College of Natural Sciences
Department of Statistics

Course title*:* Introduction Probability
Course code: Stat1012
Credit hours: 3

Instructor's Name: Reta Habtamu Bacha
Email: *habtamuretab@yahoo.com*

CHAPTER 1

INTRODUCTION

Introduction

In our daily life, it is not uncommon to hear words which express our doubts or being uncertain about the happenings of certain events. To mention some instances, “If by chance you meet her, please convey my heart-felt greeting”, “Probably, he might not take the class today", etc.- these statements show uncertainty about the happening of the event under question. In Statistics, however, sensible numerical statements can be made about uncertainty and apply different approaches to calculate probabilities.

In this chapter, there are two main points to be discussed: Possibilities and Probabilities. The first part is about techniques of counting or the methods used to determine the number of possibilities, which are indispensable to calculate probabilities. The second part is meant to introduce some basic terms in probability, followed by different methods of measuring probabilities.

Contents

Learning Outcomes

At the end of this chapter students will be able to:

- ✓ Define probability and non-probability models.
- ✓ Define Basic terms such as random experiment, sample space and events.
- ✓ Relate some common points in probability and set theories.
- ✓ Define probability.
- ✓ Solve Problems related to probability of a certain events.

RESOURCES:

- 🕮 Mathematical Statistics, John E. Freund, 6th Edition, pages 1-72.
- 🕮 Modern Elementary Statistics, Freund and Simon, 9th Edition, pages 105-192.
- 🕮 Statistical Methods, S.P. Gupta, 12th edition, pages A-1.3 -1.56.

A deterministic model is one in which every set of variable states is uniquely determined by parameters in the model and by sets of previous states of these variables. Hypothesize exact relationships and it will be suitable when prediction error is negligible.

In a non-deterministic (stochastic/probabilistic) model, randomness is present, and variable states are not described by unique values, but rather by probability distributions. Hence, there will be a defined pattern or regularity appears to construct a precise mathematical model. Hypothesize two components, which is deterministic and random error.

Example 1.1:

a. Energy contained in a body moving in a vacuum with a speed of light $E = mc^2$
b. If the price of an item increases, then the demand for that item will decrease.
c. Body mass index (BMI) is measure of body fat $BMI = \frac{\text{Weight in Kilograms}}{(\text{Height in Meters})^2}$
d. Systolic blood pressure of newborns is 6 Times the Age in days + Random Error

$$SBP = 6\ age(d) + \varepsilon$$

1.2 Random Experiments, Sample Space and Events

Random experiments

An *experiment* is the process by which an observation (measurement) is obtained. Results of experiments may not be the same even through conditions which are identical. Such experiments are called *random experiments*.

Example 1.2:

a. If we aretossing a fair die the result of the experiment is that it will come up with one of the following numbers in the set S = {1, 2, 3, 4, 5, 6}
b. If an experiment consists of measuring “lifetimes” of electric light bulbs produced by a company, then the result of the experiment is a time t in hours that lies in some interval say, $0 \leq t \leq 4000$ where we assume that no bulb lasts more than 4000 hours.

Example 1.3: In an experiment of rolling a fair die, S = {1, 2, 3, 4, 5, 6}, each sample point is an equally likely outcome. It is possible to define many events on this sample space as follows:

A = {1, 4} - the event of getting a perfect square number.

B = {2, 4, 6} - the event of getting an even number.

C = {1, 3, 5} - the event of getting an odd number.

D = the event of getting even or odd number.

E = the event of getting number 8.

Then, $A^c = \{2,3,5,6\}$; B and C are complementary(first it is better to define what complement mean); D is certain ; and E is an impossible event.

Example 1.4: Roll a fair die and flip a balanced coin.

Let A = the die shows an even number, B = the coin shows head. Then, A and B are independent events

Sample space and events

A set S which consists of all possible outcomes of a random experiment is called a *sample space* and each outcome is called *sample point*.

Example 1.4: In tossing a fair die the sample space or a set which contains all the possible outcomes is denoted by S = {1, 2, 3, 4, 5, 6}

Event is any subset of sample space S.

Example 1.5: a) In tossing a coin the sample space S is S = {Head, Tail}. The events will be

A = {Head, Tail}, B = {Head}, C = {Tail} and D = {}.

b) In example 1.3 a. above set A = {1,2,3,4,5,6}, B = {1,2,3,4,5}, C = {1,2,3,4}, D = {1,2,3,4,5}, E = {1,2,3,4,5} and F = {1,2,3,4,5} are events.

Mutually exclusive events*:* Two events A and B are said to be mutually exclusive if they cannot occur simultaneously; *i.e.,* $A \cap B = \phi$.. The intersection of two mutually exclusive sets is empty set.

Event: An event is a subset of the sample space. That is, an event is a collection of sample points, denoted by A, B, C, D, E etc.

- **Simple event**: If an event E consists of a single outcome, then it is called a simple or elementary event.

- **Compound event**: This is a subset containing two or more points in a sample space. It is also called a composite event.

- **Certain event**: This is an event which is sure to occur.
- **Impossible event**: This is an event which will never occur.
- **Complement of an event**: The complement of event A (denoted by A^c *or* A'), consists of all the sample points in the sample space that are not in A.
- **Independent events**: Two events are said to be independent if the occurrence of one is not affected by, and does not affect, the other. If two events are not independent, then they are said to be **dependent**.
- **Equally likely out comes**: If each out come in an experiment has the same chance to occur, then the outcomes are said to be equally likely.

1.3 Review of set theory

Definition1.3:

Set is a collection of well-defined objects. These objects are called elements. Sets usually denoted by capital letters and elements by small letters. Membership for a given set can be denoted by $\in$ to show belongingness and $\notin$ to say not belong to the set.

Description of sets: Sets can be described by any of the following three ways. That is the complete listing method (all element of the set are listed), the partial listing method (the elements of the set can be indicated by listing some of the elements of the set) and the set builder method (using an open proposition to describe elements that belongs to the set).

Example 1.2: The possible outcomes in tossing a six side die

S = {1, 2, 3, 4, 5, 6} or S = {1, 2, . . ., 6} or S = {x: x is an outcome in tossing a six side die}

Types of set

***Universal set*:** is a set that contains all elements of the set that can be considered the objects of that particular discussion.

***Empty or null set*:** is a set which has no element, denoted by {} or ϕ

***Finite set*:** is a set which contains a finite number of elements. (eg.{x: x is an integer, 0 < x < 5})

Infinite set: is a set which contains an infinite number of elements. (eg. $\{x : x \in \Re , x > 0\}$)

Sub set: If every element of set A is also elements of set B, set A is called sub sets of B, and denoted by $A \subseteq B$.

***Proper subset*:** For two sets A and B if A is subset of B and B is not sub set of A, then A is said to be a proper subset of B. Denoted by A $\subset$ B.

Equal sets: two sets A and B are said to be equal if elements of set A are also elements of set B.

***Equivalent sets*:** Two sets A and B are said to be equivalent if there is a one to one correspondence between elements of the two sets.

Set Operation and their Properties

There are many ways of operating two or more set to get another set. Some of them are discussed below.

***Union of sets**:* The union of two sets A and B is a set which contains elements which belongs to either of the two sets. Union of two sets denoted by$\cup$, A $\cup$ B (A union B).

***Intersection of sets**:* The intersection of two sets A and B is a set which contains elements which belongs to both sets A and B. Intersection of two sets denoted by$\cap$, A $\cap$B (A intersection B).

Disjoint sets: are two sets whose intersection is empty set.

***Absolute complement or complement**:* Let U is the universal set and A be the subset of U, then the complement of set A is denoted by A\` is a set which contains elements in U but does not belong in A.

***Relative complement (or differences)*:** The *difference* of set A with respected to set B, written as A\B (or A – B) is a set which contain elements in A that doesn\`t belong in B.

Symmetric difference: of two sets A and B denoted by A Δ B is a set which contain elements which belong in A but not in B and contain elements which belong in B but not in A. That is, A Δ B is a set which equals to (A\B) $\cup$ (B\ A).

Basic Properties of the Set Operations

Let U be the universal set and sets A, B, C are sets in the universe, the following properties will hold true.

1. A $\cup$ B = B $\cup$A (Union of sets is commutative)
2. A $\cup$ (B $\cup$ C) = (A $\cup$B) $\cup$C = A $\cup$B $\cup$ C (Union of sets is associative)
3. A $\cap$ B = B $\cap$ A (Intersection of sets is commutative)

4. $A \cap (B \cap C) = (A \cap B) \cap C = A \cap B \cap C$ (Intersection of sets is associative)
5. $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$ (union of sets is distributive over Intersection)
6. $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$ (Intersection of sets is distributive over union)
7. $A - B = A \setminus B = A \cap B`$
8. If $A \subseteq B$, then $B` \subseteq A`$ or if $A \subset B$ then $B` \subset A$
9. $A \cup \phi = A$ and $A \cap \phi = \phi$
10. $A \cup U = U$ and $A \cap U = A$
11. $(A \cup B)` = A` \cap B`$ De Morgan's first rule
12. $(A \cap B)` = A` \cup B`$ De Morgan's second rule
13. $A = (A \cap B) \cup (A \cap B`)$

Corresponding statement in set theory and probability

Set theory	Probability theory
Universal set, U	Sample space S, sure event
Empty set ϕ	Impossible event
Elements a, b,…	Sample point a, b, c… (Or simple events)
Set A, B, C, . .	Event A, B, C, . .
A	Event A occur
A`	Event A doesn`t occur
$A \cup B$	At least one of event A and B occur
$A \cap B$	Both event A and B occur
$A \subseteq B$	The occurrence of A necessarily imply the occurrence of B
$A \cap B = \phi$	A and B are mutually exclusive (That is, they cannot occur simultaneously)

In many problems of probability, we are interested in events that are actually combinations of two or more events formed by unions, intersections, and complements. Since the concept of set theory is of vital importance in probability theory, we need a brief review.

- The union of two sets A and B, $A \cup B$, is the set with all elements in A or B or both.
- The intersection of A and B, $A \cap B$, is the set that contains all elements in both A & B.

- The complement of A, A^c, is the set that contains all elements in the universal set $\bigcup$ that are not found in A. Some similarities between notions in set theory and that of probability theory are:

In probability Theory	***In set Theory***
i. Event A or Event B	$A \cup B$
ii. Event A and Event B	$A \cap B$
iii. Event A is impossible	$A = \phi$
iv. Event A is certain	$A = \bigcup$
v. Events A and B are mutually exclusive	$A \cap B = \phi$

Again, using Venn-diagram, one can easily verify the following relationships:

1. $A \cup B = (A \cap B') \cup (A \cap B) \cup (B \cap A'),$ noting that the three are mutually exclusive;

$A \cap B' = A - B \, and \, B \cap A' = B - A.$

2. $A \cup B = A \cup (B \cap A')$, *again mutually exclusive*

3. $A = (A \cap B') \cup (A \cap B) \quad and \quad B = (B \cap A') \cup (A \cap B)$.

1.4. Finite and infinite sample space

If a sample space has finite number of points, it is called a *finite sample space*. If it has as many point as natural numbers1, 2, 3,…it is called a *countableinfinite sample space*. If it has as many point as there are in some interval the *x*-axis, such as $0 < x < 1$, it is called a *noncountable infinite sample space*. A sample space which is finite or countable infinite is often called a *discrete sample space* while a set which is non countable infinite is called *non discrete* or *continuous sample space*.

Example 1.6: a) The result of the experiment making bolts observing defective. Thus, the outcome will be a member of the set {defective, non defective}.

b) The lifetime of a bulb in example 1.3 b.

1.5. Equally Likely Outcomes

Equally likely outcomes are outcomes of an experiment which has equal chance (equally probable) to appear. In most cases it is commonly assumed finite or countable infinite sample space is equally likely.

If we have n equally likely outcomes in the sample space then the probability of the i^{th} sample point x_i is p (x_i) = $\frac{1}{n}$, where $x_i$can be the first, second,... or the n^{th} outcome.

Example 1.7: In an experiment tossing a fair die, the outcomes are equally likely (each outcomeis equally probable. Hence,

$$P(x_i = 1) = P(x_i = 2) = P(x_i = 3) = P(x_i = 4) = P(x_i = 5) = P(x_i = 6) = \frac{1}{6}$$

In order to determine the number of outcomes, we can use several rules of counting.

- **The addition rule**
- **The multiplication rule**
- Permutation rule
- Combination rule

1.6. Counting Techniques

In many cases the number of sample points in a sample space is not very large, and so direct enumeration or counting of sample points used to obtain probabilities is not difficult. However, problems arise where direct counting becomes a practical impossibility. To avoid such difficulties we apply the fundamental principles of counting (counting techniques).

Multiplication Rule

Suppose a task is completed in K stages by carrying out a number of subtasks in each one of the K stages. If in the first stage the task can be accomplished in n_1 different ways and after this in the second stage the task can be accomplished in n_2 different ways, . . . , and finally in the K^{th} stage the task can be accomplished in n_k different ways, then the overall task can be done in

$n_1 \times n_2 \times \cdot \cdot \cdot \times n_k$ different ways.

Example 1.8: Suppose that a person has 2 different pairs of trousers and 3 shirts. In how many ways can he wear his trousers and shirts?

Solution: He can choose the trousers $in\ n_1 = 2$ ways, and shirts in $n_2 = 3$ ways. Therefore, he can wear in $n_1 \times n_2 = 2 \times 3 = 6$ possible ways.

Example 1.9:You are eating at a restaurant and the waiter informs you that you have (a) two choices for appetizers: soup or salad; (b) three for the main course: a chicken, fish, or hamberger dish; and (c) two for dessert: ice cream or cake.How many possible choices do you have for your complete meal?

Solution: The menu is decided in three stages at each stage the number of possible choices does not depend on what is chosen in the previous stages. The total number of choices will be the product of the number of choices at each stage. Hence 2 x 3 x 2 = 12 possible menus are available.

Sometimes a tree diagram can be used to illustrate the general counting technique.

