2\frac{\left(t_1 + t_2\rho\left(\frac{\sigma_2}{\sigma_1}\right)\right)^2}{2}\right\}$$

Or, equivalently,

$$M(t_1, t_2) = \exp\left(\mu_1 t_1 + \mu_2 t_2 + \frac{\sigma_1{}^2 t_1{}^2 + 2\rho\sigma_1\sigma_2 t_1 t_2 + \sigma_2{}^2 t_2{}^2}{2}\right).$$

It is interesting to note that if, in this moment-generating function $M(t_1, t_2)$, the correlation coefficient ρ is set equal to zero, then

$$M(t_1, t_2) = M(t_1, 0)M(0, t_2).$$

Thus X and Y are stochastically independent when $\rho = 0$. If, conversely, $M(t_1, t_2) = M(t_1, 0)M(0, t_2)$, we have $e^{\rho\sigma_1\sigma_2 t_1 t_2} = 1$. Since each of σ_1 *and* $\sigma_2$is positive, then $\rho = 0$.

Exercise

1. Prove that $P(C_1 \cap C_2 \cap C_3 \cap C_4) = P(C_1)P(C_2/C_1)P(C_3/C_1 \cap C_2)P(C_4/C_1 \cap C_2 \cap C_3)$.
2. A hand of 13 cards is to be dealt at random and without replacement from an ordinary deck of playing cards. Find the conditional probability that there are at least three kings in the hand relative to the hypothesis that the hand contains at least two kings.
3. A bowl contains 10 chips. Four of the chips are red, 5 are white, and 1 is blue. If 3 chips are taken at random and without replacement, compute the conditional probability that there is 1 chip of each color relative to the hypothesis that thee is exactly 1 red chip among the 3.
4. Let X_1 and X_2 have the joint p.d.f.$f(x_1, x_2) = x_1 + x_2,\ \ 0 < x_1 < 1, 0 < x_2 < 1$, *zero elsewhere* . Find the conditional mean and variance of X_2 given by $X_1 = x_1, 0 < x_1 < 1$.

5. If X_1 and X_2 are random variables of the discrete type having p.d.f. $f(x_1, x_2) = \frac{x_1+2x_2}{18}$, $(x_1, x_2) = (1,1), (1,2), (2,1), (2,2), zero\ elsewhere.$ Determine the conditional mean and variance X_2 given by $X_1 = x_1, x_1 = 1\ or\ 2.$
6. Let $f(x,y) = 2, 0 < x < 1, 0 < y < 1, zero\ elsewhere,$ be the joint p.d.f. of X and Y. Show that the conditional means are, respectively, $\frac{1+x}{2}, 0 < x < 1,$ $and \frac{y}{2}, 0 < y < 1.$ Show that the correlation coefficient of X and Y is $\rho = 1/2.$
7. If the random variables X_1, X_2 have the joint p.d.f. $f(x_1, x_2) = 2e^{-x_1-x_2},\ 0 < x_1 \infty, 0 < x_2 < \infty, zero\ elsewhere,$ show that X_1 and X_2 are stochastically dependent.
8. If the moment-generating function of a random variable X is $(\frac{1}{3} + \frac{2}{3^{et}})^5, fin$ $\Pr(X = 2\ or 3).$
9. If X is b(n,p), show that $E\left(\frac{X}{n}\right) = p\ and\ E\left[\left(\frac{X}{n-p}\right)^2\right] = \frac{p(1-p)}{n}$
10. Let y be the number of success in n independent repetitions of a random experiment having the probability of success p = 2/3. If n = 3 , compute Pr(2<Y); if n =5, compute Pr(3≤Y).
11. Let X be b(2,p) and let Y be b(4,p). If $\Pr(X \geq 1) = 5/9$, find $\Pr(Y \geq 1).$
12. If a fair coin tossed at random five independent times, find the conditional probability of five heads relative to the hypothesis that are at least four heads.
13. If the random variable X has a poisson distribution such that $\Pr(X = 1) = \Pr(X = 2)$, find Pr(X=4).
14. Compute the measures of skewness and kurtosis of the poisson distribution with mean μ.
15. If X is $X^2(5)$, determine the constants c and d so that Pr(c<X<d) = 0.95 and Pr(X<c)=0.025.
16. Compute the measures of skewness and kurtosis of a gamma distribution with parameters α and β.

DISTRIBUTIONS OF FUNCTIONS OF RANDOM VARIABLES

3.4 Sampling Theory

Definition 1. A function of one or more random variables that does not depend upon any *unknown* parameter is called a *statistic.*

Definition 2. Let $X_1, X_2, \ldots X_n$ denote n mutually stochastically independent random variables, each of which has the same but possibly unknown p.d.f. f(x); that is, the probability density functions of $X_1, X_2, \ldots X_n$ are, respectively, $f_1(x_1) = f(x_1), f_2(x_2) = f(x_2), \ldots, f_n(x_n) = f(x_n)$ so that the joint p.d.f. is $f(x_1)f(x_2) \ldots f(x_n)$. The random variables $X_1, X_2, \ldots X_n$ are then said to constitute a random sample from a distribution that has p.d.f. f(x).

Definition 3. Let $X_1, X_2, \ldots X_n$ denote a random sample of size *n* from a given distribution. The statistic

$$X = \frac{X_1 + X_2 + \cdots + X_n}{n} = \sum_{i=1}^{n} \frac{X_i}{n},$$

is called the mean of the random sample, and the statistic

$$S^2 = \sum_{i=1}^{n} \frac{(X_i - \bar{X})^2}{n} = \sum_{i=1}^{n} \frac{X_i^2}{n} - \bar{X}^2$$

is called the variance of the random sample.

***Example* 1**. Let the random variable Y be distributed uniformly over the unit interval $0 < y < 1$; that is, the distribution function of Y is

$$G(y) = \begin{cases} 0, & y \le 0, \\ y, & 0 < y < 1, \\ 1, & 1 \le y. \end{cases}$$

Suppose that *F(x)* is a distribution function of the continuous type which is strictly increasing when $0 < F(x) < 1$. If we define the random variable *X* by the relationship Y = *F(X),* we now show that X has a distribution which corresponds to F(x). If 0 < F(x) < 1, the inequalities
$X \le x$ and $F(X) \le F(x)$ are equivalent. Thus, with 0 < F(x) < 1, the distribution function of
X is

$$\Pr(X \le x) = \Pr[F(X) \le F(x)] = \Pr\,[Y \le F(x)]$$

because Y = F(x). However, $\Pr(Y \le y) = G(y)$, so we have

$$\Pr(X \le x) = G[F(x)] = F(x), \quad 0 < F(x) < 1.$$

That is the distribution function of X is F(x).

This result permits us to *simulate* random variables of different types.

3.5 Transformations of Variables of the Discrete Type

An alternative method of finding the distribution of a function of one or more random variables is called the *change of variable technique.*

Let X is have the poisson p.d.f.

$$f(x)=\begin{cases}\frac{\mu^x e^{-\mu}}{x!}, & x=0,1,2,\dots,\\ 0 & elsewhere.\end{cases}$$

Let A denote the space A = {x; x = 0,1,2, ... },so that d is the set where f(x) > 0. Define a new random variable Y by Y = 4X. We wish to find the p.d.f. of Y by the change-of-variable
technique. Let y = 4x. We call y = 4x a transformation from x to y,and we say that the transformation maps the space A onto the space B = {y; y = 0, 4, 8, 12, ... }. The space B is obtained by transforming each point in d in accordance with y = 4x.

The p.d.f. g(y) of the discrete type

$$g(y)=\Pr(Y=y)=\Pr\left(X=\frac{y}{4}\right)=\frac{\mu^{y/4}e^{-\mu}}{(\frac{y}{4})!},\quad y=0,4,8,\dots,.$$

$$=0\quad elsewhere.$$

***Example* 1.** Let X have the binomial p.d.f.

$$f(x)=\begin{cases}\frac{3!}{x!\,(3-x)!}\left(\frac{2}{3}\right)^x\left(\frac{1}{3}\right)^{3-x}, & x=0,1,2,3,\\ 0 & elsewhere.\end{cases}$$

We seek the p.d.f. g(y) of the random variable $Y = X^2$. The transformation $Y = u(x) = x^2$ maps A= {x; x = 0, 1, 2, 3} onto B = {y; Y = 0, 1,4, 9}.In general, $Y = x^2$ does not define a one-to-one transformation; here, however, it does, for there are no negative values of x in
A = {x; x = 0, 1,2,3}.That is, we have the single-valued inverse function
$x=w(y)=\sqrt{y}\ (not\ \sqrt{-y})$, and so

$$g(y)=f(\sqrt{y})=\begin{cases}\frac{3!}{(\sqrt{y})!\,(3-\sqrt{y})!}\left(\frac{2}{3}\right)^{\sqrt{y}}\left(\frac{1}{3}\right)^{3-\sqrt{y}}, & y=0,1,4,9,\\ 0 & elsewhere.\end{cases}$$

***Example* 2**. Let X_1 and X_2 be two stochastically independent random variables that have Poisson distributions with means μ_1 and μ_2 respectively.