Permutations

Suppose that we are given *n* distinct objects and wish to arrange *r* of these objects in a line. Since there are *n* ways of choosing the 1^{st} object, and after this is done, *n - 1* ways of choosing the 2^{nd} object, . . . , and finally *n - r + 1* ways of choosing the r^{th} object, it follows by the fundamental principle of counting that the number of different arrangements or permutations is given by

$n(n - 1)(n - 2) \dots (n - r + 1) = nPr$

where it is noted that the product has r factors. We call *nPr* the number of permutations of *n* objects taken *r* at a time.

In the particular case where *r = n*, the above equation becomes

$nPn = \frac{n!}{(n-n)!} = n(n - 1)(n - 2) \dots 1 = n!$

which is called *n* factorial.

Moreover, we can rewrite nPr in terms of factorials as follow

$nPr = \frac{n!}{(n-r)!}$

If $r = n$, $(n – r)! = 0!$ and we take as the definition of 0!, 0! = 1

Example 1.10: a) The number of different arrangements, or permutations, consisting of 3 letters each that can be formed from the 7 letters A, B, C, D, E, F, G is

${}_7P_3 = \frac{7!}{(7-3)!}$

b) Wonder Woman's invisible plane has 3 chairs. There are 3 people who need a lift. How many seating options are there?

Solution; ${}_3P_3 = \frac{3!}{(3-3)!}$ = 6 different ways.

Remark

- ✍ If a set consists of n objects of which n_1 are of one type (i.e., indistinguishable from each other), n_2 are of a second type, . . . , n_k are of a k^{th} type. Then the number of different permutations of the objects is given by: $_nP_{n_1,n_2,\ldots,n_k} = \frac{n!}{n_1!\,n_2!\,.\,.\,.\,n_k!}$

Example 1.11: The number of different permutations of the 11 letters of the word "M I S S I S S I P P I", which consists of 1 M, 4 I's, 4 S's, and 2 P's, is

$\frac{11!}{1!4!4!2!}$ = 34,650

Combinations

In a permutation we are interested in the order of arrangement of the objects. For example, abc is different permutation from bca. In many problems, however, we are interested only in selecting or choosing objects without regard to order. Such selections are called combinations. For example, abc and bca are the same combination. The total number of combinations of r objects selected from n (also called the combinations of n things taken r at a time) is denoted by $_nC_r$ or $\binom{n}{r}$ or C_r^n.

$$\binom{n}{r} = \frac{n!}{r!\,(n-r)!} = \frac{n(n-1)(n-2).\,.\,.(n-r+1)}{r!} = \frac{_nP_r}{r!}$$

Moreover, we can show that $\binom{n}{r} = \binom{n}{n-r}$

Example 1.12: The number of ways in which 3 cards can be chosen or selected from a total of 8 different cards is

Solution: $\binom{8}{3} = \binom{8}{8-3} = \binom{8}{5}$ = 56

Example 1.13: In how many ways can a committee of 2 students be formed out of 6?

Solution: We substitute n = 6 and r =2 in Formula 1.6, to get $\binom{6}{2} = \frac{6!}{2!.4!} = \frac{6\times 5}{2!} = 15$.

Example 1.14: A committee consisting of 5 candidates is to be formed out of 10, of which 4 are girls and 6 are boys. How many committees can be formed if 2 girls are to be included?

Solution: It can be seen as a two-stage selection. Since 2 of the 4 girls can be selected in $n_1 = {}_4C_2$ =6 ways, and 3 of the 6 boys in $n_2 = {}_6C_3$ =20 ways, then using Formula 1.1, the total number of committees is $n_1 \times n_2 = {}_4C_2 \times {}_6C_3 = 6 \times 20 = 120$.

ACTIVITY 1.1

1. There is no simple formula for the number of combinations and permutations of n non -identical objects when fewer than n are selected.

   a) List all combinations of four letters from the word “bubble".

   b) Find the number of permutations of the list in (a).**[Ans.(a) Hint: use a tree -diagram b) 6 cases.]**

2. In how many different ways can 2 green, 4 yellow, and 3 red bulbs be arranged in a Christmas tree light with 9 sockets?**[Ans. 1260]**

CHECKLIST 1.1

Put a tick mark (√) for each of the following questions if you can solve the problems, and an X otherwise. Can you

1. State the fundamental principles of counting? ☐
2. Discuss about permutations and combinations? ☐
3. Differentiate between permutations and combinations? ☐
4. Write the formulae for the number of permutations and combinations? ☐
5. Partition a set of n distinct objects into k subsets? ☐

Exercise 1.1

1. A psychiatrist is preparing three-letter nonsense words for use in a memory test. The first letter is chosen from k, m, w, and z; the second from a, i and u; the last form b, d, f, k, m, and t.

   a) How many words can he construct? b) How many of them begin with z?

c) How many of them end with k or m? d) How many of them begin and end with the same letter?

2. How many horizontal flags can be formed using 3 colors out of 5 when

   a) Repetition is allowed? b) Repetition is not allowed?

3. In how many ways can five students line up for lunch at café? In how many ways can they line up if two of them refuse to follow each other?

4. Among the seven nominees for two vacancies are 3 men and 4 women. In how many ways can these identical vacancies be filled with: a) any two of the 7 nominees; b)any two of the 4 women; c) one of the men & one of the women?

5. A shipment of 10 TV sets includes three that are defective. In how many ways can a   hotelpurchase four of these sets and receive at least two of the defectives?

6. How many different 3-digit numbers can be made with 3 fours, 4 twos, and 2 threes?

7. In how many ways can 10 objects be split in to two groups containing 4 and 6 objects respectively?

8. From 8 consonants and 4 vowels, how many words can be formed consisting of 3 different consonants and 2 different vowels? The words need not have meaning.

9. In how many ways can 3 men be selected out of 15 if one is always included and two are always excluded?

10. A college team plays 10 football games during a season. In how many ways can it end   the season with five wins, four losses, and one tie?

11. How many different sums of money can be drawn from a pocket containing 1, 5, 10, 25, and 50 cent coins by taking: a) at least one coin; b) at least two coins?

1.7 Concept and Definitions of Probability

In any random experiment there is always uncertainty as to whether a particular event will or will not occur. As a measure of the chance, or probability, with which we can expect the event to occur, it is convenient to assign a number between 0 and 1. If we are sure or certain that the event will occur, we say that its probability is 100% or 1, but if we are sure that the event will not occur, we say that its probability is zero.

There are different procedures by means of which we can define or estimate the probability of an event. These procedures are discussed below:

1. **Classical Approach or Definition of Probability**

Let S be a sample space, associated with a certain random experiment and consisting of finitely many sample points *m*, say, each of which is equally likely to occur whenever the random experiment is carried out. Then the probability of any event A, consisting of *n* sample points ($0 \leq n \leq m$), is given by: $P(A) = \frac{n}{m}$

Example 1.15: What is the probability that a 3 or 5 will turn up in rolling a fair die ?

Solution: S ={1, 2, 3, 4, 5, 6}; let E ={3, 5}. For a fair die, P(1)=P(2) = … =P(6)=1/6; then,

$$P(E) = \frac{m}{n} = \frac{2}{6} = \frac{1}{3}.$$

Example 1.16: In an experiment of tossing a fair coin three times, find the probability of getting

a) exactly two heads; b)at least two heads.

Solution: For each toss, there are two possible outcomes, head (H) or tail (T). Thus, the number of possible outcomes is n =2x2x2=8. And the sample space is given below (a tree-diagram will facilitate listing down elements of S).

S = {HHH, HHT, HTH, HTT, THH, THT, TTH, TTT}

If E_1 = an event of getting 2 heads, then E_1 = {HHT, HTH, THH}, & n(E_1)= m = 3.

Therefore, $P(E_1) = \frac{m}{n} = \frac{3}{4}$.

2. **Relative Frequency Approach or Definition of Probability:**

Let *N(A)* be the number of times an event A occurs in *N* repetitions of a random experiment, and assume that the relative frequency of A, $\frac{N(A)}{N}$, converges to a limit as $N \rightarrow \infty$. This limit is denoted by *P(A)* and is called the probability of A.

Both the classical and frequency approaches have serious drawbacks, the first because the words “equally likely” are vague and the second because the “large number” involved is vague. Because of these difficulties, to avoid such ambiguity an axiomatic approach to probability is preferred.

Example 1.17: Bits & Bytes Computer Shop tracks the number of desktop computer systems it sells over a month (30 days):

Desktops sold	0	1	2	3	4
number of days	1	2	10	12	5

From this we can construct the "estimated" probabilities of an event (i.e. the Number of desktop sold on a given day).

Desktops sold (X)	number of days	Probability of desktop sold P(X)
0	1	1/30 = 0.30 = *P(X=0)*
1	2	2/30 = 0.30 = *P(X=0)*
2	10	10/30 = 0.30 = *P(X=0)*
3	12	12/30 = 0.30 = *P(X=0)*
4	5	5/30 = 0.30 = *P(X=0)*
		$\sum P(X) = 1$

3. **Axiomatic Approach or Definition of Probability:**

Probability is a function, denoted by *P*, defined for each event of a sample space S, taking on values in the real line $\Re$, and satisfying the following three properties (or *axioms of probability*):

Axiom 1: $P(A) \geq 0$ for every event A (probability of an event is nonnegative)

Axiom 2: For the sure or certain event S, $P(S) = 1$

Axiom 3: For countable infinite many pair wise disjoint (mutually exclusive) events A_i, i =1, 2, 3,….., $A_i \cap A_j = \phi$, $i \neq j$, it holds

$P\ (A_1 \cup A_2 \cup A_3\ .\ .\ .) = P(A_1) + P(A_2) + P(A_3)\ .\ .\ .$ or = $P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$

1.8. Some Important Theorems on Probability

Theorem 1: If A\` is the complement of A, then *P(A\`) = 1 - P(A)*

Theorem 2: If $A = A_1 \cup A_2 \cup \ldots \cup A_n$, where $A_1, A_2, \ldots, A_n$ are mutually exclusive events, then, $P(A) = P(A_1) + P(A_2) + \ldots + P(A_n)$

In particular, if A = S, the sample space, then $P(A_1) + P(A_2) + \ldots + P(A_n) = 1$

Theorem 3: If A and B are any two events, then $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

More generally, if A_1, A_2, A_3 are any three events, then

$P(A_1 \cup A_2 \cup A_3) = P(A_1) + P(A_2) + P(A_3) - P(A_1 \cap A_2) - P(A_2 \cap A_3) - P(A_3 \cap A_1) + P(A_1 \cap A_2 \cap A_3)$

Generalizations to n events can also be made.

Theorem 4: For any events A and B, $P(A) = P(A \cap B) + P(A \cap B`)$, (since $A \cap B$) and $(A \cap B`)$ are mutually .exclusive.

Theorem 5: If an event A must result in the occurrence of one of the mutually exclusive events $A_1, A_2, \ldots, A_n$, then

$$P(A) = P(A \cap A_1) + P(A \cap A_2) + \ldots + P(A \cap A_n).$$

ACTIVITY 1.2

In a class of 200 students, 138 are enrolled in a Mathematics course, 115 are enrolled in Statistics, and 91 are enrolled in both. How many of these students take

a) either course; b) neither course; c) Statistics but not Mathematics;d) Mathematics but not Statistics?

CHECKLIST 1.2

Put a tick mark (√) for each of the following questions if you can solve the problems, and an X otherwise. Can you

1.Define the relative frequency and empirical approaches to measure probability? ☐

2. State the shortcomings of both concepts? ☐

3. State the three postulates of probability? ☐

EXERCISE 1.2 (SAME WITH EXERCISE 1.1??? NO SOLUTIONS)

1. If $P(A) = 0.7, P(B) = 0.6$, and $P(A \cup B) = 0.8$, find

   a) $P(A \cap B)$; b) $P(A' \cap B)$; c) $P(A' \cap B')$

2. In a certain city, out of every 100, 000 people, 700 of them own cars, 42,000 own bike, and 450 own both. Find the probabilities that a person randomly selected form this city owns

   a) a car; b) car and bike; c) a bike; d) neither a car nor a bike.

3. A coin is loaded so that $P(H) = 0.52$, and $P(T) = 0.48$. If the coin is tossed three times, what is the probability of getting?

a) all heads; b) two tails and a head in this order; c) two tails & a head in any order?

4. 25% of students in a college graduate with honor while 20 % of them were honor graduates who got good jobs. Find the probability of a randomly chosen graduate to get a good job if he/she graduates with honors?

5. A student is known to answer 3 questions out of 5, and another student 5 out of 7. If a problem is given to both of them, assuming independent work, find the probabilities that a) both; b) any one; c) none; d) only one of them will solve it.

SUMMARY

- ☞ The fundamental principle of counting states that: if a choice has k steps with n_1 different ways for the first, and after this n_2 different ways for the second,…, and finally n_k different ways for the k^{th} step, then the whole choice can be done in $n_1*n_2*n_2*\ldots*n_k$ different ways.
- ☞ The number of permutations of r objects selected out of n different objects is given by

$$_nP_r = \frac{n!}{(n-r)!} \quad for \quad r = 0,\ 1,\ 2, \ldots, n,$$

.

- ☞ All of n different objects can be arranged in n! different ways.

- The number combinations of n distinct objects selecting r of them at a time is: ${}_nC_r = \binom{n}{r} = \frac{n!}{r!(n-r)!} \quad for\ r = 0, 1, 2, \ldots, n.$, and thisgives the number of subsets with r objects.
- **Classical probability concept**: The probability of an event is m/n if it can occur in m ways out of a total of n equally likely ways.
- The **relative frequency concept of probability**: The probability of the occurrence of an event equals its relative frequency.
- The three **axioms of probability** are:
  - Probability is non – negative(always $0 \leq P(A) \leq 1$).
  - Probability of a sample space is unity.
  - $P(A \cup B) = P(A) + P(B)$ if A & Bare mutually excursive.
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$, for any two events.
- For any three events A, B, and C,

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C).$$

CHAPTER 2

CONDITIONAL PROBABILITY AND INDEPENDENCE

Introduction

This chapter is the continuous part of the first one which consists of conditional probability, multiplication rule, partition of a set and independent events.