The joint p.d.f. of X_1 and X_2 is

$$\frac{\mu_1^{x_1}\mu_2^{x_2}e^{-\mu_1-\mu_2}}{x_1!\,x_2!}, \quad x_1 = 0,1,2,3,\ldots, x_2 = 0,1,2,3,\ldots,$$

and is zero elsewhere. Thus the space A is the set of points (x_1, x_2) where each of X_1 and $X_2$is a nonnegative integer. We wish to find the p.d.f. of $Y_1 = X_1 + X_2$. If we use the change of variable technique, we need to define a second random variable Y_2. Because Y_2 is of no interest to us, let us choose it in such a way that we have a simple one-to-one transformation. For example, take $Y_2 = X_2$. Then $y_1 = x_1 + x_2$ and $y_2 = x_2$ represent one-to-one transformation that maps A onto

$$B = \{(y_1,y_2);\ y_2 = 0,1,\ldots,y_1 \ \text{ and } \ y_1 = 0,1,2,\ldots\}.$$

Note that, if $(y_1, y_2)\epsilon\, B$, then $0 \le y_2 < y_1$.The inverse functions are givenby $x_1 = y_1 - y_2$ and $x_2 = y_2$.Thus the joint p.d.f. of Y_1 and $Y_2$is

$$g(y_1, y_2) = \frac{\mu_1^{y_1-y_2}\mu_2^{y_2}e^{-\mu_1-\mu_2}}{(y_1 - y_2)!\,y_2!}, \qquad (y_1, y_2)\epsilon B,$$

and is zero elsewhere. Consequently, the marginal p.d.f. of Y_1 is given by

$$g_1(y_1) = \sum_{y_2=0}^{y_1} g(y_1, y_2)$$

$$= \frac{e^{-\mu_1-\mu_2}}{y_1!}\sum_{y_2=0}^{y_1}\frac{y_1!}{(y_1 - y_2)!\,y_2!}\mu_1^{y_1-y_2}\mu_2^{y_2}$$

$$= \frac{(\mu_1 + \mu_2)^{y_1}e^{-\mu_1-\mu_2}}{y_1!}, \qquad y_1 = 0,1,2,\ldots,$$

And is zero elsewhere. That is, $Y_1 = X_1 + X_2$ has a poisson distribution with the parameter $\mu_1 + \mu_2$.

3.6 Transformations of Variables of the Continuous Type

***Example* 1**. Let X be a random variable of the continuous type, having p.d.f.

$$f(x) = \begin{cases} 2x, & 0 < x < 1, \\ 0 & elsewhere. \end{cases}$$

Here B is the space $\{x;\ 0 < x < I\}$, where $f(x) > 0$. Define the random variable Y by $Y = 8X^3$ and consider the transformation $y = 8x^2$.Under the transformation $Y = 8x^2$, the set B is mapped onto the set B=$\{y;\ 0 < Y < 8\}$,and, moreover, the transformation is one-to-one. For every $0 < a < b < 8$, the event $a < Y < b$ will occur when, and only when, the event $\frac{1}{2}\sqrt[3]{a} < x < \frac{1}{2}\sqrt[3]{b}$ occurs because there is a one to one correspondence between the points of a and b. Thus,

$$\Pr(a < Y < b) = \Pr\,(\frac{1}{2}\sqrt[3]{a} < X < \frac{1}{2}\sqrt[3]{b}$$

$$= \int_{\sqrt[3]{a}/2}^{\sqrt[3]{b}/2} 2x dx.$$

Let us rewrite this integral by changing the variable of integration from x to y by writing $y = 8x^3 \;\; or \; x = \frac{1}{2}\sqrt[3]{y}$. Now

$$\frac{dx}{dy} = \frac{1}{6y^{2/3}},$$

And, accordingly we have

$$\Pr(a < Y < b) = \int_a^b 2\left(\frac{\sqrt[3]{y}}{2}\right)\frac{1}{6y^{2/3}}dy$$

$$= \int_a^b \frac{1}{6y^{1/3}}dy.$$

Since this is true for ever 0 < a< b< 8, the p.d.f. g(y) of Y is the integrand; that is

$$g(y) = \begin{cases} \frac{1}{6y^{1/3}}, & 0 < y < 8, \\ 0 & elsewhere. \end{cases}$$

***Example* 2**. Let *X* have the p.d.f.

$$f(x) = \begin{cases} 1, & 0 < x < 1, \\ 0 & elsewhere. \end{cases}$$

We are to show that the random variable Y = -21n X has a chi-square distribution with 2 degrees of freedom. Here the transformation is Y = u(x) =-21n X, so that x = w(y) = $e^{-y/2}$. The space A is A = {x; 0 <x<1}, which the one-to-one transformation Y = -21n X maps onto B = {y; 0 <y<∞}.The Jacobian of the transformation is

$$J = \frac{dx}{xy} = w'(y) = \frac{1}{2}e^{-y/2}\,.$$

Accordingly, the p.d.f.g(y) of Y = -21n X is

$$g(y) = f\left(e^{-\frac{y}{2}}\right)|J| = \frac{1}{2}e^{-y/2}, \qquad 0 < y < \infty,$$

$$= 0 \ elsewhere.$$

a p.d.f. that is chi-square with 2 degree of freedom. This method of finding the p.d.f. of a function of one random variable of the continuous type will now be extended to functions of two random variables of this type. Again, only functions that define a one-to-one transformation will be considered at this time. Let $Y_1 = \mu_1(x_1, x_2)$ and $Y_2 = \mu_2(x_1, x_2)$ define a one-to-one transformation that maps a (two-dimensional) set A in the x_1x_2-plane onto a (two-dimensional) set A in the y_1y_2-plane. If we express each of X_1 and X_2 in terms of Y_1 and Y_2 we can write $x_1 = w_1(y_1, y_2), x_2 = w_2(y_1, y_2)$.The determinant of order 2,

$$\begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} \end{vmatrix}$$

is called the Jacobian of the transformation and will be denoted by the symbol J.

***Example* 4**. Let the random variable X have the p.d.f.

$$f(x) = \begin{cases} 1, & 0 < x < 1, \\ 0 & elsewhere. \end{cases}$$

and let X_1, X_2 denote a random sample from this distribution. The joint p.d.f. of X_1 and X_2 is then

$$\varphi(x_1, x_2) = f(x_1)f(x_2) = 1, \quad 0 < x_1 < 1, 0 < x_2 < 1,$$

$$= 0 \ elsewhere.$$

Consider the two random variables $Y_1 = X_1 + X_2$ and $Y_2 = X_1 + X_2$, we wish to find the joint p.d.f of Y_1 and Y_2.Here the two-dimensional space A in thex_1x_2 -plane is that of Example 3 of this section. The one-to-one transformation$y_1 = x_1 + x_2$, $y_2 = x_1 + x_2$ maps A onto the space B of that example. Moreover, the Jacobian of that transformation has been shown to be J = -1/2. Thus

$$g(y_1, y_2) = \varphi\left[\frac{1}{2}(y_1 + y_2), \frac{1}{2}(y_1 - y_2)\right]|J|$$

$$= f\left[\frac{1}{2}(y_1 + y_2)\right] f\left[\frac{1}{2}(y_1 - y_2)\right] |J|$$

$$= \frac{1}{2}(y_1, y_2) \in B,$$

$$= 0 \quad elsewhere$$

Because B is not a product space, the random variables Y_1 and Y_2 are stochastically dependent. The marginal p.d.f. of Y_1 is given by

$$g_1(y_1) = \int_{-y_1}^{y_1} \frac{1}{2} dy_2 = y_1, \quad 0 < y_1 \le 1,$$

$$= \int_{y_1-2}^{2-y_1} \frac{1}{2} dy_2 = 2 - y_1, \quad 1 < y_1 \le 2,$$

$$= 0 \quad elsewhere.$$

In a similar manner, the marginal p.d.f.$g_2(y_2)$ is given by

$$g_2(y_2) = \int_{-y_2}^{y_2+2} \frac{1}{2} dy_1 = y_2 + 1, \quad -1 < y_2 \le 0,$$

$$= \int_{y_2}^{2-y_2} \frac{1}{2} dy_1 = 1 - y_2, \quad 0 < y_2 \le 1$$

$$= 0 \quad elsewhere.$$

Example **6**. Let $Y_1 = (X_1, -X_2)$, where X_1 and X_2 are stochastically independent random variables, each being $X^2(2)$. The joint p.d.f. of X_1 andX_2 is

$$f(x_1)f(x_2) = \frac{1}{4}\exp\left(-\frac{x_1 + x_2}{2}\right), \quad 0 < x_1 < \infty, 0 < x_2 < \infty,$$

$$= 0 \qquad elsewhere.$$

Let $Y_2 = X_2$ so that $y_1 = \frac{1}{2}(x_1 - x_2), y_2 = x_2, or\ x_1 = 2y_1 + y_2. x_2 = y_2$ define a one-to-one transformation from A=$\{(x_1, x_2); 0 < x_1 < \infty, 0 < x_2 < \infty\}$ onto $B = \{(y_1, y_2);\ -2y_1 < y_2\ and\ 0 < y_2, -\infty < y_1 < \infty\}$. The Jacobian of the transformation is

$$J = \begin{vmatrix} 2 & 1 \\ 0 & 1 \end{vmatrix} = 2;$$

Hence the joint p.d.f. of Y_1 and Y_2 is

$$g(y_1, y_2) = \frac{|2|}{4} e^{-y_1 - y_2}, \quad (y_1, y_2) \epsilon\, B,$$

$$= 0 \qquad elsewhere.$$

Thus the p.d.f. of Y_1 is given by

$$g_1(y_1) = \int_{-2y_1}^{\infty} \frac{1}{2} e^{-y_1 - y_2}\, dy_2 = \frac{1}{2} e^{y_1}, \ -\infty < y_1 \le 0,$$

$$= \int_{0}^{\infty} \frac{1}{2} e^{-y_1 - y_2}\, dy_2 = \frac{1}{2} e^{-y_1}, \ 0 \le y_1 < \infty$$

or

$$g_1(y_1) = \frac{1}{2} e^{-|y_1|}, \qquad -\infty < y_1 < \infty.$$

This p.d.f. is called the double exponential p.d.f.