Contents

2.1 Conditional Probability

2.2 Multiplication rule

2.3 Total probability & Bayes' Theorems and their Applications

2.4 Independent Events

Learning out comes

At the end of this chapter students will be able to:

- ✓ Define conditional probability.
- ✓ Apply multiplication rule to solve problems
- ✓ Apply total probability theorem to solve problems
- ✓ State Baye`s theorem
- ✓ Define independency of two or more events

RESOURCES:

- 🕮 Mathematical Statistics, John E. Freund, 6th Edition, pages 1-72.
- 🕮 Modern Elementary Statistics, Freund and Simon, 9th Edition, pages 105-192.
- 🕮 Statistical Methods, S.P. Gupta, 12th edition, pages A-1.3 -1.56.

2.1. Conditional Probability

Definition 2.1:

The *conditional probability* of an event A, given that event B *has occurred* with $P(B) > 0$, is denoted by $P(A|B)$ and is defined by: $P(A|B) = \frac{P(A\cap B)}{P(B)}$. $P(B)\neq 0$

$Similarly$ the $conditional probability$ of an event B, given that event $A has occurred$ with $P(A) > 0$ is defined as $P(B\backslash A) = \frac{P(A\cap B)}{P(A)}$

- ⇨ $P(A \cap B) = P(A|B)P(B) = P(B)P(A|B)$
- ⇨ $P(A \cap B) = P(B|A)P(A) = P(A)P(B|A)$

Example 2.1: We toss a fair coin three successive times. We wish to find the conditional probability $P(A|B)$ where A and B are the events A = {more heads than tails come up}, B = {1st toss is a head}.

2.2 Multiplication Rule

Multiplication rule (Multiplicative Theorem)

For any n events $A_1, A_2, \ldots, A_n$ with $P(\bigcap_{j=1}^{n-1} A_j) > 0$ it holds

$$P\left(\bigcap_{j=1}^{n} A_j\right) = P(A_n \mid A_1 \cap \ldots \cap A_{n-1})P(A_{n-1} \mid A_1 \cap \ldots \cap A_{n-2}) \ldots P(A_2 \mid A_1)P(A_1).$$

Its significance is that we can calculate the probability of the intersection of n events, step by step, by means of conditional probabilities. The calculation of these conditional probabilities is far easier. Here is a simple example which simply illustrates the point.

Example 2.2: The completion of a construction job may be delayed because of a strike. The probabilities are 0.60 that there will be a strike, 0.85 that the construction job will be completed on time if there is no strike, and 0.35 that the construction job will be completed on time if there is a strike. What is the probability that the construction job will be completed on time?

Solution: If A is the event that the construction job will be completed on time and B is the event that there will be a strike, we are given P(B)=0.60,P(A/B^C)=0.85, and P(A/B)=0.35. Making use of the multiplication rule, we can write

$P(A)= P\{(A \cap B) \cup (A \cap B^C)\} = P(A \cap B) + P(A \cap B^C)$

$= P(B)*P(A/B) + P(B^C)*P(A/B^C) = (0.60)(0.35) + (0.40)(0.85) = 0.55.$

P(A)= 0.55.

2.3 Partition Theorem, Bayes' Theorem and Applications

Theorem 2.1

Let $B_1, B_2, \ldots, B_n$ be a partition of the sample space S, if (mutually exclusive and exhaustive??)

(i) $B_i \cap B_j = \emptyset$ for $i \neq j$ where $i, j = 1, 2, 3, \ldots, n$

(ii) $\bigcup_{i=1}^{n} B_i = S$

(iii) $P(B_i) > 0$ for all i.

Then, for any event A in S, (total thm is not stated here??? Which very crucial for Bayes thm?)

$P(A) = P(A \cap B_1) + P(A \cap B_2) + P(A \cap B_3) + \ldots + P(A \cap B_n)$

$= P(B_1).P(A \backslash B_1) + P(B_2).P(A \backslash B_2) + \ldots + P(B_n).P(A \backslash B_n)$

Example 2.3: The members of a consulting firm rent cars from three rental agencies: 60% from agency I, 30% from agency II, and 10% from agency III. If 9% of the cars from agency I need a tune up, 20% of the cars from agency II need a tune-up, and 6% of the cars from agency III need a tune-up, what is the probability that a rental car delivered to the firm will need a tune-up?

Solution: If A is the event that the car needs a tune-up, and B_1, B_2, and B_3 are the events that the car comes from rental agencies I, II, or III, we have $P(B_1)=0.60$, $P(B_2)=0.30$, $P(B_3)=0.10$, $P(A/B_1)=0.09$, $P(A/B_2)=0.20$, and $P(A/B_3)=0.06$. Then, using the rule of total probability,

$P(A)= P(B_1)* P(A/B_1)+ P(B_2)* P(A/B_2) + P(B_3)* P(A/B_3)$

$= (0.60)(0.09) + (0.30)(0.20) + (0.10)(0.06) = 0.12.$

Thus, 12% of all the rental cars delivered to this firm will need a tune-up.

Baye`s Theorem or Rule

Suppose that $B_1, B_2, . . ., B_n$ are mutually exclusive events whose union is the sample space S. Then if A is any event, we have the following important theorem:

$$P(B_i \setminus A) = \frac{P(B_i)P(A \setminus B_i)}{\sum_{j=1}^{n} P(B_j)P(A \setminus B_j)}$$

Example 2.4: With reference to example 2.3, if a rental car delivered to the consulting firm needs a tune-up, the probability that it came from rental agency II is:

$$P(B_2/A) = \frac{P(B_2) * P(A/B_2)}{P(A)} = \frac{(0.30)(0.20)}{0.120} = 0.5.$$

2.3 Independent Event

Definition 2.2: Two events A_1 and A_2 are said to be independent (statistically or stochastically or in the probability sense), if $P(A_1 \cap A_2) = P(A_1) P(A_2)$.

Two events A_1 and A_2 are said to be dependent when $P(A_1 \cap A_2) \neq P(A_1) P(A_2)$

In other words, two events A_1 and A_2 are independent means the occurrence of one event A_1 is not affected by the occurrence or non-occurrence of A_2 and vice versa.

Remark

- ✍ If two events A and B are independent then $P(B \setminus A) = P(B)$, for $P(A) > 0$ and $P(A|B) = P(A)$ where $P(B) > 0$.

The definition of independent event can be extended in two more than two event as follow:

Definition 2.3:

The events $A_1, A_2, . . . , A_n$ are said to be independent (statistically or stochastically or in the probability sense) if, for all possible choices of k out of n events ($2 \leq k \leq n$), the probability of their intersection equals the product of their probabilities.

More formally, for any k with $2 \leq k \leq n$ and any integer $j_1, j_2 . . . , j_k$ with

$1 \leq j_1 < \cdot \cdot \cdot < j_k \leq n$, we have: $P(\bigcap_{i=1}^{k} A_{j_i}) = \prod_{i=1}^{k} P(A_{j_i})$

NB: If at least one of the relations violates the above equation, the events are said to be dependent.

The intuition behind the independence of a collection of events is analogous to the case of two events. Independence means that the occurrence or non-occurrence of any number of the events from that collection carries no information on the remaining events or their complements.

If we have a collection of three events, A_1, A_2 and A_3, independence amounts in satisfying the four conditions

$P(A_1 \cap A_2) = P(A_1)\, P(A_2)$,

$P(A_1 \cap A_3) = P(A_1)\, P(A_3)$,

$P(A_2 \cap A_3) = P(A_2)\, P(A_3)$,

$P(A_1 \cap A_2 \cap A_3) = P(A_1)\, P(A_2)\, P(A_3)$.

The first three conditions simply assert that any two events are independent, a property known as pair wise independence. But the fourth condition is also important and does not follow from the first three. Conversely, the fourth condition does not imply the first three conditions.

Example 2.5: Pair wise independence does not imply independence. Consider two independent fair coin tosses, and the following events:

H_1 = {1st toss is a head},

H_2 = {2nd toss is a head},

D = {the two tosses have different results}.

The events H_1 and H_2 are independent. To see that H_1 and D are independent, we note that

$$P(D|H_1) = \frac{P(H_1 \cap D)}{P(H_1)} = \frac{1/4}{1/2} = \frac{1}{2} = P(D)$$

Similarly, H_2 and D are independent. On the other hand, we have

$$P(H_1 \cap H_2 \cap D) = 0 \neq \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = P(H_1)P(H_2)P(D)$$

This shows that these three events are not independent.

NB: The equality $P(A_1 \cap A_2 \cap A_3) = P(A_1)P(A_2)P(A_3)$ is not enough for independence.

Example 2.6: Consider two independent rolls of a fair die, and the following events:

A = {1st roll is 1, 2, or 3},

B = {1st roll is 3, 4, or 5},

C = {the sum of the two rolls is 9}.

We have $P(A \cap B) = \frac{1}{6} \neq \frac{1}{2} \cdot \frac{1}{2} = P(A)P(B)$;

$$P(A \cap C) = \frac{1}{36} \neq \frac{1}{2} \cdot \frac{4}{36} = P(A)P(C)$$

$$P(B \cap C) = \frac{1}{12} \neq \frac{1}{2} \cdot \frac{4}{36} = P(B)P(C)$$

Thus the three events A, B, and C are not independent, and indeed no two of these events are independent. On the other hand, we have

$$P(A \cap B \cap C) = \frac{1}{36} = \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{4}{36} = P(A)P(B)P(C)$$

Theorem 2.1:

i. If the events A_1, A_2 are independent, then so are all three sets of events: A_1, $A`_2$; $A`_1$, A2; $A`_1$, $A`_2$.

ii. More generally, if the events $A_1, A_2, \ldots, A_n$ are independent, then so are the events $A`_1, A`_2, \ldots, A`_n$, where $A`_i$ stands either for A_i or $A`_i$, $i = 1, \ldots, n$.

Remark

✍ Let events A_1, A_2, A_3, A_4 are independent, one obtains relations such as $P(A_1 \cup A_2 | A_3 \cap A_4) = P(A_1 \cup A_2)$ Or $P(A_1 \cup A`_2 | A`_3 \cap A_4) = P(A_1 \cup A`_2)$

CHECKLIST 2.1

Put a tick mark (√) for each of the following questions if you can solve the problems and an X otherwise. Can you

1. State the other rules of probability like complementary rule, general addition rule, multiplication rule, etc? ☐

2. Define and compute conditional probability and that ofindependent events? ☐

EXERCISE 2.1

1. Four candidates are short-listed for a vacancy. If A is twice as likely to be elected as B, and B and C are given about the same chance to be elected, while C is twice as likely to be elected as D, what are the probability that (a) C will win; (b) A won't win?

2. In a group of 200 students, 138 are enrolled in a Mathematics course, 115 are enrolled in a Physics course, and 91 are enrolled in both. Draw a suitable Venn-diagram and fill in the numbers associated with the various regions. Then, find the number of students who are enrolled for (a) only Mathematics; b) only Physics; c) either course; d) neither course.

SUMMARY

☞ Conditional probability:

$$P(A/B)=\frac{P(A\cap B)}{P(B)} \text{ if } P(B)\neq 0 \quad \text{or } P(B/A)=\frac{P(A\cap B)}{P(A)},\ if\ \ P(A)\neq 0.$$

☞ Multiplication rule: $P(A\cap B)=P(A)P(B/A)=P(B)P(A/B)$.

☞ A & B are independent if and only if

$P(A\cap B)=P(A)P(B)\ or\ P(A/B)=P(A)\ or\ P(B/A)=P(B).$

CHAPTER 3
ONE-DIMENSIONAL RANDOM VARIABLES

Introduction

In this chapter, we shall study meaning of random variables which may be both discrete and continuous and distribution of these random variables including their properties.

Contents

Learning out comes

At the end of this chapter students will be able to

- ✓ Define Random Variable.
- ✓ Define Distribution Functions for Random Variables
- ✓ List Properties of distribution functions, *F(x)*
- ✓ Define Discrete Random Variables
- ✓ Define Probability Distribution of Discrete Random Variables
- ✓ List properties of Probability Distribution of Discrete Random Variables of apply to solve problems
- ✓ Distribution Functions for Discrete Random Variables

RESOURCES:

- 🕮 Mathematical Statistics, John E. Freund, 6th Edition, pages 1-72.
- 🕮 Modern Elementary Statistics, Freund and Simon, 9th Edition, pages 105-192.
- 🕮 Statistical Methods, S.P. Gupta, 12th edition, pages A-1.3 -1.56.

3.1 Definitions of Random Variables

Definition 3.1:

Let S be a sample space of an experiment and X is a real valued function defined over the sample space S, then X is called a random variable (or *stochastic variable*).

A random variable, usually shortened to r.v. (rv), is a function defined on a sample space S and taking values in the real line $\Re$, and denoted by capital letters, such as X, Y, Z. Thus, the value of the r.v. X at the sample point s is X(s), and the set of all values of X, that is, the range of X, is usually denoted by X(S) or R_X.

The difference between a r.v. and a function is that, the domain of a r.v. is a sample space S, unlike the usual concept of a function, whose domain is a subset of $\Re$ or of a Euclidean space of higher dimension. The usage of the term "random variable" employed here rather than that of a function may be explained by the fact that a r.v is associated with the outcomes of a random experiment. Of course, on the same sample space, one may define many distinct r.vs.

Example 3.1: Assume tossing of three distinct coins once, so that the sample space is S = {HHH, HHT, HTH, THH, HTT, THT, TTH, TTT}. Then, the random variable X can be defined as X(s), X(s) = the number of heads (H's) in S.

Example 3.2:In rolling two distinct dice once. The sample space S is S = {(1, 1), (1, 2), . . . , (2, 1), . . . , (6, 1), (6, 2), . .. , (6, 6)}, a r.v. X of interest may be defined by X(s) = sum of the numbers in the pair S.

Example 3.3:Recordingthe lifetime of an electronic device, or of an electrical appliance. Here S is the interval (0, T) or for some justifiable reasons, S = (0, ∞), a r.v. X of interest is X(s) = s, $s \in S$.

Example 3.4: Measuring the dosage of a certain medication administered to a patient, until a positive reaction is observed. Here S = (0, D) for some suitable D. X(s) = s, s ∈ S, or X(s)= the No of days the patient get sick.

In the examples discussed above we saw r.v.s with different values. Hence, random variables can be categorized in to two broad categories such as discrete and continuous random variables.

3.2 Discrete Random Variables

Definition 3.2:

A random variable X is called discrete (or of the discrete type), if X takes on a finite or countably infinite number of values; that is, either finitely many values such as $x_1, \ldots, x_n$, or countably infinite many values such as $x_0, x_1, x_2, \ldots$.

Or we can describe discrete random variable as, it

- Take whole numbers (like 0, 1, 2, 3 etc.)
- Take finite or countably infinite number of values
- Jump from one value to the next and cannot take any values in between.

Example 3.5: In Example 3.1 and 3.2 above, the random variables defined are discrete r.v.s.

Example 3.6:

Experiment	**Random Variable (X)**	**Variable values**
Children of one gender in a family	Number of girls	0, 1, 2, …
Answer 23 questions of an exam	Number of correct	0, 1, 2, ..., 23
Count cars at toll between 11:00 am &1:00 pm	Number of cars arriving	0, 1, 2, ..., *n*

Probability Distribution of Discrete Random Variables

Definition 3.3:

If X is a discrete random variable, the function given by $f(x) = P(X = x)$ for each x within the range of $\boldsymbol{X}$ is called the *probability distribution* or *probability mass function* of $\boldsymbol{X}$.

Example 3.7: Find the probability mass function corresponding to the random variable X of (what???)

Example 3.1. That is the r.v X = {0, 1, 2, 3}.

Remark

- ✍ The probability distribution (mass) function *f(x)*, of a discrete random variable X, satisfy the following two conditions

1. $f(x) \geq 0$
2. $\sum_x f(x) = 1$, The summation is taken over all possible values of *x*.

Example 3.8: Find the formula of probability distribution of the total number of heads obtained in four tosses of balanced coin?

Solution: $f(x=0) = \frac{\binom{4}{0}}{16} = \frac{1}{16}$,

$$f(x=1) = \frac{\binom{4}{1}}{16} = \frac{4}{16}, \qquad f(x=2) = \frac{\binom{4}{2}}{16} = \frac{6}{16}$$

$$f(x=3) = \frac{\binom{4}{3}}{16} = \frac{4}{16}, \; f(x=4) = \frac{\binom{4}{4}}{16} = \frac{1}{16}$$

Example 3.9: Check whether the function given by $f(x) = \frac{x+2}{25}$, for *x* = 1, 2, 3, 4, 5 is a *p.m.f*?

(is it exercise??? No solution)

3.3 Continuous Random Variables(Distribution functions for Random Variables and Their Properties, doesn't much with the content!!!!)

Definition 3.4:

- ✍ A r.v X is called continuous (or of the continuous type) if *X* takes all values in a proper interval I ⊆ ℜ.

Or we can describe continuous random variables as follows:

- ➢ Take whole or fractional number.
- ➢ Obtained by measuring.
- ➢ Take infinite number of values in an interval.
- ➢ Too many to list like discrete variable

Example 3.10: In Example 3.3 and 3.4 above, the random variables defined are continuous r.v.s

Example 3.11: The following examples are continuous r.v.s

Experiment	Random Variable X	Variable values

Weigh 100 People	Weight	45.1, 78, ...
Measure Part Life	Hours	900, 875.9, …
Ask Food Spending	Spending	54.12, 42, ...
Measure Time Between Arrivals	Inter-Arrival time	0, 1.3, 2.78, ...

Probability Density Function of Continuous Random Variables

Definition 3.5:

A function with values *f(x)*, defined over the set of all real numbers, is called a probability density function of the continuous random variable X if and only if

$P\ (a \leq x \leq b) = \int_a^b f(x)dx$ for any real constant $a \leq b$.

Probability density function also referred as probability densities (*p.d.f.*), probability function, or simply densities.

Remarks

- ✍ The probability density function *f (x)* of the continuous random variable X, has the following properties (satisfy the conditions)
  1. $f(x) \geq 0$ for all *x*, or for $-\infty < x < \infty$
  2. $f(x) = \int_{-\infty}^{\infty} f(x)\,dx = 1$
- ✍ If X is a continuous random variable and *a* and *b* are real constants with $a \leq b$, then

  $P\ (a \leq x \leq b) = P\ (a < x \leq b) = P\ (a \leq x < b) = P\ (a < x < b)$

Example 3.12: Suppose that the r-v X is continuous with the pdf of $f(x) = \begin{cases} 2x, o < x < 1, \\ 0, otherwise \end{cases}$

a) Check that $f(x)$ is a pdf

b) b) Find $P(X < 0.5)$;

c) Evaluate $P\left(X < \frac{1}{2}\ given\ that\ \frac{1}{3} < X < \frac{2}{3}\right)$.(why conditional here??????)

Solution: a) Obviously, for o < X< 1, f(x) >0, and

$$\int_{-\infty}^{\infty} f(x)dx = \int_0^1 f(x)dx = \int_0^1 2xdx = x^2\Big|_0^1 = 1.$$

Hence, $f(x)$ is the pdf of some r-v X.

Note: a) $\int_{-\infty}^{\infty} f(x)dx = \int_{0}^{1} f(x)dx,$ since f(x) is zero in the other two intervals: $(-\infty, 0] \cup [1, \infty)$.

b) $P(X < 0.5) = \int_{0}^{0.5} f(x)dx = \int_{0}^{0.5} 2xdx = x^2\Big|_0^{0.5} = 0.25.$

c) Let $A = \left\{X < \frac{1}{2}\right\}, B = \left\{\frac{1}{3} < X < \frac{2}{3}\right\}, so that A \cap B = \left\{\frac{1}{3} < X < \frac{1}{2}\right\}.$

Then, $P\left(X < \frac{1}{2} \middle| \frac{1}{3} < X < \frac{2}{3}\right) = P(A/B) = \frac{P(A \cap B)}{P(B)}$, where $P(A \cap B) = \int_{1/3}^{1/2} 2xdx = \frac{5}{36}$,

and $P(B) = \int_{1/3}^{2/3} 2xdx = \frac{1}{3}$.

$\therefore P(A/B) = \frac{5/36}{1/3} = \frac{5}{36} \times 3 = \frac{5}{12}.$

1.1. 3.4 Cumulative distribution function and its properties(Probability distribution for Discrete Random Variables ???)

***Definition 3.6:*Distribution Functions for Random Variables**

The cumulative distribution function, or the distribution function, for a random variable X is a function defined by: $F(x) = P(X \leq x)$

Where *x* is any real number, i.e., - ∞ < x < ∞. Thus, the distribution function specifies, for all real values *x*, the probability that the random variable is less than or equal to *x*.

Properties of distribution functions, *F(x)*

1. *0 ≤ F(x) ≤ 1 for all x in R*
2. *F(x)* is non-decreasing [i.e., *F(x) ≤ F(y) if x ≤ y*].
3. F(x) is continuous from the right [i.e., $\lim_{h \to 0^+} F(x+h) = F(x)$ for all x]
4. $\lim_{x \to -\infty} F(x) = 0$ and $\lim_{x \to \infty} F(x) = 1$ (where are the proofs?)

Distribution Functions for Discrete Random Variables

Definition 3.7:

If X is a discrete random variable, the function given by: $F(x) = P(X \leq x) = \sum_{t \leq x} f(t)$ For all x in $\Re$ and t ∈*X, where f(t)* is the value of probability distribution or *p.m.f* of X at *t,* is called the distribution function, or the cumulative distribution function of X.

If X takes on only a finite number of values $x_1, x_2, \ldots, x_n$, then the distribution function is given by:

$$F(x) = \begin{cases} 0 & -\infty < x < x_1 \\ f(x_1) & x_1 \le x < x_2 \\ f(x_1) + f(x_2) & x_2 \le x < x_3 \\ \vdots & \vdots \\ f(x_1) + \cdots + f(x_n) & x_n \le x < \infty \end{cases}$$

Example 3.13: Let X be a continuous r-v with pdf $f(x) = \begin{cases} x, & 0 < x < 1 \\ 2 - x, & 1 \le x < 2 \\ 0, & \text{elsewhere} \end{cases}$

a) Check that $f(x)$ defines a pdf;

Where is b???

Solution:

a) $\int_{-\infty}^{\infty} f(x)dx = \int_0^1 xdx + \int_1^2 (2-x)dx = \left.\frac{x^2}{2}\right|_0^1 + \left.\left(2x - \frac{x^2}{2}\right)\right|_1^2 = \frac{1}{2} + (4-2) - \left(2 - \frac{1}{2}\right) = 1.$

And since $f(x) > 0$ for all x, the function is the pdf of X.

b) $P(0.8 < X < 1.2) = \int_{0.8}^{1.2} f(x)dx = \int_{0.8}^{1} xdx + \int_1^{1.2} (2-x)dx$

$= \left.\frac{x^2}{2}\right|_{0.8}^1 + \left.\left(2x - \frac{x^2}{2}\right)\right|_1^{1.2} = \left(\frac{1}{2} - 0.32\right) + (2.4 - 0.72) - (2 - 0.5) = 0.36.$

Definition 3.8:**Distribution Functions of Continuous Random Variables**

If X is a continuous random variable and the value of its probability density is *f (t),* then function given by $F(x) = P(X \le x) = \int_{-\infty}^{x} f(t)\,dt$ is called the distribution function, or the cumulative distribution of the continuous r.v. X.

Theorem 3.4: If *f (x)* and *F(x)* are the values of the probability density and the distribution function of X at *x*, then $P\ (a \le x \le b) = F(b) - F(a)$

For any real constant a and b with a ≤ b, and

$$f(x) = \frac{dF(x)}{dx} \text{ Where the derivative exist.}$$

Example 3.15: Let X be a continuous r-v with pdf $f(x) = \begin{cases} x, & 0 < x < 1 \\ 2-x, & 1 \le x < 2 \\ 0, & \text{elsewhere} \end{cases}$

Find $P(0.8 < X < 1.2)$. (Redundancy with the above example)

Solution: $P(0.8 < X < 1.2) = \int_{0.8}^{1.2} f(x)dx = \int_{0.8}^{1} xdx + \int_{1}^{1.2} (2-x)dx$

$$= \left.\frac{x^2}{2}\right|_{0.8}^{1} + \left.\left(2x - \frac{x^2}{2}\right)\right|_{1}^{1.2} = \left(\frac{1}{2} - 0.32\right) + (2.4 - 0.72) - (2 - 0.5) = 0.36.$$

Example 3.16: (a)Find the constant C such that the function *f(x)* is the density function of a r.v. *X*, where *f(x)* is given by $f(x) = \begin{cases} Cx^2 & 0 < x < 3 \\ 0 & otherwise \end{cases}$ (b) Compute $P(1 < x < 2)$?

Solution:a) $P(0 < X < 3) = \int_{0}^{3} f(x)dx = \int_{0}^{3} cx^2 dx = 1$

$$= \left. c\frac{x^3}{3}\right|_{0.}^{3} = 1 \quad , \frac{27c}{3} = 1 \quad , \quad c = 1/9$$

b) $P(1 < x < 2) = P(1 < X < 2) = \int_{1}^{2} f(x)dx = \int_{1}^{2} cx^2 dx = \left. c\frac{x^3}{3}\right|_{1}^{2}$ =1/27(8-1)=7/27 ???

not from the question????

EXERCISE 3.1

1. A lot of 12 TV sets includes 2 that are defectives. If 3 sets are selected at random for shipment, how many defective sets are expected? If X has a pdf of $f(x) = 3x^2$, for 0 <x <1, and o elsewhere, find

   a) P(X < 0.5); b) E(X) and V(X);

c) a if $P(X > a) = 0.05$; d) b if $P(X \le b) = P(X > b)$.

2. The amount of bread X (in hundreds of kg) that a certain bakery is able to sell in a day is found to be a continuous r-v with a pdf given as below:

$$f(x) = \begin{cases} kx & , \ 0 \le x < 5 \\ k(10-x) & , \ 5 \le x < 10 \\ 0 & , \text{ otherwise} \end{cases}$$

a) Find k; b) Find the probability that the amount of bread that will be sold tomorrow is

i) More than 500kg, ii) between 250 and 750 kg;

no solutions for all questions ????

SUMMARY

☞ If X is continuous with a pdf $f(x)$ such that $f(x) \ge 0$ and $\int f(x)dx = 1$,over the domain of $f(x)$, then, $p(a < X < b) = \int_a^b f(x)dx$.

☞ For a continuous, r-v X, $P(X \le a) = P(X < a); \quad and \quad P(X \ge b) = P(X > b)$

☞ The value of its probability density is *f (t),* then function given by $F(x) = P(X \le x) = \int_{-\infty}^{x} f(t)\,dt$

CHAPTER 4

FUNCTIONS OF RANDOM VARIABLES

Introduction

In standard statistical methods, the result of statistical hypothesis testing, estimation, or even statistical graphics does not involve a single random variable but, rather, *functions of one or more random variables*. As a result, statistical inference requires the distributions of these functions. In many situations in statistics, we may be interested (it is necessary) to derive the probability distribution of a function of one or more random variables. For instance, a probability model of today's weather, let the random variable X be the temperature in degrees Celsius, and consider the transformation Y = 1.8X + 32, which gives the temperature in degrees Fahrenheit. In this example, Y is a linear function of X, of the form Y = g(X) = a X + b, or the use of **averages of random variables** is common. In addition, sums and more general linear combinations are important. We are often interested in the distribution of sums of squares of random variables, particularly in the use of analysis of variance techniques. In the following sections the probability distribution (pmf and pdf) of a function of one random variable will be discussed.

Contents

4.1 Equivalent events

4.2 Functions of discrete random variables

4.3 Functions of continuous random variables

Learning Outcomes

At the end of this chapter students will be able to:

- ✓ Define Functions of Random Variables.
- ✓ Define Equivalent events
- ✓ Relate if two events are equivalent then their probabilities are equal
- ✓ Apply Theorems on Functions of discrete random variables

RESOURCES:

📖 Statistics for Engineering and the Sciences, William M. & Terry S. (2007), 5th ed., pp. 221-225.

📖 Probability & Statistics for Engineers & Scientists, Sharon L. Myers et al. (2012), 9th ed., pp. 211-218.