Unit IV

4.1 The β, *t* and F Distributions

Let W denote a random variable that is n (0,1); let V denote a random variable that is $X^2(r)$; and let W and V be stochastically independent, Then the joint p.d.f. of W and that of V or

$$\varphi(w,v) = \frac{1}{\sqrt{2\pi}} e^{-w^2/2} \frac{1}{\Gamma(\frac{r}{2})2^{r/2}} v^{r/2-1} e^{-v/2}, -\infty < w < \infty, -\infty < r < \infty,$$

$$= 0 \quad elsewhere.$$

Define a new random variable T by writing

$$T = \frac{W}{\sqrt{V/r}}$$

The change-of-variable technique will be used to obtain the p.d.f. $g_1(t)$ of Y. The equations

$$t = \frac{w}{\sqrt{v/r}} \quad and \quad u = v.$$

define a one-to-one transformation that maps A = {(w, v); -∞< w<∞,0 < v <∞} onto B = {(t, u); -∞< t <∞,0 < u <∞}, Since $w = t\sqrt{u}/\sqrt{r},$ v = u, the absolute value of the Jacobian of the transformation is $|J| = \sqrt{u}/\sqrt{r}$. Accordingly, the joint p.d.f. of T and U =V is given by

$$g(t,u) = \varphi\left(\frac{t\sqrt{u}}{\sqrt{r}}, u\right)|J|$$

$$= \frac{1}{\sqrt{2\pi}\Gamma(\frac{r}{2})2^{r/2}} u^{r/2-1} \exp\left[\frac{-u}{2}\left(1+\frac{t^2}{r}\right)\frac{\sqrt{u}}{\sqrt{r}}\right] - \infty < t < \infty, 0 < u < \infty,$$

$$= 0 \quad elsewhere.$$

The marginal p.d.f. of T is then

$$g_1(t) = \int_{-\infty}^{\infty} g(t,u)du$$

$$= \int_0^{\infty} \frac{1}{\sqrt{2\pi r}\Gamma(\frac{r}{2})2^{r/2}} u^{(r+1)/2-1} \exp\left[\frac{-u}{2}\left(1+\frac{t^2}{r}\right)\right] du$$

In this integral, let $z = u[1 + (\frac{t^2}{2})]/2$, and it is seen that

$$g_1(t) = \int_0^{\infty} \frac{1}{\sqrt{2\pi r}\Gamma\left(\frac{r}{2}\right)2^{\frac{r}{2}}} \left(\frac{2z}{1+\frac{t^2}{r}}\right)^{(r+1)/2-1} e^{-z}\left(\frac{2}{1+\frac{t^2}{r}}\right) dz$$

$$= \frac{\Gamma[\frac{r+1}{2}]}{\sqrt{\pi r}\Gamma(\frac{r}{2})} \frac{1}{\left(1+\frac{t^2}{r}\right)^{(r+1)/2}}, \qquad -\infty < t < \infty.$$

Thus, if W is n (0,1), if V is X^2(r), and if W and V are stochastically independent, then

$$T = \frac{W}{\sqrt{V/r}}.$$

4.2 Extensions of the Change-of-Variable Technique

Consider an integral of the form

$$\int_{\dots A} \int \varphi(x_1, x_2, \dots, x_n) dx_1 dx_2 \dots dx_n$$

taken over a subset A of n- dimensional space A. Let

$$y_1 = u_1(x_1, x_2, \dots, x_n), y_2 = u_2(x_1, x_2, \dots x_n), \dots, y_n = u_n(x_1, x_2, \dots x_n),$$

together with the inverse functions

$$x_1 = w_1(y_1, \dots., y_n), x_2 = w_2(y_1, \dots., y_n), \dots, x_n = w_n(y_1, \dots., y_n)$$

define a one-to-one transformation that maps A onto B in the$y_1, y_2, \dots., y_n$space (and hence maps the subset A of A onto a subset B of B). Let the first partial derivatives of the inverse functions be continuous and let the *n* by *n* determinant (called the Jacobian)

$$J = \begin{bmatrix} \frac{\partial x_1}{\partial y_1} \frac{\partial x_1}{\partial y_2} & \cdots & \frac{\partial x_1}{\partial y_n} \\ \vdots & & \vdots \\ \frac{\partial x_n}{\partial y_1} \frac{\partial x_n}{\partial y_2} & \cdots & \frac{\partial x_n}{\partial y_n} \end{bmatrix}$$

not vanish identically in B. Then

$$\int_A \cdots \int \varphi(x_1, x_2, \dots, x_n) dx_1 dx_2 \dots dx_n$$
$$= \int_B \cdots \int \varphi[w_1(y_1, \dots, y_n), w_2(y_1, \dots, y_n), \dots, w_n(y_1, \dots, y_n)] \times |J| dy_1 dy_2 \dots dy_n$$

The joint p.d.f. of the random variables $Y_1 = u_1(X_1, X_2, \dots, X_n), Y_2 = u_2(X_1, X_2, \dots, X_n), \dots, Y_n = u_n(X_1, X_2, \dots, X_n)$ – where the joint p.d.f. of $X_1, X_2, \dots, X_n$ is $\varphi(x_1, x_2, \dots, x_n)$ – is given by

$$g(y_1, \dots, y_n) = |J| \varphi[w_1(y_1, \dots, y_n), w_2(y_1, \dots, y_n), \dots, w_n(y_1, \dots, y_n)],$$

When $(y_1, \dots, y_n) \in B$, and is zero elsewhere.

Example 1. Let $X_1 X_2$, … , X_{k+1} be mutually stochastically independent random variables, each having a gamma distribution with β= 1. The joint p.d.f. of these variables may be written as

$$\varphi(x_1, x_2, \dots, x_{k+1}) = \prod_{i=1}^{k+1} \frac{1}{\Gamma(\alpha_1)} x_i^{\alpha_i - 1} e^{-x_i}, \quad 0 < x_i < \infty,$$

$$= 0 \qquad elsewhere.$$

Let

$$Y_i = \frac{X_i}{X_1 + X_2 + \cdots + X_{k+1}}, \quad i = 1,2, \dots, k,$$

and $Y_{k+1} = X_1 + X_2 + \cdots + X_{k+1}$ denote k + 1 new random variables. The associated transformation maps A $=\{(x_1, x_2, \dots, x_{k+1});\ 0<x_i<\infty,\ i=1,\dots,k+1\}$ onto the space
B = $\{(y_1, \dots, y_k, y_{k+1});\ 0<y_i,\ i=1,\dots,k,\ y_1 + \cdots + y_k < 1, 0 < y_{k+1} < \infty\}$.

The single-valued inverse functions are $x_1 = y_1 y_{k+1}, \dots, x_k = y_k y_{k+1}$, $x_{k+1} = y_{k+1}(1 - y_1 - \cdots - y_k)$, so that the Jacobian is

$$J = \begin{vmatrix} y_{k+1} & 0 & \dots & 0 & y_1 \\ 0 & y_{k+1} \dots & 0 & & y_2 \\ & & \vdots\vdots & & \\ & 0 & 0 & \dots y_{k+1} y_k & \\ -y_{k+1} & -y_{k+1} \dots & -y_{k+1}(1-y_1-\dots-y_k) & & \end{vmatrix} = y_{k+1}^k.$$

Hence the joint p.d.f. of $Y_1, \dots, Y_k, Y_{k+1}$ is given by

$$\frac{y_{k+1}^{\alpha_1+\dots+\alpha_{k+1}-1} y_1^{\alpha_1-1} \dots y_k^{\alpha_k-1}(1-y_1-\dots-y_k)^{\alpha_{k+1}-1} e^{-y_{k+1}}}{\Gamma(\alpha_1)\dots\Gamma(\alpha_k)\Gamma(\alpha_{k+1})},$$

provided that $(y_1, \dots, y_k, y_{k+1}) \in B$ and is equal to zero elsewhere. The p.d.f. of $Y_1, \dots, Y_k$ is

$$g(y_1, \dots, y_k) = \frac{\Gamma(\alpha_1+\dots+\alpha_{k+1})}{\Gamma(\alpha_1)\dots\Gamma(\alpha_{k+1})} y_1^{\alpha_1-1} \dots y_k^{\alpha_k-1}(1-y_1-\dots-y_k)^{\alpha_{k+1}-1},$$

When $0 < y_i,\ i = 1, \dots, k,\ y_1 + \dots + y_k < 1$, while the function g is equal to zero elsewhere. Random variables $Y_1,\dots,Y_k$ that have a joint p.d.f. of this form are said to have a *Dirichlet distribution* with parameters $a_1, \dots, a_k, a_{k+1}$ and any such $g(y_1, \dots, y_k)$ is called a *Dirichlet p.d.f.* It is seen, in the special case of $k = 1$, that the Dirichlet p.d.f. becomes a beta p.d.f.
Moreover, it is also clear from the joint p.d.f. of $Y_1,\dots,Y_k,Y_{k+1}$ that Y_{k+1} has a gamma distribution with parameters $\alpha_1 + \dots + \alpha_{k+1}$ and $\beta = 1$ and that Y_{k+1} is stochastically independent of $Y_1,\dots,Y_k$.

Now, let X have the Cauchy p.d.f.

$$f(x) = \frac{1}{\pi(1+x^2)}, \quad -\infty < x < \infty,$$

and let $Y = X^2$. We seek the p.d.f. $g(y)$ of Y. Consider the transformation $y = x^2$. This transformation maps the space of X, d = $\{x;\ -\infty< x <\infty\}$, onto B= $\{y;\ 0 < Y <\infty\}$, However, the transformation is not one-to-one. To each $y \in B$, with the exception of y = 0, there correspond two points $x \in A$. For example, if y= 4, we may have either $x = 2$ or $x = -2$. In such an instance, we represent d as the union of two disjoint sets A_1 and A_2 such that $Y = x^2$ defines a one-to-one transformation that maps each of A_1 and A_2 onto B. If we take A_1 to be $\{x;\ -\infty< x <0\}$ and A_2 to be $\{x;\ 0 < x <\infty\}$, we see that A_1 is mapped onto $\{y;\ 0 < Y <\infty\}$, where as A_2 is mapped onto $\{y;\ 0 < Y <\infty\}$, and these sets are not the same.

Take $A_1 = \{x;\ -\infty< x <0\}$ and $A_2 = \{x;\ 0 < x <\infty\}$. Thus $y = x^2$ with the inverse $x = -\sqrt{y}$, maps A_1 onto B = $\{y;\ 0 < Y <\infty\}$ and the transformation is one-to-one. Moreover, the transformation $y = x^2$ with the inverse $x = -\sqrt{y}$, maps A_2 onto B = $\{y;\ 0 < Y <\infty\}$ and the transformation is one-to-one.

Consider the probability Pr (Y ∈ B), where B $\subset B$. Let $A_3 = \{x; x = -\sqrt{y}, y \in B\} \subset A_1$ and let $A_4 = \{x; x = \sqrt{y}, y \in B\} \subset A_2$. Thus we have

$$\Pr(Y \in B) = \Pr(X \in A_3) + \Pr(X \in A_4)$$

$$= \int_{A_3} f(x)dx + \int_{A_4} f(x)dx.$$

In the first of these integrals, let $x = -\sqrt{y}$. Thus the Jacobian, say J_1, is $-1/2\sqrt{y}$; moreover, the set A_3 is mapped onto *B*. In the second integral let $x = \sqrt{y}$. Thus the Jacobian, say J_2 is $1/2\sqrt{y}$ moreover, the set A_4 is also mapped onto *B*. Finally,

$$\Pr(Y \in B) = \int_B f(-\sqrt{y}) \left| -\frac{1}{2\sqrt{y}} \right| dy + \int_B f(\sqrt{y}) \frac{1}{2\sqrt{y}} dy$$

$$= \int_B [f(-\sqrt{y}) + f(\sqrt{y})] \frac{1}{2\sqrt{y}} dy.$$

Hence the p.d.f. of Y is given by

$$g(y) = \frac{1}{2\sqrt{y}} [f(-\sqrt{y}) + f(\sqrt{y})], \quad y \in B.$$

With f (x) the Cauchy p.d.f. we have

$$g(y) = \frac{1}{\pi(1+y)\sqrt{y}}, \quad 0 < y < \infty,$$
$$= 0 \ elsewhere.$$

Let $\varphi(x_1, x_2, \ldots, x_n)$,be the joint p.d.f. of $X_1, X_2, \ldots, X_n$,which are random variables of the continuous type. Let A be the n-dimensional space where $y_1 = \mu_1(x_1, x_2, \ldots, x_n), \ldots, y_n = \mu_n(x_1, x_2, \ldots, x_n)$ which maps A onto B on the $y_1, \ldots, y_n$ space. To each point of A there will correspond, of course, but one point in B; but to a point in B there may correspond more than one point in A. That is, the transformation may not be one-to-one. Suppose, however, that we can represent A as the union of a finite number, say *k,* of mutually disjoint sets $A_1, A_2, \ldots, A_k$ so that

$$y_1 = \mu_1(x_1, x_2, \ldots, x_n), \ldots, \quad y_n = \mu_n(x_1, x_2, \ldots, x_n)$$

Define a one-to-one transformation of each A_i onto B. Thus, to each point in B there will correspond exactly one point in each $A_1, A_2, \ldots, A_k$. Let

$$x_1 = w_{1i}(y_1, \ldots, y_n), x_2 = w_{2i}(y_1, \ldots, y_n), \ldots, x_n = w_{ni}(y_1, \ldots, y_n), i = 1,2, \ldots, k.$$

denote the k groups of n inverse functions, one group for each of these k transformations. Let the first partial derivatives be continuous and let each

$$J_i = \begin{vmatrix} \frac{\partial w_{1i}}{\partial y_1} \frac{\partial w_{1i}}{\partial y_2} \cdots & \frac{\partial w_{1i}}{\partial y_n} \\ \frac{\partial w_{2i}}{\partial y_1} \frac{\partial w_{2i}}{\partial y_2} \cdots & \frac{\partial w_{2i}}{\partial y_n} \\ \vdots \quad \vdots & \vdots \\ \frac{\partial w_{ni}}{\partial y_1} \frac{\partial w_{ni}}{\partial y_2} \cdots & \frac{\partial w_{ni}}{\partial y_n} \end{vmatrix}, \quad i = 1,2, \ldots, k,$$

be not identically equal to zero in B. From a consideration of the probability of the union of k mutually exclusive events and by applying the change of variable technique to the probability of each of these events, it can be seen that the joint p.d.f. of $Y_1 = u_1(X_1, X_2, \ldots, X_n), \ldots, Y_n = u_n(X_1, X_2, \ldots, X_n)$, is given by

$$g(y_1, \ldots, y_n) = \sum_{i=1}^{k} |J_i| \varphi[w_{1i}(y_1, \ldots, y_n), \ldots, w_{ni}(y_1, \ldots, y_n)],$$

provided that $(y_1, \ldots, y_n) \in B$, and equals to zero elsewhere. The p.d.f. of any Y_i, say Y_1, is then

$$g_1(y_1) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(y_1, \ldots, y_n) d\, y_2 \ldots dy_n.$$

Example 2. To illustrate the result just obtained, take*n* = 2and letX_1, X_2denote a random sample of size 2 from a distribution that is*n(0, 1).*The joint p.d.f. of X_1 and X_2 is

$$f(x_1, x_2) = \frac{1}{2\pi} \exp\left(-\frac{x_1^2 + x_2^2}{2}\right), \qquad -\infty < x_1 < \infty, -\infty < x_2 < \infty.$$

Let Y_1 denote the mean and let Y_2 denote twice the variance of the random sample. The associated transformation is

$$y_1 = \frac{x_1 + x_2}{2},$$

$$y_1 = \frac{(x_1 - x_2)^2}{2}.$$

The transformation maps $A = \{(x_1, x_2);\ -\infty < x_1 < \infty, -\infty < x_2 < \infty\}$ onto $B = \{(y_1, y_2); -\infty < y_1 < \infty, 0 \le y_2 < \infty\}$. But the transformation is not one-to-one because, to each point in B, exclusive of points where $y_2 = 0$, there correspond two points in A. In fact, the two groups of inverse functions are

$$x_1 = y_1 - \sqrt{\frac{y_2}{2}} x_2 = y_1 + \sqrt{\frac{y_2}{2}},$$

And

$$x_1 = y_1 + \sqrt{\frac{y_2}{2}} x_2 = y_1 - \sqrt{\frac{y_2}{2}}.$$

Moreover the set A cannot be represented as the union of two disjoint sets, each of which under our transformation maps onto B. Our difficulty is caused by those points of A that lie on the line whose equation is $x_2 = x_1$.At each of these points we have$y_2 = 0$. However, we can define $f(x_1, x_2)$to be zero at each point where $x_1 = x_2$.We can do this without altering the distribution of probability, because the probability measure of this set is zero. Thus we have a new $A = \{(x_1, x_2);\ -\infty < x_1 < \infty, -\infty < x_2 < \infty,\ x_1 \ne x_2\}$. This space is the union of the two disjoint sets $A_1 = \{(x_1, x_2); x_2 > x_1\} and\ A_2 = \{(x_1, x_2); x_2 < x_1\}$. Moreover, our transformation now define a one-to-one transformation of each A_j, j = 1,2, onto the new $B = \{(y_1, y_2); -\infty < y_1 < \infty, 0 \le y_2 < \infty\}$. We can now find the joint p.d.f. say $g(y_1, y_2)$, of the mean Y_1 and twice the variance Y_2 of our random sample.