4.1. Equivalent Events

Let X be a random variable defined on a sample space, S, and let Y be a function of X. then Y is also a random variable. Define R_x and R_y called the range space of X and Ycan take. Let $C \subset R_y$ and $B \subset R_x$ defined as: $B = \{X \in R_x: Y(X) \in C\}$ then the event B and C are called equivalent events. Or if B and C are two events defined on different sample spaces, saying they are equivalent means that one occurs if and only if the other one occurs.

Definition 4.1.1:

Let E be an experiment and S be its sample space. Let X be a random variable defined on S and let R_x be its range space. Let B be an event with respected to R_x, that is, $B \subseteq R_x$, suppose that A is defined as $A = \{s \in S: X(s) \in B\}$, and we say A and B are equivalent events.

Example 4.1: In tossing two coins the sample space S = {HH, HT, TH, TT}. Let the random variable X = Number of heads, $R_x = \{0, 1, 2\}$. Let $B \subseteq R_x$ and B = {1}. Moreover X (HT) = X (TH) = 1. If A = {HT, TH} then A and B are equivalent events.

Example 4.2: Let X is a discrete random variable on scores of a die and $Y = X^2$, then Y is a discrete random variable as X is discrete. Therefore, the range sample space of X is $R_x = \{1, 2, 3, 4, 5, 6,\}$ and the range sample space of Y is $R_y = \{1, 4, 9, 16, 25, 36\}$. Now,

$\{Y = 4\}$ is equivalent to $\{X = 2\}$

$\{Y < 9\}$ is equivalent to $\{X < 3\}$

$\{Y \leq 25\}$ is equivalent to $\{X \leq 5\}$etc.

Example 4.3: Let X be a continuous random variable taking value in [0,2] and $Y = X^2 + 1$. Now,

$\{Y = 3\}$ is equivalent to $\{X = \sqrt{2}\}$

$\{Y > 4\}$ is equivalent to $\{X > \sqrt{3}\}$

$\{4 < Y \leq 5\}$ is equivalent to $\{\sqrt{3} < X \leq 2\}$ etc.

Definition 4.1.2:

Let B be an event in the range space R_x of the random variable X, we define *P(B)* as *P(B) = P(A)* where $A = \{s \in S: X(s) \in B\}$.From this definition, we saw that if two events are equivalent then their probabilities are equal.

Definition 4.1.3:

Let X be a random variable defined on the sample space S, let R_x be the range space of Xand let H be the real valued function and consider the random variable H(x) = Y with range space R_y, for any event $C \subseteq R_y$, we define *P(C)* as *P(C)* = *P(*{x ε R_x: H(x) ε C}). This means the probability of an event associated with the sample space Y is defined as the probability of equivalent event in the range space of X.

ACTIVITY 4.1

Let X is a continuous random variable with *p.d.f.* $f(x) = \begin{cases} e^{-x} & x > 0 \\ 0 \text{ elsewhere} \end{cases}$ (a) If Y = H(x) = 2x +1 (a)determine the range space of Y. (b) suppose event C is defined as C = {y ≥ 5}, determine the event B = {x ε R_x: H(x) ε C} (c) Determine P(y ≥ 5) from event B?**[Ans.(a) Ry =(y > 0), (b) B = {x≥ 2} (c) e^{-2}]**

4.2. Functions of discrete random variables

If X is a discrete or continuous random variable and Y is a function of X, then it follows immediately that Y is also discrete or continuous.

Definition 4.2:

Suppose that *X* is a **discrete** random variable with probability distribution p(*x*).Let *Y* = g(*X*) define a one-to-one transformation between the values of *X* and*Y* so that the equation *y* = *g*(*x*) can be uniquely solved for *x* in terms of *y*, say*x* = *w*(*y*). Then the probability distribution of *Y* is

p(*y*) = *p*[*w*(*y*)].

Example 4.4: Let *X* be a random variable with probability distribution $p(x) = \left(\frac{3}{4}\right)\left(\frac{1}{4}\right)^{x}$, *x*= 1, 2, 3, . . . then find the probability distribution of the random variable $Y = X^2$.

Solution: Since the values of *X* are all positive, the transformation defines a one-to-onecorrespondence between the *x* and *y* values, $y = x^2$and $x = \sqrt{y}$. Hence $p(y) = p(\sqrt{y}) = \left(\frac{3}{4}\right)\left(\frac{1}{4}\right)^{\sqrt{y}}$, *y*= 1, 4, 9, . . . , and 0, elsewhere.

Example 4.5: If X is the number of heads obtained in four tosses of a balanced coin, find the probability distribution of $H(X) = \frac{1}{1+X}$.

Solution:The sample space S = {HHHH, HHHT, HHTH, HTHH, THHH, HHTT, HTHT, HTTH, TTHH, THTH, THHT, HTTT, TTTH, TTHT, THTT, TTTT}

x	*0*	*1*	*2*	*3*	*4*

p(x)	1/16	4/16	6/16	4/16	1/16

Then, using the relation y = 1/ (1 + x) to substitute values of Y for values of X, we find the probability distribution of Y

y	1	1/2	1/3	1/4	1/5
p(y)	1/16	4/16	6/16	4/16	1/16

Example 4.6: Let X be random variable which assumes -1, 0 and 1 with probability values 1/3, ½ and 1/6 respectively. Let H(x) = 3x + 1 then what is the respective possible values of H(x)?

Solution:the possible values of H(X) = -2, 1 and 4 with probability 1/3, ½ and 1/6 respectively.

4.3. Functions of continuous random variables

A straight forward method of obtaining the probability density function of continuous random variables consists of first finding its distribution function and then the probability density by differentiation. Thus, if X is a continuous random variable with probability density *f(x)*, then the probability density of Y = H(X) is obtained by first determining an expression for the probability

$G(y) = P(Y \leq y) = P(H(X) \leq y)$ and then differentiating

$$g(y) = \frac{d\,G(y)}{dy}$$

Finally determine the values of y where $g(y) > 0$.

To find the probability distribution of the random variable $Y = u(X)$ when X is a continuous random variable and the transformation is one-to-one, we shallneed the following definition.

Definition 4.3:

Suppose that X is a **continuous** random variable with probability distribution $f(x)$. Let $Y = g(X)$ define a one-to-one correspondence between the values of X and Y so that the equation $y = g(x)$ can be uniquely solved for x in terms of y, say $x = w(y)$. Then the probability distribution of Y is $f(y) = f[w(y)]|J|$, where $J = w'(y)$ and is called the **Jacobian** of the transformation.

Remarks

- ✍ Suppose that X_1 and X_2 are **discrete** random variables with joint probability distribution $p(x_1, x_2)$. Let $Y_1 = g_1(X_1, X_2)$ and $Y_2 = g_2(X_1, X_2)$ define a one-to-one transformation between the points (x_1, x_2) and (y_1, y_2) so that the equations $y_1 = g_1(x_1, x_2)$ and $y_2 = g_2(x_1, x_2)$ may be uniquely solved for x_1 and x_2 in terms of y_1 and y_2, say $x_1 = w_1(y_1, y_2)$ and $x_2 = w_2(y_1, y_2)$. Then the joint probability distribution of Y_1 and Y_2 is $g(y_1, y_2) = p[w_1(y_1, y_2), w_2(y_1, y_2)]$.

Dear student, to find the joint probability distribution of the random variables $Y_1 = g_1(X_1, X_2)$ and $Y_2 = g_2(X_1, X_2)$ when X_1 and X_2 are continuous and the transformation is one-to-one, we need an additional definition, analogous to definition 4.3.

- ✍ Suppose that X_1 and X_2 are **continuous** random variables with joint probability distribution $f(x_1, x_2)$. Let $Y_1 = g_1(X_1, X_2)$ and $Y_2 = g_2(X_1, X_2)$ define a one-to-one transformation between the points (x_1, x_2) and (y_1, y_2) so that the equations $y_1 = g_1(x_1, x_2)$ and $y_2 = g_2(x_1, x_2)$ may be uniquely solved for x_1 and x_2 in terms of y_1 and y_2, say $x_1 = w_1(y_l, y_2)$ and $x_2 = w_2(y_1, y_2)$. Then the joint probability distribution of Y_1 and Y_2 is $g(y_1, y_2) = f[w_1(y_1, y_2), w_2(y_1, y_2)]|J|$, where the **Jacobian** is the 2 × 2 determinant

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

and $\frac{\partial x_1}{\partial y_1}$ is simply the derivative of $x_1 = w_1(y_1, y_2)$ with respect to y_1 holding y_2 constant, referred to in calculus as the partial derivative of x_1 with respect to y_1. The other partial derivatives are defined in a similar manner.

Example 4.7: Let X be a continuous random variable with probability distribution

$f(x) = \begin{cases} \frac{x}{12} \text{ for } 1 < x < 5 \\ 0 \quad\quad \text{elsewhere} \end{cases}$ Then find the probability distribution of the random variable $Y = 2X - 3$.

***Solution*:** The inverse solution of $y = 2x - 3$ yields $x = (y + 3)/2$, from which we obtain

$J = w'(y) = dx/dy = 1/2$. Therefore, using Theorem 4.3, we find the density function of Y to be

$f(y) = \frac{y+3}{12}\frac{1}{2} = \frac{y+3}{48}$, $-1 < y < 7$, *and* 0, elsewhere.

Example 4.8: Let $X1$ and $X2$ be two continuous random variables with joint probability distribution $f(x_1, x_2) = 4x_1x_2$, $0 < x1 < 1$, $0 < x2 < 1$, and 0, elsewhere. Then find the joint probability distribution of $Y_1 = X_1^2$ and $Y_2 = X_1X_2$.

***Solution*:** The inverse solutions of $y_1 = x_1^2$ and $y_2 = x_1x_2$ are $x_1 = \sqrt{y_1}$ and $x_2 = \frac{y_2}{\sqrt{y_1}}$, from which we obtain: $J = \frac{1}{2y_1}$. Finally, from the above remarks the joint probability distribution of Y_1 and Y_2 is $g(y_1, y_2) = \frac{2y_2}{y_1}$, $y_2^2 < y_1 < 1$, $0 < y_2 < 1$, *and 0*, elsewhere.

ACTIVITY 4.2:

1. If the probability density of X is given by $f(x) = \begin{cases} 6x(1-x) & \text{for } 0 < x < 1 \\ 0 & \text{elsewhere} \end{cases}$

   Find the probability density of $Y = X^3$ **[Ans. f(y) = 2($y^{-1/3}$ -1), 0 < y < 1].**
2. Let a random variale X has pdf given by $(x) = \begin{cases} 2x & 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$. Let $H(x) = e^{-x}$ be random variable defined on X. then find the *p.d.f* of H(x)? **[Ans. f(y) = 2/y(-ln(y), 1/e < y < 1].**

SUMMARY

- If two events are equivalent then their probabilities are equal.
- Suppose that *X* is a **discrete** random variable with probability distribution p(*x*) and let *Y* = g(*X*), then the probability distribution of *Y* is *p*(*y*) = *p*[*w*(*y*)].
- If X is a continuous random variable with probability density *f(x)*, then *G (y) = P(Y ≤ y) = P (H(X) ≤ y)* and then differentiating $g(y) = \frac{dG(y)}{dy}$
- Suppose that *X* is a **continuous** random variable with probability distribution *f*(*x*) and let *Y* = g(*X*), then the probability distribution of *Y* is f(*y*) = *f*[*w*(*y*)]|*J*|, where *J* = *w'*(*y*) and is called the **Jacobian** of the transformation.

CHECKLIST 4.1

Put a tick mark (√) for each of the following questions if you can solve the problems and an X otherwise. Can you

1. State equivalent events? ☐
2. Differentiate between discrete and continuous functions of random variables? ☐
3. Drive probability distributions for functions of random variables? ☐

Exercise 4.1

1. Suppose that the discrete random variable X assumes the values 1, 2 and 3 with equal probability. What is the range space of Y if Y = 2X + 3?
2. Suppose that the discrete random variable X assumes the values -1, 0 and 1 with the probabilities of 1/3, 1/2, and 1/6 respectively. (a) What is the range space of Y, if $Y = X^2$? (b) Find the probability mass function of Y.
3. Suppose X has a pdf of f(x) = 1, 0 < x < 1, then what is the pdf of Y if $Y = X^2$?
4. Let X has a pdf of f(x) = ½, -1 < x < 1, then: find the pdf of Y if (a) $Y = X^2$ (b) Y = 2x +1 and hence find P(Y < 2).
5. Suppose that X has pdf of f(x0 = 2x, 0 < x < 1, then find pdf of Y if Y = ½ x-3.

CHAPTER 5

TWO OR MORE DIMENSION RANDOM VARIABLES

Introduction

Our study of random variables and their probability distributions in the precedingsections is restricted to one-dimensional sample spaces, in that we recordedoutcomes of an experiment as values assumed by a single random variable. Therewill be situations, however, where we may find it desirable to record the simultaneous outcomes of several random variables. For example, we might measure theamount of precipitate P and volume V of gas released from a controlled chemicalexperiment, giving rise to a two-dimensional sample space consisting of the outcomes(p, v), or we might be interested in the hardness H and tensile strength $T$of cold-drawn copper, resulting in the outcomes (h, t). In a study to determine thelikelihood of success in college based on high school data, we might use a threedimensionalsample space and record for each individual his or her aptitude testscore, high school class rank, and grade-point average at the end of freshman yearin college.

Contents

5.1 Definitions of two dimensional random variables

5.2 Joint distributions

5.3 Marginal probability distribution

5.4 Conditional probability distribution

5.5 Independent random variables

5.6 n –dimensional random variables

Learning Outcomes

At the end of this lecture you will be able to

- ✓ Define two dimensional random variables.
- ✓ Differentiatetwo dimensional discrete and continuous random variable
- ✓ Give example for two dimensional discrete and continuous random variables.
- ✓ Define Joint Probability Distribution.
- ✓ Solve Problems related to joint probability distribution.