$$|J_1| = |J_2| = \frac{1}{\sqrt{2y_2}}.$$

Thus

$$g(y_1, y_2)$$
$$= \frac{1}{2\pi} \exp\left[-\frac{\left(y_1 - \sqrt{\frac{y_2}{2}}\right)^2}{2} - \frac{\left(y_1 + \sqrt{\frac{y_2}{2}}\right)^2}{2}\right] \frac{1}{\sqrt{2y_2}}$$
$$+ \frac{1}{2\pi} \exp\left[-\frac{\left(y_1 + \sqrt{\frac{y_2}{2}}\right)^2}{2} - \frac{\left(y_1 - \sqrt{\frac{y_2}{2}}\right)^2}{2}\right] \frac{1}{\sqrt{2y_2}}$$

$$= \sqrt{\frac{2}{2\pi}} e^{-y_1^2} \frac{1}{\sqrt{2}\Gamma\left(\frac{1}{2}\right)} y_2^{\frac{1}{2}-1} e^{-\frac{y_2}{2}}, -\infty < y_1 < \infty, 0 \leq y_2 < \infty.$$

The mean Y_1 of our random sample is n(0,1/2); Y_2, which is twice the variance of our sample is $X^2(1)$; and the two are stochastically independent. Thus the mean and the variance of our sample are stochastically independent.

4.3 The Moment-Generating-Function Technique

Let $\varphi(x_1, x_2, \ldots, x_n)$ denote the joint p.d.f. of the n random variables $X_1, X_2, \ldots, X_n$. These random variables may or may not be the items of a random sample from some distribution that has a given p.d.f. f(x). Let$Y_1 = u_1(X_1, X_2, \ldots, X_n)$. We seek $g(y_1)$, the p.d.f. of the random variable Y_1. Consider the moment-generating function of Y_1. If it exists, it is given by

$$M(t) = E(e^{tY_1}) = \int_{-\infty}^{\infty} e^{ty_1} g(y_1) dy_1$$

in the continuous case.

***Example* 1**. Let the stochastically independent random variables $X_1$and X_2 have the same p.d.f.

$$f(x) = \frac{x}{6}, \quad x = 1,2,3,$$

$$= 0 \; elsewhere.$$

That is the p.d.f. of X_1 is $f(x_1)$ and that of X_2 is $f(x_2)$; and so that the joint p.d.f. of X_1 and X_2 is

$$f(x_1)f(x_2) = \frac{x_1 x_2}{36}, \; x_1 = 1,2,3, x_2 = 1,2,3,$$

$$= 0 \qquad elsewhere.$$

A probability, such as $\Pr(X_1 = 2, X_2 = 3)$, can be seen immediately to be $\frac{(2)(3)}{36} = \frac{1}{6}$. However, consider a probability such as $\Pr(X_1 + X_2 = 3)$, the computation can be made by first observing that the event $X_1 + X_2 = 3$ is the union exclusive of the events with probability zero of the non mutually exclusive events $(X_1 = 1, X_2 = 2)$ and $(X_1 = 2, X_2 = 1)$. Thus

$$\Pr(X_1 + X_2 = 3) = Pr(X_1 = 1, X_2 = 2) + Pr(X_1 = 2, X_2 = 1)$$

$$= \frac{(1)(2)}{36} + \frac{(2)(1)}{36} = \frac{4}{36}.$$

More generally, let y represent any of the numbers 2,3,4,5,6. The probability of each of the events $X_1 + X_2 = y$, y= 2,3,4,5,6, can be computed as in the case y = 3. Let $g(y) = \Pr(X_1 + X_2 = y)$. Then the table

y	2	3	4	5	6
$g(y)$	$\frac{1}{36}$	$\frac{4}{36}$	$\frac{10}{36}$	$\frac{12}{36}$	$\frac{9}{36}$

Gives the values of g(y) for y = 2,3,4,5,6. For all the values of u, g(y) = 0. Now, define a new random variable Y of $X^2(r_1)$,, and then we have to calculate the p.d.f. g(y) of this random variable Y. We shall now solve the same problem and by the moment generating function technique.

Now the moment generating function of Y is

$$M(t) = E(e^{t(X_1+X_2)})$$

$$= E(e^{tX_1}e^{tX_2})$$

$$= E(e^{tX_1})E(e^{tX_2}),$$

Since X_1 and X_2 are stochastically independent.

Theorem 1. Let$X_1, X_2, \ldots, X_n$ be mutually stochastically independentrandom variables having, respectively, the normal distributions $n(\mu_1, \sigma_1^2), n(\mu_2, \sigma_2^2), \ldots and\ n(\mu_n, \sigma_n^2)$. The random variable $Y = k_1X_1 + k_2X_2 + \cdots + k_nX_n$, where $k_1, k_2, \ldots, k_n$ are real constants, is normally distributed with mean $k_1\mu_1 + \cdots + k_n\mu_n$ and variance ${k_1}^2\sigma_1^2 + \cdots + {k_n}^2\sigma_n^2$.

Proof.

Since because $X_1, X_2, \ldots, X_n$ are mutually stochastically independent the moment generating function of Y is given by

$$M(t) = E\{\exp[t(k_1X_1 + k_2X_2 + \cdots + k_nX_n)]\}$$

$$= E(e^{tk_1X_1})E(e^{tk_2X_2}) \ldots E(e^{tk_nX_n}).$$

Now

$$E(e^{tX_1}) = \exp\left(\mu_1 t + \frac{\sigma_1^2 t^2}{2}\right),$$

for all real t, i = 1,2,…,n. Hence we have

$$E(e^{tk_1X_1}) = \exp\left[\mu_i(k_i t) + \frac{\sigma_i^2 (k_i t)^2}{2}\right].$$

That is, the moment generating function of Y is

$$M(t) = \prod_{i=1}^{n} \exp\left[\mu_i(k_i t) + \frac{\sigma_i^2 (k_i t)^2}{2}\right]$$

$$= \exp\left[\left(\sum_1^n k_i\mu_i\right)t + \frac{(\sum_1^n k_i^2\sigma_i^2)t^2}{2}\right].$$

But this is the moment generating function of a distribution that is $n(\sum_1^n k_i\mu_i, \sum_1^n k_i^2\sigma_i^2)$. Hence the proof.

Theorem 2. Let $X_1, X_2, \ldots, X_n$ be mutually stochastically independent variables that have, respectively, the chi-square distributions $X^2(r_1), X^2(r_2), \ldots, and\ X^2(r_n)$. Then the random variable $Y = X_1 + X_2 + \cdots + X_n$ has a chi-square distribution with $r_1 + \ldots + r_n$ degrees of freedom; that is, Y is $X^2(r_1 + \ldots + r_n)$.

Proof.

The moment generating function of Y is

$$M(t) = E\{\exp[t(X_1 + X_2 + \cdots + X_n)]\}$$

$$= E(e^{tX_1})E(e^{tX_2}) \ldots E(e^{tX_n})$$

Because $X_1, X_2, \ldots, X_n$ are mutually stochastically independent since

$$E(e^{tX_1}) = (1-2t)^{-r_i/2}, t < \frac{1}{2}, i = 1,2,\ldots,n,$$

we have

$$M(t) = (1-2t)^{-(r_1+r_2+\cdots+r_n)/2}, \quad t < \frac{1}{2}.$$

But this is the moment generating function of a distribution that $X^2(r_1 + ... + r_n)$. Accordingly, Y has this chi-square distribution.

Also, let $X_1, X_2, \dots, X_n$ be a random sample of size n from a distribution that is $n(\mu, \sigma^2)$. Thus, each of the random variable $\frac{(X_i-\mu)^2}{\sigma^2}$, $i = 1,2,\dots,n$, is $X^2(1)$. Moreover, these random variables are mutually stochastically independent. By the Theorem 2, the random variable $Y = \sum_1^n \left[\frac{(X_i-\mu)^2}{\sigma^2}\right]$, $i = 1,2,\dots,n$ is $X^2(n)$.

4.4 The Distributions of $\bar{X}$ and nS^2/σ^2

Let $X_1, X_2, \dots, X_n$ denote a random sample of size $n \geq 2$ from a distribution that is $n(\mu, \sigma^2)$. Here we discuss about mean and the variance of this random sample that is the distribution of the two statistics

$$\bar{X} = \sum_1^n \frac{X_i}{n} \, and \;\; S^2 = \sum_1^n \frac{(X_i - \bar{X})^2}{n}.$$

The problem of the distribution of $\bar{X}$, the mean of the sample is solved by the use of Theorem 1 of Section 3.6. We have here, in the notation of the statement of that theorem $\mu_1 = \mu_2 = \cdots \mu_n = \mu, \sigma_1^2 = \sigma_2^2 = \cdots \sigma_n^2 = \sigma^2 \; and \; k_1 = k_2 = \cdots k_n = \frac{1}{n}$. Accordingly Y =X has a normal distribution with mean and variance given by

$$\sum_1^n \left(\frac{1}{n}\mu\right) = \mu, \quad \sum_1^n [\left(\frac{1}{n}\right)^2 \sigma^2 = \frac{\sigma^2}{n},$$

respectively that is X is $n(\mu, \sigma^2/n)$.

***Example* 1**. Let $\bar{X}$ be the mean of a random sample of size 25 from a distribution that is n(75, 100). Thus $\bar{X}$ is n(75, 4). Then, for instance,

$$\Pr(1 < \bar{X} < 79) = N\left(\frac{79-75}{2}\right) - N\left(\frac{71-75}{2}\right)$$

$$= N(2) - N(-2) = 0.954.$$

We now take up the problem of the distribution S^2 the variance of the random sample $X_1, X_2, \dots, X_n$ from a distribution that is $n(\mu, \sigma^2)$. Consider the joint distribution $Y_1 = X_1, Y_2 = X_2, \dots, Y_n = X_n$. The corresponding transformation

$$x_1 = ny_1 - y_2 - \cdots - y_n$$

$$x_2 = y_2$$

$$\vdots \quad \vdots$$

$$x_n = y_n$$

has Jacobian n. Since

$$\sum_1^n (x_i - \mu)^2 = \sum_1^n (x_i - \bar{x} + \bar{x} - \mu)^2$$

$$= \sum_1^n (x_i - \bar{x})^2 + n(\bar{x} - \mu)^2$$

because $2(\bar{x} - \mu)\sum_1^n (x_i - \bar{x}) = 0$, the joint p.d.f. of $X_1, X_2, \ldots, X_n$ can be written

$$\left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \exp\left[-\frac{\sum (x_i - \bar{x})^2}{2\sigma^2} - \frac{n(\bar{x} - \mu)^2}{2\sigma^2}\right],$$

where $\bar{x}$ represent $(x_1 + x_2 + \cdots + x_n)/n \ and - \infty < x_i < \infty, i = 1,2, \ldots, n.$ Accordingly, with $y_1 = \bar{x}$, we find that the joint p.d.f. of $Y_1, Y_2, \ldots, Y_n$ is

$$n\left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \exp\left[-\frac{(ny_1 - y_2 - \cdots - y_n - y_1)^2}{2\sigma^2} - \frac{\sum_2^n (y_i - y_1)^2}{2\sigma^2} - \frac{n(y_1 - \mu)^2}{2\sigma^2}\right],$$

$-\infty < y_i < \infty$, i = 1,2,…,n. The quotient of this joint p.d.f. and the p.d.f.

$$\frac{\sqrt{n}}{\sqrt{2\pi}\sigma} \exp\left[-\frac{n(y_1 - \mu)^2}{2\sigma^2}\right]$$

of $Y_1 = \bar{X}$is the conditional p.d.f.$Y_2, \ldots, Y_n$, given $Y_1 = y_1$,

$$\sqrt{n}\left(\frac{1}{\sqrt{2\pi}\sigma}\right)^{n-1} \exp\left(-\frac{q}{2\sigma^2}\right),$$

where $q = (ny_1 - y_2 - \cdots - y_n - y_1)^2 + \sum_2^n (y_i - y_1)^2$. Since this is a joint p.d.f. it must be, for all $\sigma > 0$, that

$$\int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} \sqrt{n}\left(\frac{1}{\sqrt{2\pi}\sigma}\right)^{n-1} \exp\left(-\frac{q}{2\sigma^2}\right) dy_2 \ldots dy_n = 1.$$

Now consider

$$nS^2 = \sum_1^n (X_i - \bar{X})^2$$

$$= (ny_1 - y_2 - \cdots - y_n - y_1)^2 + \sum_2^n (Y_i - Y_1)^2 = Q.$$

The conditional moment generating function of $\frac{nS^2}{\sigma^2} = \frac{Q}{\sigma^2}$, given $Y_1 = y_1$, is

$$E\left(e^{tQ/\sigma^2} \middle| y_1\right) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \sqrt{n} \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^{n-1} \exp\left(-\frac{(1-2t)q}{2\sigma^2}\right) dy_2 \dots dy_n$$

$$= \left(\frac{1}{1-2t}\right)^{(n-1)/2} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \sqrt{n} \left[\frac{1-2t}{2\pi\sigma^2}\right]^{(n-1)/2}$$
$$\times \exp\left[\left(-\frac{(1-2t)q}{2\sigma^2}\right) dy_2 \dots dy_n\right]$$

where $0 < 1 - 2t, or\ t < \frac{1}{2}$. However, this integral is exactly the same that of the conditional p.d.f. of $Y_2, \dots, Y_n$, given $Y_1 = y_1$ with σ^2 replaced by $\frac{\sigma^2}{1-2t} > 0$, and thus must equal 1. Hence the conditional moment generating functions of $\frac{nS^2}{\sigma^2}$, given $Y_1 = y_1$or equivalently $\bar{X} = \bar{x}$, is

$$E\left(e^{tnS^2/\sigma^2} \middle| \bar{x}\right) = (1-2t)^{-(n-1)/2}, \quad t < \frac{1}{2}.$$

That is, the conditional distribution of $\frac{nS^2}{\sigma^2}$, given by $\bar{X} = \bar{x}$, is X^2(n-1).Moreover, since it is clear that this conditional distribution does not depend, upon $\bar{X}, \bar{x}$ and $\frac{nS^2}{\sigma^2}$ must be stochastically independent or equivalently $\bar{X}\ and\ S^2$ are stochastically independent.

To summarize we have established. In this section, three important properties $\bar{X}\ and\ S^2$ when the sample arises from a distribution which is $n(\mu, \sigma^2)$:

a) $\bar{X}\ is\ n\left(\mu, \frac{\sigma^2}{n}\right)$.
b) $\frac{nS^2}{\sigma^2}$is X^2(n – 1).

c) $\bar{X}$ *and* S^2 are stochastically independent.

4.5 Expectation Of Functions Of Random Variables

Theorem 1. Let $X_1, X_2, \dots, X_n$ denote random variables that have means $\mu_1, \mu_2, \dots, \mu_n$ and variances $\sigma_1^2, \sigma_2^2, \dots, \sigma_n^2$. Let $\rho_{ij},\ i \neq j$, denote the correlation coefficient of X_i and X_j and let $k_1, k_2, \dots, k_n$ denote real constants. The mean and the variance of the linear function

$$Y = \sum_{1}^{n} k_i X_i,$$

are respectively,

$$\mu_Y = \sum_{1}^{n} k_i \mu_i$$

and

$$\sigma_Y^2 = \sum_{1}^{n} k_i^2 \sigma_i^2 + 2 \sum_{i<} \sum_{j} k_i k_j \rho_{ij} \sigma_i \sigma_j.$$

Corollary: Let $X_1, X_2, \dots, X_n$denote the items of a random sample of the variance of $Y = \sum_1^n k_i X_i$ are respectively $\mu_Y = (\sum_1^n k_i)\mu$ and $\sigma_Y^2 = (\sum_1^n k_i^2)\sigma^2$.

Example : Let $\bar{X} = \sum_1^n \frac{X_i}{n}$ denote the mean of a random sample of size n from a distribution that has mean μ and the variance σ_2. In accordance, with the corollary, we have $\mu_{\bar{x}} = \mu \sum_1^n \left(\frac{1}{n}\right) = \mu$ and $\sigma_{\bar{x}}^2 = \sigma^2 \sum_1^n \left(\frac{1}{n}\right)^2 = \frac{\sigma^2}{n}$. We have seen, in this section, that if our sample is from a distribution $n(\mu, \sigma^2)$, then $\bar{X}$ is $n(\mu, \sigma^2/n)$. It is interesting that $\mu_{\bar{x}} = \mu$ and $\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n}$ whether the sample is or is not from a normal distribution.

Unit V

5.1 Limiting Distribution

If $\bar{X}$ is the mean of a random sample $X_1, X_2, \dots, X_n$ from a distribution that has the p.d.f.

$$f(x) = 1, \quad 0 < x < 1,$$

$$= 0 \ elsewhere.$$

The moment generating function of $\bar{X}$ is given by $[M(t/n)]^n$, where here

$$M(t) = \int_0^1 e^{tx}\, dx = \frac{e^t - 1}{t}, \quad t \neq 0,$$

$$= 1, \qquad t = 0.$$

Hence

$$E\left(e^{t\bar{X}}\right) = \left(\frac{e^{\frac{t}{n}} - 1}{\frac{t}{n}}\right)^n, \quad t \neq 0,$$

$$= 1, \qquad t = 0.$$

For example, we shall write

$$F_n(\bar{x}) = \int_{-\infty}^{\bar{x}} \frac{1}{\sqrt{\frac{1}{n}}\sqrt{2\pi}} e^{-nw^2/2} dw$$

for the distribution function of the mean $\bar{X}_n$ of the random sample of the size n from a normal distribution with mean 0 and variance 1.

Definition: Let the distribution function $F_n(y)$ of the random variable Y_n depend upon n, a positive integer. If F(y) is a distribution and if $\lim_{n\to\infty} F(y) = F(y)$ for every point y at which F(y) is continuous, then the random variable Y_n is said to have a limiting distribution with distribution function F(y).