RESOURCES:

- 🕮 Statistics for Engineering and The Sciences, William M. & Terry S. (2007), 5th ed., pp. 211-217.
- 🕮 Probability & Statistics for Engineers & Scientists, Sharon L. Myers et al. (2012), 9th ed., pp. 94-104

5.1 Definitions of Two-dimensional Random Variables

We are often interested simultaneously in two outcomes rather than one. Then with each one of these outcomes a random variable is associated, thus we are furnished with ***two random variables*** or a ***2-dimensional random vector*** denoted by (X, Y).

Definition 5.1:

Let (X, Y) is a two-dimensional random variable. (X, Y) is called *a two dimensional discrete random variable* if the possible values of (X, Y) are finite or countably infinite. That is the possible values of (X, Y) may be represented as $(x_i, y_j), i = 1, 2, ...,n, \dots$ and $j = 1, 2, . . . , m, ...$

Let (X, Y) is a two-dimensional random variable. (X, Y) is called *a two dimensional continuous random variables* if the possible values of (X, Y) can assume all values in some non countable set of Euclidian space. That is, (X, Y) can assume values in a rectangle $\{(x,y): a \leq x \leq b \text{ and } c \leq y \leq d\}$ or in a circle $\{(x,y): x^2 + y^2 \leq 1\}$ etc.

5.2. Joint Probability Distribution

If X and Y are two random variables, the probability distribution for their simultaneous occurrence can be represented by a function with values p(x, y) for any pair of values (x, y) within the range of the random variables X and Y. It is customary to refer to this function as the **joint probability distribution**of X and Y.

Definition 5.2.1: Let (X, Y) is a two-dimensional discrete random variablesthat is the possible values of (X, Y) may be represented as $(x_i, y_j), i = 1, 2, ...,n, \dots$ and $j = 1, 2, . . . , m, ...$Hence, in the discrete case, $p(x, y) = P(X = x, Y = y)$; that is, the values $p(x, y)$ give the probability that outcomes x and y occur at the same time, then the function p(x, y) is a **joint probability distribution** or **probability mass function** of the discrete random variables X and Y*if:*

1. $P(x_i. y_j) \geq 0$ for all (x, y)
2. $\sum_x \sum_y f(x, y) = 1$

Example 5.1: Two ballpoint pens are selected at random from a box that contains 3 blue pens, 2 red pens, and 3 green pens. If X is the number of blue pens selected and Y is the number of red pens selected, then find the joint probability mass function $p(x, y)$ and verify that it is pmf.

Solution: The possible pairs of values (x, y) are (0, 0), (0, 1), (1, 0), (1, 1), (0, 2), and (2, 0). Now, $p(0, 1)$, for example, represents the probability that a red and a greenpens are selected. The total number of equally likely ways of selecting any 2pens from the 8 is $\binom{8}{2}$= 28. The number of ways of selecting 1 red from 2red pens and 1 green from 3 green pens is $\binom{2}{1}\binom{3}{1}$ = 6. Hence, p(0, 1) =$\frac{6}{28}$ = $\frac{2}{14}$. Similar calculations yield the probabilities for the other cases, whichare presented in the following Table.

Joint Probability Distribution

(Y, X)	0	1	2	$p_y(y)$
0	$\frac{3}{28}$	$\frac{9}{28}$	$\frac{3}{28}$	$\frac{15}{28}$
1	$\frac{3}{14}$	$\frac{3}{14}$	0	$\frac{3}{7}$
2	$\frac{1}{28}$	0	0	$\frac{1}{28}$
$p_x(x)$	$\frac{5}{14}$	$\frac{15}{28}$	$\frac{3}{28}$	1

The probabilities sum to 1is shows that it is probability mass function.Note that, the joint probability mass function of the above Table canbe represented by the formula: $p(x, y)$ =$\frac{\binom{3}{x}\binom{2}{y}\binom{2}{1}\binom{3}{2-x-y}}{\binom{8}{2}}$, for x = 0, 1, 2; y = 0, 1, 2; and $0 \le x + y \le 2$.

Example 5.2: Consider two discrete random variables, X and Y, where x=1 or x=2, and y=0 and y=1. The bivariate probability mass function for X and Y is defined as follows. p(x, y)= $\frac{0.25+x-y}{5}$, consider the joint probability function and then verify that the properties of a discrete joint probability mass function are satisfied.

Solution: Since X takes on two values (1 or 2) and Y takes on two values (0 or 1), there are 2x2 = 4 possible combinations of X and Y. these four (x, y) pairs are (1,0), (1,1), (2, 0), and (2, 1). Substituting these possible values of X and Y into the formula for p(x, y), we obtain the following joint probabilities.

	X	1	2
Y	0	0.25	0.45
	1	0.05	0.25

The probabilities sum to 1 and all values are nonnegativeare shows that it is probability mass function.

Definition 5.2.2:

Let (X, Y) is a two dimensional continuous random variables assuming all values in some region *R* of the Euclidian space that is, (X, Y) can assume values in a rectangle $\{(x,y): a \leq x \leq b \text{ and } c \leq y \leq d\}$ or in a circle $\{(x,y): x^2 + y^2 \leq 1\}$ etc, then the function $f(x, y)$ is a **joint density function** of the continuous random variables *X* and *Y* if: (1) $f(x, y) \geq 0$ for all (x, y) ∈ R and

(2) $\int\int f(x,y)dxdy = 1$

Examples 5.3: The joint probability function of two continuous random variables *X* and *Y* is given by $f(x, y) = c(2x + y)$, where*x* and *y* can assume all integers such that $0 \leq x \leq 2$, $0 \leq y \leq 3$, and$f(x, y) = 0$ otherwise.

a) Find the value of the constant *c*?
b) Find $P(X \leq 2, Y \leq 1)$?

Solution: (a) $\int_0^3\int_0^2 c(2x+y)dxdy = 1$ =c $\int_0^3[x^2+yx]_0^2\,dy = c\int_0^3(4+2y)\,dy = c[4y+y^2]_0^3$

= 21c then c =1/21.

(b) p(X ≤2, Y ≤1) = $\int_0^1\int_0^2 \frac{1}{21}(2x+y)dxdy = \frac{1}{21}\int_0^1[x^2+yx]_0^2\,dy$

$= \frac{1}{21}\int_0^1(4+2y)\,dy = \frac{1}{21}[4y+x^2]_0^1 = \frac{1}{21}$ 4+1) = $\frac{5}{21}$

ACTIVITY 5.1:

1. A privately owned business operates both a drive-in facility and a walk-in facility. On a randomly selected day, let *X* and *Y* , respectively, be the proportions of the time that the drive-in and the walk-in facilities are in use, and suppose that the joint density function of these random variables is: $f(x,y) = \frac{6}{7}\left(x^2 + \frac{xy}{2}\right)$, $0 < x \leq 1, \quad 0 < y \leq 2$, then verify that it is a pdf and find that p(0 < x < ½ , ¼ < y < ½). ***[Ans. 13/160]***
2. The joint density function of two continuous random variables X and Y is

$$f(x,y) = \begin{cases} cxy & 0 < x < 4,\ 1 < y < 5 \\ 0 & otherwise \end{cases}$$

(a) Find the value of the constant c?(b)Find $P(X \geq 3,\ Y \leq 2)$?**[Ans. (a) 1/96 (b) 7/128]**

Definition 5.2.3:

A function closely related to the probability distribution is the cumulative distribution function, CDF. If (X, Y) is a two-dimensional random variable, then the cumulative distribution function is defined as follows.

Let (X, Y) is a two-dimensional discrete random variable, then the *joint distribution* or *joint cumulative distribution function*,CDF of (X, Y) is defined by F(x, y) = $P(X \leq x,\ Y \leq y)$

$= \sum_{s \leq x} \sum_{t \leq y} p(s,t)$, $s \leq x$, $t \leq y$ for $-\infty < x < \infty$ and $-\infty < y < \infty$, where $p(s, t)$ is the joint probability mass function of (X, Y) at (s, t).

Let (X, Y) is a two dimensional continuous random variable, then the *joint distribution* or *joint cumulative distribution function*,CDF of (X, Y) is defined by F(x, y) = $P(X \leq x,\ Y \leq y)$

$= \int_{-\infty}^{y} \int_{-\infty}^{x} f(s,t)\, ds\, dt$ for $-\infty < x < \infty$ and $-\infty < y < \infty$, where $f(s, t)$ is the joint probability density function of (X, Y) at (s, t).

Remark:

- ✍ If F(x, y) is joint cumulative distribution function of a two dimensional random variable (X, Y) with joint p.d.f $f(x, y)$, then: $f(x,y) = \frac{d^2 F(x,y)}{dx\, dy}$.

ACTIVITY5.2:

Suppose F(x, y) = ½ x^2y, for 0 < x < 2, 0 < y < 1, then verify f(x, y) and find p(Y <$\frac{X}{2}$). **[Ans. it is pdf and p(Y <$\frac{X}{2}$) = ½]**

5.3 Marginal Probability Distributions

In a two dimensional random variable (X, Y) we associated two one dimensional random variables X and Y. Sometime we may be interested in the probability distribution of X or Y. Given the joint probability distribution $p(x, y)$ of the discrete random variables$X$ and Y, the probability distribution $p_x(x)$ of X alone is obtained by summing p(x, y) over the values of Y. Similarly, the probability distribution $p_y(y)$ of Y aloneis obtained by summing p(x, y) over the values of X. We define $p_x(x)$ and $p_y(y)$ tobe the **marginal distributions** of X and Y , respectively. When X and Y arecontinuous random variables, summations are replaced by integrals. Dear students, we can nowmake the following general definition for the marginal distributions.

Definition 5.3.1:

If X and Y are two-dimensional discrete random variables and *p(x, y)* is the value of their joint probability mass function at (x, y), the function given by $p_x(x) = \sum_y p(x, y)$ for each *y* within the range of X is called *the marginal distributionof X*. Similarly, the function given by$p_y(y) = \sum_x p(x, y)$ for each *x* within the range of Y is called *the marginal distribution of Y*.

The term *marginal* is used here because, in the discrete case, the values of *g*(*x*)and *h*(*y*) are just the marginal totals of the respective columns and rows when thevalues of *f*(*x, y*) are displayed in a rectangular table.

Examples 5.4:Consider two discrete random variables, X and Y with the joint probability mass function of X and Y:

Y	X	1	2
	0	0.25	0.45
	1	0.05	0.25

Then construct the marginal probability mass function of X and Y.

Solution:

x	1	2	Total	y	0	1	Total
$P_x(x)$	0.3	0.7	1	$P_y(y)$	0.7	0.3	1

Example 5.5: Two ballpoint pens are selected at random from a box that contains 3 blue pens, 2 red pens, and 3 green pens. If *X* is the number of blue pens selected and *Y* is the number of red pens selected have the joint probability mass function *p*(*x*, *y*) as shown below. Then verify that the column and row totals are the marginal of X and Y, respectively.

(X, Y)	0	1	2
0	$\frac{3}{28}$	$\frac{9}{28}$	$\frac{3}{28}$
1	$\frac{3}{14}$	$\frac{3}{14}$	0
2	$\frac{1}{28}$	0	0

Solution:

X	0	1	2	Total	y	0	1	2	Total
$p_x(x)$	$\frac{5}{14}$	$\frac{15}{28}$	$\frac{3}{28}$	1	$P_y(y)$	$\frac{15}{28}$	$\frac{3}{7}$	$\frac{1}{28}$	1

Definition 5.3.2:

If X and Y are two-dimensional continuous random variables and *f(x, y)* is the value of their joint probability density function at (x, y), the function given by $f_x(x) = \int_{-\infty}^{\infty} f(x, y)\, dy$ *for* $-\infty \le x \le \infty$ is called *the marginal distributionof X*. Similarly, the function given by$f_y(y) = \int_{-\infty}^{\infty} f(x, y)\, dx$ *for* $-\infty \le y \le \infty$ is called *the marginal distribution of Y*.

ACTIVITY5.3:

1. Let X and Y be continuous random variables having joint density function

$$f(x, y) = \begin{cases} c(x^2 + y^2) & 0 \le x \le 1, \;\; 0 \le y \le 1 \\ 0 & otherwise \end{cases}$$

, then find the constant c and the marginal distribution functions of X and Y.**[Ans. c = 3/2, $f_x(x) = 3/2(x^2 + 1/3)$ and $fy(y) = 3/2(1/3 + y^2)$]**

2. The joint probability mass function of two discrete random variables *X* and *Y* is given by

$$f(x, y) = \begin{cases} c\,x, \; for\; 0 \le x^2 \le y \le x \le 1 \\ 0 \qquad\qquad \text{otherwise} \end{cases}$$

a) Determine a constant c. **[Ans. c = 12]**

b) Construct the marginal probability functions of X and Y. **[Ans. $f_x(x) = 12(x^2 – x^3)$]**

c) find P(y < ½ / x = ½) **[Ans. 1/5]**

Remark

- ✍ The fact that the marginal distributions $p_x(x)$ and $p_y(y)$ are indeed the probability distributions of the individual variables X and Y alone can be verified by showing that the conditions of probability distributions stated in the one-dimensional case are satisfied.

5.4 Conditional Probability Distributions

In one-dimensional random variable case, we stated that the value x of the random variable X representsan event that is a subset of the sample space. If we use the definition of conditionalprobability as stated in Chapter 2, P(B/A) = $\frac{A \cap B}{p(A)}$, provided p(A) > 0, where A and B are now the events defined by $X = x$ and $Y = y$, respectively, then

$P(Y = y \mid X = x) = \frac{p(X=x, Y=y)}{p(X=x)} = \frac{p(x,\ y)}{px(x)}$, provided $p_x(x) > 0$,where X and Y are discrete random variables.

It is clear that the function $\frac{p(x,\ y)}{px(x)}$, which is strictly a functionof y with x fixed, satisfies all the conditions of a probability distribution. Thisis also true when f(x, y) and $fx(x)$ are the joint probability density function and marginal distribution,respectively, of continuous random variables. As a result, it is extremely importantthat we make use of the special type of distribution of the form $\frac{f(x,\ y)}{fx(x)}$, inorder to be able to effectively compute conditional probabilities. This type of distributionis called a **conditional probability distribution**; the formal definitions are given as follows.