***Example* 1.** Let Y_n denote the n^{th} order statistic of a random sample $X_1, X_2, \dots, X_n$ from a distribution having p.d.f.

$$f(x) = \frac{1}{\theta}, \quad 0 < \theta < \infty,$$

$$= 0 \; elsewhere.$$

The p.d.f. of Y_n is

$$g_n(y) = \frac{ny^{n-1}}{\theta^n}, \; 0 < y < \theta,$$

$$= 0 \; elsewhere.$$

and the distribution function of Y_n is

$$F_n(y) = 0, \quad y < 0,$$

$$= \int_0^y \frac{nz^{n-1}}{\theta^n} dz = \left(\frac{y}{\theta}\right)^n, \; 0 \le y < \theta,$$

$$= 1, \quad \theta \le y < \infty.$$

Then

$$\lim_{n \to \infty} F_n(y) = 0, \quad -\infty < y < \theta,$$

$$= 1, \qquad \theta \le y < \infty.$$

Now

$$F(y) = 0, \quad -\infty < y < \theta,$$

$$= 1, \qquad \theta \le y < \infty.$$

is a distribution function.

***Example* 2.** Let X_n have the distribution function

$$F_n(\bar{x}) = \int_{-\infty}^{\bar{x}} \frac{1}{\sqrt{\frac{1}{n}}\sqrt{2\pi}} e^{-nw^2/2} dw.$$

If the change of the variable $v = \sqrt{n}w$ is made, we have

$$F_n(\bar{x}) = \int_{-\infty}^{\sqrt{n}\bar{x}} \frac{1}{\sqrt{2\pi}} e^{-v^2/2} dv.$$

It is clear that

$$\lim_{n\to\infty} F_n(\bar{x}) = 0, \quad \bar{x} < 0,$$

$$= \frac{1}{2}, \quad \bar{x} = 0,$$

$$= 1, \quad \bar{x} > 0.$$

Now the function

$$F(\bar{x}) = 0, \quad \bar{x} < 0,$$

$$= 1, \quad \bar{x} \geq 0,$$

is a distribution function and $\lim_{n\to\infty} F_n(\bar{x}) = F(\bar{x})$ at every point of continuity of F($\bar{x}$).

Accordingly the random variable X_n has a limiting distribution with distribution function F($\bar{x}$). Again this limiting distribution is degenerate and has all the probability at one point $\bar{x} = 0$.

***Example* 3.** The fact that limiting distributions, if they exist cannot general be determined by taking the limit of p.d.f. will now illustrated let X_n have the p.d.f.

$$f_n(x) = 1, \quad x = 2 + \frac{1}{n},$$

$$= 0 \quad elsewhere.$$

Clearly $\lim_{n\to\infty} f_n(x) = 0$ all values of x. This may suggest that X_n is

$$F_n(x) = 0, \quad x < 2 + \frac{1}{n},$$

$$= 1, \quad x \geq 2 + \frac{1}{n}.$$

and

$$\lim_{n\to\infty} F_n(x) = 0, \quad x \leq 2,$$

$$= 1, \quad x > 2.$$

Since

$$F(x) = 0, \quad x < 2,$$

$$= 1, \quad x \geq 2,$$

Is a distribution function, and since $\lim_{n\to\infty} F_n(x) = F(x)$ at all points of continuity of F(x), there is a limiting distribution of X_n with distribution function F(x).

5.2 Stochastic Convergence

Theorem : Let $F_n(y)$ denote the distribution function of a random variable Y_n whose distribution depends upon the positive integer n. Let c denote a constant which does not depend on n. The random variable Y_n converges stochastically to the constant c if and only if, for ever $\in > 0$, the

$$\lim_{n\to\infty} \Pr\left(|Y_n - c| < \epsilon\right) = 1.$$

Proof.

Let

$$\lim_{n\to\infty} \Pr\left(|Y_n - c| < \epsilon\right) = 1.$$

for every $\in > 0$. We have to prove that the random variable Y_n converges stochastically to the constant c. That is we have to prove that

$$\lim_{n\to\infty} F_n(y) = 0, \quad y < c,$$

$$= 1, \quad y > c.$$

If the limit of $F_n(y)$ is indicated, then Y_n has a limiting distribution with distribution function

$$F(y) = 0, \quad y < c,$$

$$= 1, \quad y \geq c.$$

Now

$$\Pr(|Y_n - c| < \epsilon) = F_n[(c+\epsilon)-] - F_n(c-\epsilon),$$

where $F_n[(c+\epsilon)-]$ is the left-hand limit of $F_n(y)$ at $y = c+\epsilon$. Thus we have

$$1 = \lim_{n\to\infty} \Pr(|Y_n - c| < \epsilon) = \lim_{n\to\infty} F_n[(c+\epsilon)-] - \lim_{n\to\infty} F_n(c-\epsilon).$$

Because $0 \leq F_n(y) \leq 1$ for all values of y and for every positive integer n, it must be that

$$\lim_{n\to\infty} F_n(c-\epsilon) = 0, \quad \lim_{n\to\infty} F_n[(c+\epsilon)-] = 1.$$

Since this is true for every $\in > 0$, we have

$$\lim_{n\to\infty} F_n(y) = 0, \quad y < c,$$

$$= 1, \quad y > c.$$

Now, we assume that

$$\lim_{n\to\infty} F_n(y) = 0, \quad y < c,$$

$$= 1, \quad y > c.$$

We are to prove that $\lim_{n\to\infty} \Pr(|Y_n - c| < \epsilon) = 1$, for every $\in > 0$. Because

$$\lim_{n\to\infty} \Pr(|Y_n - c| < \epsilon) = \lim_{n\to\infty} F_n[(c+\epsilon)-] - \lim_{n\to\infty} F_n(c-\epsilon),$$

and because it is given that

$$\lim_{n\to\infty} F_n(c-\epsilon) = 0, \quad \lim_{n\to\infty} F_n[(c+\epsilon)-] = 1.$$

for every $\in > 0$, we have the desired result. This completes the proof of the theorem.

That is this last limit is also a necessary and sufficient condition for the stochastic convergence of the random variable Y_n to the constant c.

Example: Let $\bar{X}_n$ denote the meanof a random sample of size n from a distribution that has a mean μ and positive variance σ^2.Then the mean and variance of $\bar{X}_n$ are $\mu \; and \frac{\sigma^2}{\mathrm{n}}$. Consider for every fixed $\in > 0$, the probability

$$\Pr(|\bar{X}_n - \mu| < \epsilon) = \Pr\left(|\bar{X}_n - \mu| \geq \frac{k\sigma}{\sqrt{n}}\right),$$

Where $k = \epsilon\sqrt{n}/\sigma$. In accordance with the inequality of Chebyshev, this probability is less than or equal to $\frac{1}{k^2} = \sigma^2/n\epsilon^2$. So, for every fixed $\in > 0$, we have

$$\lim_{n\to\infty} \Pr\left(|\bar{X}_n - \mu| \geq \epsilon\right) =\leq \lim_{n\to\infty} \frac{\sigma^2}{n\epsilon^2} = 0.$$

Hence X_n converges stochastically to μ if σ^2 is finite.

5.3 Limiting Moment-Generating Functions

Result:

Let the random variable $Y_n$have the distribution function $F_n(y)$ and the moment generating function M(t;n) that exists for -h < t < h for all n. If there exists a distribution F(y), with corresponding moment-generating function M(t), defined for $|t| \leq h_1 < h$, such that $\lim_{n\to\infty} M(t;n) = M(t)$, then Y_n has a limiting distribution with distribution function F(y).

***Example* 1.** Let Y_n have a distribution that b(n,p). Suppose that the mean $\mu = np$ is the same for every n; that is $p = \frac{\mu}{n}$, where μ is a constant. We shall find the limiting distribution of the binomial distribution, when $p = \frac{\mu}{n}$, by finding the limit of M(t;n). Now

$$M(t;n) = E(e^{tY_n}) = [(1-p) + pe^t]^n = \left[1 + \frac{\mu(e^t - 1)}{n}\right]^n$$

for all real values of t. Henc we have

$$\lim_{n\to\infty} M(t;n) = e^{\mu(e^t - 1)}$$

for all real values of t. Since there exists a distribution, namely the poisson distribution with mean μ that has this moment generating function $e^{\mu(e^t-1)}$, then in accordance with the theorem and under the conditions stated, it is seen that Y_n has a limiting poisson distribution with mean μ.