Definition 5.4.1:

The probability of numerical event X, given that the event Y occurred, is the conditional probability of X given Y = y. A table, graph or formula that gives these probabilities for all values of Y is called the conditional probability distribution for X given Y and is denoted by the symbol p(x/y).

Therefore, let X and Y be discrete random variables and let p(x, y) be their joint probability mass function, then the conditional probability distributions for X and Y is defined as: p(x/y) = $\frac{p(x,y)}{p_y(y)}$, provided $p_y(y) > 0$. Similarly, the conditional probability distribution of X given that $Y = y$ is defined as: p(y/x) = $\frac{p(x,y)}{p_x(x)}$, provided $p_x(x) > 0$.

Again, let X and Y be continuous random variables and let f(x, y) be their joint probability density function, then the conditional probability distributions for X and Y is defined as: $f(x/y) = \frac{f(x,y)}{f_y(y)}$, provided $f_y(y) > 0$. Similarly, the conditional probability distribution of *X* given that *Y* = *y* is defined as: $f(y/x) = \frac{f(x,y)}{f_x(x)}$, provided $f_x(x) > 0$.

Examples 5.6:The joint probability mass function of two discrete random variables X and Y is given by *p(x, y)* = *cxy* for *x* = 1, 2, 3and *y* = 1, 2, 3, and zero otherwise. Then find the conditional probability distribution of X given Yand Y given X.

Solution: first $\sum\sum cxy = 1$ = c(1x1 + 1x2 + …+ 3x2 + 3x3) = 1, then c = 1/36 and finally P(x,y) = (xy)/36. Therefore, $p(X/Y) = \frac{p(x,y)}{p_y(y)} = \frac{xy/36}{\sum_{\forall x} p(x,y)} = \frac{y}{6}$, y = 1, 2, 3 and $p(Y/X) = \frac{xy/36}{\sum_{\forall y} p(x,y)} = \frac{x}{6}$, x = 1, 2, 3.

Example 5.7:A software program is designed to perform two tasks, A and B. let X represent the number of IF-THEN statement in the code for task A and let Y represent the number of IF-THEN statements in the code for task B. the joint probability distribution p(x, y) for the two discrete random variables is given in the accompanying table.

		X					
Y		0 0	1	2	3	4	5
	0	0.000	0.050	0.025	0.000	0.025	0.000
	1	0.200	0.050	0.000	0.300	0.000	0.000
	2	0.100	0.000	0.000	0.000	0.100	0.150

Then construct the conditional probability distribution of X=0 given Y= 1 and Y=2 given X =5.

Solution: $p(X=0/Y=1) = \frac{p(x=0,y=1)}{p_y(y=1)} = \frac{0.2}{0.55}$ = 4/11

Example 5.8: The joint density function for the random variables (X, Y), where X is the unit temperature change and Y is the proportion of spectrum shift that a certain atomic particle produces, is $f(x, y) = 10xy^2$, $0 < x < y < 1$, *and* 0, elsewhere,then

(a) construct the conditional probability distribution of Y given X.

(b) Find the probability that the spectrum shifts more than half of the total observations, given that the temperature is increased by 0.25 units.

Solution: (a) $f(y/x) = \frac{f(x,y)}{f_X(x)} = \frac{10xy^2}{\int_x^1 f(x,y)dy} = \frac{10xy^2}{\int_x^1 10xy^2 dy} = \frac{10xy^2}{\int_x^1 10xy^2 dy} = \frac{3y^2}{1-x^3}$, $0 < x < y < 1$

(b) $p(Y > ½ /x = ¼) = \frac{\int_0^1 \int_{1/2}^1 10xy^2 dydx}{f_X(x=1/4)} = \int_{1/2}^1 f\left(y / x = \frac{1}{4}\right) dy = 8/9.$

Example 5.9: Let X and Y be continuous random variables with joint density function

$$f(x, y) = \begin{cases} e^{-(x+y)} & x \geq 0, \quad y \geq 0 \\ 0 & otherwise \end{cases}$$

, then find the conditional density function of X given Y.

Solution: $f(x/y) = \frac{f(x,y)}{f_Y(y)} = \frac{e^{-(x+y)}}{\int_0^\infty f(x,y)dx} = \frac{e^{-(x+y)}}{e^{-y}} = e^{-x}$ $x \geq 0$.

ACTIVITY 5.4:

Given the joint probability density function of two continuous random variablesX and Y

$$f(x, y) = \begin{cases} \frac{2}{3}(x + 2y) & for \quad 0 < x < 1, \quad 0 < y < 1 \\ 0 & otherwise \end{cases}$$

(a) Find the conditional density function of X given Y.**[Ans. f(x/y) = $\frac{x+2y}{1+x}$]**

(b) Use the result in (a) to find$P(X = \frac{1}{2}|Y=\frac{1}{2})$**[Ans. 1]**

5.5Statistical Independent

If the conditional probability distribution of X given Y does not depend on *y*, then the joint probability distribution of X and Y is become the product of the marginal distributions of X and Y.It should make sense to the reader that if the conditional probability distribution of X given Y does not depend on *y*, then ofcourse the outcome of the random variable *Y* has no impact on the outcome of therandom variable *X*. In other words, we say that *X* and *Y* are independent randomvariables. We now offer the following formal definition of statistical independence.

Definition 5.5.1:

Let X and Y be two discrete random variables with joint probability mass function of p(x, y) and marginal distributions $p_x(x)$ and $p_y(y)$, respectively.The random variables X and Y are said to be **statistically independent** if andonly if p(x, y) = $f_x(x)f_y(y)$, for all (x, y) within their range.

Definition 5.5.2:

Let X and Y be two continuous random variables with joint probability density function$f(x, y)$ and marginal distributions $f_x(x)$ and $f_y(y)$, respectively.The random variables X and Y are said to be **statistically independent** if andonly if$f(x, y) = f_x(x)f_y(y)$, for all (x, y) within their range.

Note that, checking for statistical independence of discrete random variablesrequires a more thorough investigation, since it is possible to have the product ofthe marginal distributions equal to the joint probability distribution for some butnot all combinations of (x, y). If you can find any point (x, y) for which $p(x, y)$is defined such that $p(x, y) \neq p_x(x)p_y(y)$, the discrete variables X and Y are notstatistically independent.

Remark

- ✍ If we know the joint probability distribution of X and Y, we can find the marginal probability distributions, but if we have the marginal probability distributions, we may not have the joint probability distribution unless X and Y are statistically independent.

Theorem 5.1:

a. Let (X, Y) be a two dimensional discrete random variable. Then, X and Y are independent *if and only if*$P(x_i \mid y_j) = P_{xi}(x_i)$ for all i and j and $P(y_j \mid x_i) = P_{yj}(y_j)$ for all i and j.
b. Let (X, Y) be a two dimensional continuous random variable. Then, X and Y are independent *if and only if*$f(x \mid y) = f_x(x)$for all $(x, y)$and equivalently $f(y \mid x)= f_y(y)$ for all (x, y).

Examples 5.10:Let X and Y are binary random variables; that is 0 or 1 are the only possible outcomes for each of X and Y. p(0, 0) = 0.3; p(1, 1) = 0.2 and the marginal probability mass function of x = 0 and x= 1 are 0.6 and 0.4, respectively. Then

(a) construct the joint probability mass function of X and Y;

(b) calculate the marginal probability mass function of Y.

Solution: ***(a)*** ***(b)***

	X	***0***	***1***	***$P_y(y)$***
	0	***0.3***	***0.2***	***0.5***
Y	***1***	***0.3***	***0.2***	***0.5***
	$P_x(x)$	***0.6***	***0.4***	***1***

Example 5.11: Let X and Y are the life length of two electronic devices. Suppose that their joint $p.d.f$ is given by $f(x,y) = \begin{cases} e^{-(x+y)} & x \geq 0 \text{ and } y > 0 \\ 0 & elsewhere \end{cases}$, can these two random variables independent?

Solution:if X and Y are independent, then the product of their marginal distributions should equal to the joint pdf. So, $f_x(x) = e^{-x}$ $x \geq 0$ and $f_y(y) = e^{-y}$ $y \geq 0$.

Now $f(x, y) = f_x(x)\, f_y(y) = e^{-x}e^{-y} = e^{-(x+y)}$ $x \geq 0, y \geq 0$. Implies X and Y are statistically independent.

5.6 *n*–dimensional Random Variables

Instead of two random variables X and Y Sometimes we may be interested or concerned on three or more simultaneous numerical characteristics (random variables). Most of the concepts introduced for the two dimensional random variables can be extended to the n-dimensional one. Suppose that $(X_1, X_2, \ldots, X_n)$ may assume all values in some region on n-dimensional space. That is, the value is the n-dimensional vector $(X_1(s), X_2(s),\ldots, X_n(s))$, the probability density function of $(X_1, X_2, \ldots, X_n)$ is defined as follows.

There exists a joint probability mass function $p(x_1, x_2, \ldots, x_n)$for discrete random variables $X_1, X_2, \ldots, X_n$ satisfying the following conditions:

1. $f(x_1, x_2, \ldots, x_n) \geq 0$ for all $(x_1, x_2, \ldots, x_n)$
2. $\sum_{\infty} \ldots \sum_{\infty} f(x_1, x_2, \ldots, x_n) = 1$

Similarly for continuous random variables$X_1, X_2, \ldots, X_n$, the joint probability density function f of $(X_1, X_2, \ldots, X_n)$ satisfying the following conditions:

1. $f(x_1, x_2, \ldots, x_n) \geq 0$ for all $(x_1, x_2, \ldots, x_n)$
2. $f(x_1, x_2, \ldots, x_n) = \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} f(x_1, x_2, \ldots, x_n)\, dx_1 dx_2 \ldots dx_n = 1$

Remarks

All the preceding definitions concerning two random variables can be generalizedto the case of *n* random variables.

- ✍ If the joint probability mass function of discrete random variables $X_1, X_2, \ldots, X_n$ is $p(x_1, x_2, \ldots, x_n)$, then the marginal distribution of X_1 alone is given by: $p_{x1}(x_1) = \sum_{x_1} \cdots \sum_{x_n} p(x_1, x_2, \ldots, x_n)$ for all values with the range of X_1. More over the marginal distribution of X_1, X_2 and X_3 is given by: $p_{x1, x2, x3}(x_{1}, x_{2}, x_{3}) = \sum_{x_4} \cdots \sum_{x_n} p(x_4, x_5, \ldots, x_n)$

- ✍ If joint probability density function of continuous random variables $X_1, X_2, \ldots, X_n$ is $f(x_1, x_2, \ldots, x_n)$the marginal density of one of the random variable (Let X_3 alone) is given by: $f_{x3}(x_3) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, x_2, \ldots, x_n) dx_1 dx_2 dx_4 dx_5 \cdots dx_n$, for $-\infty \leq x_3 \leq \infty$.

  Moreover, the marginal density function of X_1, X_2 and X_n is given by:

  $\varphi(x_1, x_2, x_n) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, x_2, \ldots, x_n) dx_3 dx_4 dx_5 \cdots dx_{n-1}$ for $-\infty \leq x_1 \leq \infty$, for $-\infty \leq x_2 \leq \infty$, for $-\infty \leq x_n \leq \infty$ and so forth.

ACTIVITY 5.5:

If the probability density function of random variables X_1, X_2, X_3 is given by

$$f(x_1, x_2, x_3) = \begin{cases} (x_1 + x_2)e^{-x_3} & 0 < x_1 < 1, \ 0 < x_2 < 1, \ x_3 > 0 \\ 0 & otherwise \end{cases}$$

Find the marginal distribution of x_1 and x_2, x_2 and x_3, and x_1 and x_3?**[Ans. $(x_1 + x_2)$; $(1/2 + x_2)e^{x_3}$ and $(x_1 + ½)e^{x_3}$]**

- ✍ Let $f(x_1, x_2, \ldots, x_n)$ be the joint probability function of the random variables $X_1, X_2, \ldots, X_n$.We could consider numerous conditional distributions. For example, the **joint**

conditional distribution of X_1, X_2, and X_3, given that $X_4 = x_4, X_5 = x_5, \ldots, X_n = x_n$, is written: $f(x_1, x_2, x_3/ x_4, x_5, \ldots, x_n) = \frac{f(x_1, x_2, \ldots x_n)}{fx(x_4, x_5, \ldots x_n)}$, where $f(x_4, x_5, \ldots, x_n)$ is the joint marginal distribution of the random variables $X_4, X_5, \ldots, X_n$.

✍ Let $X_1, X_2, \ldots, X_n$ be n random variables, discrete or continuous, with joint probability distribution $f(x_1, x_2, \ldots, x_n)$ and marginal distribution $f_{X_1}(x_1), f_{X_2}(x_2), \ldots, f_{X_n}(x_n)$, respectively. The random variables $X_1, X_2, \ldots, X_n$ are said to be mutually **statistically independent** if and only if $f(x_1, x_2, \ldots, x_n) = f_{X_1}(x_1) f_{X_2}(x_2) \cdots f_{X_n}(x_n)$ for all $(x_1, x_2, \ldots, x_n)$ within their range.

SUMMARY

☞ If X and Y are both discrete random variables, then their *joint probability mass function* is defined by $p(x, y) = P\{X = x, Y = y\}$.

☞ The marginal mass functions are $P_x(x) = \sum_{\forall y} p(x, y)$ in the range of X and py(y) = $\sum_{\forall x} p(x, y)$ in the range of Y.

☞ The random variables X and Y are said to be *jointly continuous* if there is a pdf of(x, y), such that for any two-dimensional set C, $p\{(X, Y) \in C\} = \iint f(x, y)dxdy$.

☞ If X and Y are jointly continuous, then they are individually continuous with marginal density functions $f_X(x) = \int f(x, y)dy$ and $f_Y(y) = \int f(x, y)dx$ for X and Y respectively.

☞ The random variables X and Y are *independent* if, for all sets A and B, $P\{X \in A, Y \in B\} = P\{X \in A\}P\{Y \in B\}$.

☞ If X and Y are discrete random variables, then the *conditional probability mass function* of X given that $Y = y$ and Y given that X = x aredefined by $P(Y = y \mid X = x) = \frac{p(x, y)}{px(x)}$ and $P(X = x \mid Y = y) = \frac{p(x, y)}{py(y)}$, respectively. Similarly, $f(Y/X) = \frac{f(x, y)}{fx(x)}$ and $f(X/Y) = \frac{f(x, y)}{fy(y)}$, for joint continuous random variables.