***Example* 2.** Let Z_n be $X^2(n)$. Then the moment generating function of Z_n is $(1 - 2t)^{-n/2}, t < 1/2$. The mean and the variance of Z_n are respectively n and 2n. The limiting distribution of the random variable $Y_n = (Z_n - n)/\sqrt{2n}$will be investigated. Now the moment generating function of Y_n is

$$M(t;n) = E\left\{\exp\left[t\left(\frac{Z_n - n}{\sqrt{2n}}\right)\right]\right\}$$

$$= e^{-\frac{tn}{\sqrt{2n}}} E\left(e^{\frac{tZ_n}{\sqrt{2n}}}\right)$$

$$= \exp\left[-\left(t\sqrt{\frac{2}{n}}\right)\left(\frac{n}{2}\right)\right]\left(1 - 2\frac{t}{\sqrt{2n}}\right)^{-n/2},$$

$$t < \frac{\sqrt{2n}}{2}.$$

This may be written in the form

$$M(t;n) = \left(e^{t\sqrt{2/n}} - t\sqrt{\frac{2}{n}}e^{t\sqrt{2/n}}\right)^{-n/2}, \quad t < \sqrt{\frac{n}{2}}.$$

In accordance with Taylor's formula, there exists a number $\varepsilon(n)$, between 0 and $t\sqrt{\frac{2}{n}}$, such that

$$e^{t\sqrt{2/n}} = 1 + t\sqrt{\frac{2}{n}} + \frac{1}{2}\left(t\sqrt{\frac{2}{n}}\right)^2 + \frac{e^{\varepsilon(n)}}{6}\left(t\sqrt{\frac{2}{n}}\right)^3.$$

If this sum is substituted for $e^{t\sqrt{2/n}}$ in the last expression for M(t;n), it is seen that

$$M(t;n) = \left(1 - \frac{t^2}{n} + \frac{\psi(n)}{n}\right)^{-n/2},$$

where

$$\psi(n) = \frac{\sqrt{2}t^3 e^{\varepsilon(n)}}{3\sqrt{n}} - \frac{\sqrt{2}t^3}{\sqrt{n}} - \frac{2t^4 e^{\varepsilon(n)}}{3n}.$$

Since $e^{\varepsilon(n)} \to 0 \; as \; n \to \infty$, then $\lim \psi(n) = 0$ for every fix value of t. Also

$$\lim_{n\to\infty} M(t;n) = e^{t^2/2}$$

for all real values of t. That is the random variable $Y_n = (Z_n - n)/\sqrt{2n}$ has a limiting normal distribution with mean zero and variance 1.

5.4 The Central Limit Theorem

Statement: Let $X_1, X_2, \dots, X_n$ denote the items of random sample from a distribution that mean μ and positive variance σ^2. Then the random variable $Y_n = \frac{\sum_1^n (X_i - n\mu)}{\sqrt{n}\sigma} = \sqrt{n}(\bar{X}_n - \mu)/\sigma$ has the limiting distribution that is normal with mean zero and variance 1.

Proof.

We assume the existence of the moment generating function $M(t) = E(e^{tx}), -h < t < h,$ of the distribution.

The function

$$m(t) = E\left[e^{t(X-\mu)}\right] = e^{-\mu t}M(t)$$

also exists for -h <t <h. Since, m(y) is the moment generating function X - μ, it must follow that $m(0) = 1, m'(0) = E(X - \mu) = 0 \; and \; m''(0) = E[(X - \mu)^2] = \sigma^2$.

By Taylor's formula, there exist a number ε between 0 and t such that

$$m(t) = m(0) + m'(0)t + \frac{m''(\varepsilon)t^2}{2}$$

$$= 1 + \frac{m''(\varepsilon)t^2}{2}.$$

If $\frac{\sigma^2 t^2}{2}$ is added and subtracted then

$$m(t) = 1 + \frac{\sigma^2 t^2}{2} + \frac{[m''(\varepsilon) - \sigma^2]t^2}{2}.$$

Next consider M(t;n), wherr

$$M(t;n) = E\left[\exp\left(\frac{\sum_1^n X_i - n\mu}{\sqrt{n}\sigma}\right)\right]$$

$$= E\left[\exp\left(t\frac{X_1 - \mu}{\sqrt{n}\sigma}\right)\exp\left(t\frac{X_2 - \mu}{\sqrt{n}\sigma}\right) \dots \exp\left(t\frac{X_n - \mu}{\sqrt{n}\sigma}\right)\right]$$

$$= E\left[\exp\left(t\frac{X_1 - \mu}{\sqrt{n}\sigma}\right)\right] \dots E\left[\exp\left(t\frac{X_n - \mu}{\sqrt{n}\sigma}\right)\right]$$

$$= \left\{ E \exp\left(t \frac{X-\mu}{\sqrt{n}\sigma} \right) \right\}^n$$

$$= \left[m\left(\frac{t}{\sqrt{n}\sigma} \right) \right]^n, \quad -h < \frac{t}{\sqrt{n}\sigma} < h.$$

In m(t) replace t by $\frac{t}{\sqrt{n}\sigma}$ to obtain

$$m\left(\frac{t}{\sqrt{n}\sigma} \right) = 1 + \frac{t^2}{2n} + \frac{[m''(\varepsilon) - \sigma^2]t^2}{2n\sigma^2},$$

where now ε is between 0 and $\frac{t}{\sqrt{n}\sigma}$ with $-h\sqrt{n}\sigma < t < h\sqrt{n}\sigma$. Accordingly

$$M(t;n) = \left\{ 1 + \frac{t^2}{2n} + \frac{[m''(\varepsilon) - \sigma^2]t^2}{2n\sigma^2} \right\}^n.$$

Since $m''(t)$ is continuous at t = 0 and since $\varepsilon \to 0 \ as \ n \to \infty$, we have

$$\lim_{n\to\infty} [m''(\varepsilon) - \sigma^2] = 0.$$

Thus,

$$\lim_{n\to\infty} M(t;n) = e^{t^2/2}$$

for all real values of t. This proves that the random variable $Y_n = \sqrt{n}(\bar{X}_n - \mu)/\sigma$ has a limiting normal distribution with mean zero and variance 1.

1.5. Some Theorems on limiting distributions

Result:

Let $F_n(u)$ denote the distribution function of a random variable U_n whose distribution depends upon the positive integer n. Let U_n converge stochastically to the constant $c \neq 0$. The random variable U_n/c converges stochastically to 1.

Theorem : Let $F_n(u)$ denote the distribution function of a random variable U_n whose distribution depends upon the positive integer n. Further, let U_n converge stochastically to the positive constant and let $Pr(U_n<0) = 0$ for every n. The random variable $\sqrt{U_n}$ converges stochastically to $\sqrt{c}$.

Proof.

We are given that the $\lim_{n\to\infty} \Pr\left|\sqrt{U_n} - \sqrt{c}\right| \geq \varepsilon = 0$ for every $\varepsilon > 0$. We have to prove that the

$\lim_{n\to\infty} \Pr\left|\sqrt{U_n} - \sqrt{c}\right| \geq \varepsilon' = 0$, for every $\varepsilon' > 0$. Now the probability

$$\lim_{n\to\infty} \Pr\left|\sqrt{U_n} - \sqrt{c}\right| \geq \varepsilon = \Pr[\left|\left(\sqrt{U_n} - \sqrt{c}\right)\left(\sqrt{U_n} + \sqrt{c}\right) \geq \varepsilon\right|]$$

$$= \Pr\left(\left|\sqrt{U_n} - \sqrt{c}\right| \geq \frac{\varepsilon}{\left(\sqrt{U_n} + \sqrt{c}\right)}\right)$$

$$\geq \Pr\left(\left|\sqrt{U_n} - \sqrt{c}\right| \geq \frac{\varepsilon}{\sqrt{c}} \geq 0.\right.$$

If we let $\varepsilon' = \frac{\varepsilon}{\sqrt{c}}$, and if we take the limit as n becomes infinite, we have

$$0 = \lim_{n\to\infty} \Pr(|U_n - c| \geq \varepsilon) \geq \lim_{n\to\infty} \Pr\left|\sqrt{U_n} - \sqrt{c}\right| \geq \varepsilon' = 0$$

for every $\varepsilon' > 0$. This completes the proof.

Exercise

1. Show that

$$S^2 = \frac{1}{n}\sum_{1}^{n}(X_i - X)^2 = \frac{1}{n}\sum_{1}^{n} X_i^2 - X^2,$$

   Where $X = \sum_1^n \frac{X_i}{n}$.

2. Find the probability that exactly four items of a random sample of size 5 from the distribution having p.d.f. f(x) = (x+1)/2, -1<x<1, zero elsewhere exceed zero.
3. Let X_1,X_2 be a random sample from the distribution having p.d.f. f(x) = 2x, 0<x<1, zero elsewhere. Find $\Pr\left(\frac{X_1}{X_2} \leq \frac{1}{2}\right)$.
4. If the sample size is n = 2, find the constant c so that $s^2 = c(X_1 - X_2)^2$.
5. Let X have a p.d.f. f(x) = (1/3)x, x = 1,2,3, zero elsewhere. Find the p.d.f. of Y = 2X + 1.
6. Let X have a p.d.f. f(x) = (1/3)x, x = 1,2,3, zero elsewhere. Find the p.d.f. of Y = X_3.
7. If the p.d.f. of X is $f(x) = 2xe^{-x^2}, 0 < x < \infty$, zero elsewhere determine the p.d.f. of $Y = X^2$.
8. Let the stochastically independent random variables X_1 and X_2 have the p.d.f.$f(x) = \frac{1}{6}, x = 1,2,3,4,5,6, \mathit{zero\ elsewhere}$. Find the p.d.f. of Y = $X_1 + X_2$. Note under the appropriate assumptions 11 that Y may be interpreted as the sum of the spots that appear when two dice are cast.

9. Let X_n denote the mean of a random sample if size n from distribution that is $n(\mu, \sigma^2)$. Find the limiting the distribution of X_n.
10. Let X be $X_2(50)$. Appropriate Pr $(40 < X < 60)$.

Study Learning Material Prepared by :

Mrs. K. Jenita Devanesam, M.Sc., M.Phil.,

Assistant Professor of Mathematics,

Pope's College, Sawyerpuram