CHECKLIST 5.1

Put a tick mark (√) for each of the following questions if you can solve the problems and an X otherwise. Can you

1. Joint probability distributions? ☐
2. Differentiate between discrete and continuous distributions? ☐

3. Drive the marginal probability distributions from the joint one? ☐

4. Drive the conditional probability distributions from the joint one? ☐

5. Identify independent probability distributions of two-dimensional probability distributions? ☐

EXERCISE 5.1:

1. Suppose (X, Y) be a joint vandom variables having pdf of f(x, y) =(x^2 + (xy)/2) , 0 < x < 1, 0 < y < 2, then find (a) p(X + Y < 1) (b) p(X > ½) (c) p(Y < X)
2. Suppose F(x, y) = ½ x^2y, 0 < x < 2; 0 < y < 1, then find (a) its pdf (b) p(Y < x/2).
3. Let f(x, y) be a joint pdf of f(x, y) = k e-(x+y), 0 < x, y < 1then calculate (a) the marginal probability density functions of X and Y (b) the conditional probability density function of Y given that X.
4. Statistical department gives an interview and exam to all employees. If X and Y are respectively the proportion of correct answers that an employee gets on the interview and exam, the joint pdf of X and Y can be approximated with f(x, y) = $\frac{2}{5}$(2x +3y), 0 < x, y < 1 and 0 elsewhere, then (a) find the marginal pdf of X and Y (b) find the conditional pdf of X and Y (c) p(Y < 0.5/X > 0.8) (d) are X and Y independent?

Chapter 6

Expectation

Introduction

The ideal situation in life would be to know with certainty what is going tohappen next. This being almost never the case, the element of chance entersin all aspects of our life. That is, we would like to know the *probability distribution* of *X*. In reallife, often, even this is not feasible. Instead, we are forced to settle for somenumerical characteristics of the distribution of *X*. This line of arguments leadsus to the concepts of the mathematical expectation and variance of a random variable. Therefore, ***expectation is nothing but the mean of a random variable.*** For example, consider a discrete random variable X with distinct values $x_1, x_2, x_3, \ldots, x_n$ having $p_1, p_2, p_3, \ldots, p_n$ corresponding probabilities, then the expected value of X is given by: $x_1 p_1 + x_2 p_2 + x_3 p_3 + \ldots + x_n p_n = \sum x_i p_i$.

Contents

a. Expectation of a random variable
b. Expectation of a function of a random variable
c. Properties of expectation
d. Variance of a random variable and its Properties
e. Chebyshev's Inequality
f. Covariance and Correlation Coefficient
g. Conditional Expectation

Learning Outcomes

At the end of this chapter students will be able to

- Evaluate the expectation and variance of random variables,
- Evaluate the expectation and variance of a bivariate random variables,
- Verify the covariance and correlation coefficient of random variables,

- Identify the conditional expectation of random variables.

RESOURCES:

- Statistics for Engineering and The Sciences, William M. & Terry S. (2007), 5th ed., pp. 211-217.
- Probability & Statistics for Engineers & Scientists, Sharon L. Myers et al. (2012), 9th ed., pp. 111-138
- A First Course in Probability, Sheldon Ross (2010). 8th ed., pp. 263-266 & pp. 297-349.

Consequently, we can describe process data with numerical descriptive measures, such as its mean and variance. Therefore, the expectation of X is very often called the mean of X and is denoted by E(X). The mean, or expectation, of the random variable X gives a single value that acts as a representative or average of the values of X, and for this reason it is often called a measure of central tendency.

Definition 6.1.1:

Let X be a discrete random variable which takes values x_i ($x_1, \ldots, x_n$) with corresponding probabilities $P(X = x_i) = p(x_i)$, $i = 1, \ldots, n$. Then the *expectation* of X (or *mathematical expectation* or *mean value* of X) is denoted by $E(X)$ and is defined as:

$$E(X) = x_1p(x_1) + \ldots + x_np(x_n) = \sum_{i=1}^{n} x_i\, p(x_i) = \sum_{x} x\, p(x)$$

The last summation is taken over all appropriate values of x.

If the random variable X takes on (countable) infinite many values x_i with corresponding probabilities $p(x_i)$, $i = 1, 2, \ldots$, then the expectation of X is defined by:

$$E(X) = \sum_{i=1}^{\infty} x_i\, p(x_i) \text{ provided that } \sum_{i=1}^{\infty} |x_i|\, p(x_{ii}) < \infty$$

Note that, as a special case where the probabilitiesare all equal, we have $E(x) = \frac{x_1 + x_2 + \ldots + x_n}{n}$ which is called the arithmetic mean, or simply the mean, of $x_1, x_2, \ldots, x_n$.

Example 6.1: A school class of 120 students is driven in 3 buses to a symphonic performance. There are 36 students in one of the buses, 40 in another, and 44 in the third bus. When the buses arrive, one of the 120 students is randomly chosen. Let X denote the number of students on the bus of that randomly chosen student, and find $E[X]$.

Solution: Since the randomly chosen student is equally likely to be any of the 120students, it follows that: $P\{X = 36\} = \frac{36}{120}$, $P\{X = 40\} = \frac{40}{120}$, $P\{X = 44\} = \frac{44}{120}$.

Hence $E(X) = 36\text{x}\frac{3}{10} + 40\text{x}\frac{1}{3} + 44\text{x}\frac{11}{30} = \frac{1208}{30} = 40.2667$.

Example 6.2: Let a fair die be rolled once. Find the mean number rolled, say X.

Solution: Since S = { 1, 2, 3, 4, 5, 6} and all are equally likely with prob. of 1/6, we have

$$E(X) = 1.\frac{1}{6} + 2.\frac{1}{6} + 3.\frac{1}{6} + 4.\frac{1}{6} + 5.\frac{1}{6} + 6.\frac{1}{6} = \frac{21}{6} = 3.5.$$

Example 6.3:A lot of 12 TV sets includes two which are defectives. If two of the sets are chosen at random, find the expected number of defective sets.

Solution: Let X= the number of defective sets.

Then, the possible values of X are 0, 1, 2. Using conditional probability rule, we get,

$P(X = 0) =$ P (both non defective) = $\frac{10}{12} \times \frac{9}{11} = \frac{15}{22}$, $P(X = 2) =$ P (both defective) = $\frac{2}{12} \times \frac{1}{11} = \frac{1}{66}$,

$P(X = 1) =$ P(one defective)

= P (first defective and second good) + P (first good and second defective)

$$= \frac{2}{12} \times \frac{10}{11} + \frac{10}{12} \times \frac{2}{11} = \frac{10}{66} + \frac{10}{66} = \frac{10}{33}.$$

Or, since $P(X = 0) + P(X = 1) + P(X = 2) = 1$, we can use, $P(X = 1) = 1 - \frac{45}{66} - \frac{1}{66} = \frac{10}{33}$.

$$\therefore \; E(X) = \sum_{i=0}^{2} x_i P(X = x_i) = 0 \times \frac{15}{22} + 1 \times \frac{10}{33} + 2 \times \frac{1}{66} = \frac{1}{3}.$$

Example 6.4: Let X be a random variable with the pmf of:

X	0	1	2	3	4
P(x)	0.2	0.1	0.4	0.2	0.1

Then find its expected value.

Solution: E(X) = $\sum xp(x)$ = 0x0.2 + 1x0.1 + 2x0.4 + 3x 0.2 + 4x0.1 = 1.9

Example 6.5: Find the expectation of a discrete random variable X whose probability function is given by $p(x) = \left(\frac{1}{2}\right)^x$, for x = 1, 2, 3, . . .

Solution:E(X) = $\sum xp(x)$ = ½ + (1/2)^2 + (1/2)^3 +…= ½ (1 + ½ + (1/2)2 +…) = (½) /(1- ½) = 1

Dear students, the above three examples are designed to allow the reader to gain some insightinto what we mean by the expected value of a random variable. In all cases therandom variables are discrete. We follow with an example involving a continuous random variable.

Definition 6.1.2:

The mathematical expectations, in general, of a continuous r-v are defined in a similar way with those of a discrete r-v with the exception that summations have to be replaced by integrations on specified domains. Let the random variable*X* is continuous with *p.d.f. f(x),* its expectation is defined by: $E(X) = \int_{-\infty}^{\infty} x\, f(x)\, dx$, provided this integral exists.

Example 6.6: The density function of a random variable X is given by:

$$f(x) = \begin{cases} \frac{1}{2}x & 0 < x < 2 \\ 0 & otherwise \end{cases}$$

Then, find the expected value of X?

Solution: E(X) = $\int_0^2 xf(x)dx = \int_0^2 \frac{1}{2}x^2 dx$ = [1/6 $x^3]_0^2$ = 4/3.

Example 6.7: Find the expected value of the random variable X with the CDF of F(x) = x^3,0 < x< 1.

Solution:E(X) = $\int_0^1 xf(x)dx = \int_0^1 x^4 dx = \frac{1}{5}[x^5]_0^1 = \frac{1}{5}$.

Example 6.8: Let X has the following pdf f(x) = $\frac{2x}{c^2}$, 0 < x < c and you are told that E(X)=6, then find c.

Solution:E(X) = 6 = $\int_0^c xf(x)dx = \frac{2}{c^2}\int_0^c x^2 dx = \frac{2}{3c^2}[x^3]_0^c = \frac{2c^3}{3c^2}$ then c = 9.

ACTIVITY 6.1:

Suppose that X is a continuous random variable with pdf of f(x) = $\begin{cases} 1 + x, -1 \le x < 0 \\ 1 - x, \;\; 0 \le x \le 1 \end{cases}$, then find E(X).**[Ans. E(X) = 0]**

6.2 Expectation of a Function of a Random Variable

The Statistics that we will subsequently use for making inferences are computed from the data contained in a sample. The sample measurements can be viewed as observations on n random samples,x_1, x_2, x_3, …, x_n. Since the sample Statistics are functions of the random variables x_1, x_2, x_3, …, x_n, they also will be random variables and will possess probability distributions. To describes these distributions, we will define the expected value (or mean) of functions of random variables.

Definition 6.2:

Now let us consider a new random variable *g*(*X*), which depends on *X*; that is, each value of *g*(*X*) is determined by the value of *X*. In particular, let X be a discrete random variable with probability function p(x). Then Y = g(X) is also a discrete random variable, and the probability function of Y isp(y) = P(Y = y) = $\sum_{\{x|\, g(x)=y\}} P(X=x) = \sum_{\{x|\, g(x)=y\}} f(x)$ and hence we can define expectation of functions of random variables as:

Let X be a random variable and let Y = g(X), then

(a) If X is a discrete random variable and $p(x_i) = P(X=x_i)$ is the *p.m.f,* we will have

$$E(Y) = E(g(X)) = \sum_{i=1}^{\infty} g(x_i)\, p(x_i)$$

(b) If X is a continuous random variable with *p.d.f, f(x),* we will have

$$E(Y) = E(g(X)) = \int_{-\infty}^{\infty} g(x) f(x)\, dx$$

The reader should note that the way to calculate the expected value, or mean,shown here is different from the way to calculate the sample mean described in Introduction to Statistics, where the sample mean is obtained by using data. Here is in random variable, the expected value is calculated by using the probability distribution. However, the mean is usually understood as a “center” value of the underlyingdistribution if we use the expected value.

Example 6.9: Suppose that a balanced die is rolled once. If X is the number that shows up,find the expected value of $g(X) = 2X^2 + 1$.

Solution: Since each possible outcome has the probability 1/6, we get,

$$E(g(X)) = \sum_{x=1}^{6} (2x^2+1).\frac{1}{6} = (2\times 1^2+1).\frac{1}{6} + \ldots + (2\times 6^2+1)\frac{1}{6} = \frac{94}{3}.$$

The determination of expectations is often simplified using some properties.

ACTIVITY 6.2:

1. If the random variable X is the top of a rolled die. Find the expected value of g(x) = $2X^2+1$?**[Ans. 150.5]**

2. Let *X* be a random variable with p.d.f. $f(x) = 3x^2$, $0 < x < 1$. Thencalculate the *E(Y)*:*(a)* If $Y = 3X - 2$. *(b)* If $Y = 3x^2 - 2x$. **[Ans. (a) E(Y) = ¼ (b) E(Y) = 0.3]**

6.3 Expectation of Two Dimensional Random Variables

Generalizations can be easily made to functions of two or more random variables from the definition of expectation for one dimensional random variable.

Definition 6.3:

Let *X* and *Y* be random variables with joint probability distribution *p*(*x*, *y*) [or f(x, y)] and let H = g(x, y) be a real valued function of (X, Y), then themean, or expected value, of the random variable (x,Y) and *g*(*X*, *Y*) are:

- ☞ $E(XY) = \sum_x \sum_y xyp(x,y)$ if X and Y are discrete random variables.
- ☞ $E[XY] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf(x,y)\,dx\,dy$ if X and Y are continuous random variables.
- ☞ $E[g(X,Y)] = \sum_x \sum_y g(x,y)p(x,y)$ if X and Y are discrete random variables.
- ☞ $E[g(X,Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)f(x,y)\,dx\,dy$ if X and Y are continuous random variables.

Example 6.10: If the joint probability density function of X and Y given by

$$f(x) = \begin{cases} \frac{2}{7}(x+2y) & 0 < x < 1, 1 < y < 2 \\ 0 & otherwise \end{cases}$$

Then find the expected value of g(X, Y) = X/Y?

Solution: $E\{g(x,y)\} = \int\int g(x,y)f(x,y)dxdy = \frac{2}{7}\int_1^2\int_0^1 \frac{x}{y}(x+2y)dxdy = \frac{2}{7}\int_1^2\int_0^1\{\frac{x^2}{y} + 2x\}dxdy$

$$= \frac{2}{7}\int_1^2\int_0^1\{\frac{x^3}{3y} + x^2\}dy = \frac{2}{7}\int_1^2\int_0^1\{\frac{1}{3y} + 1\}dy = \frac{2}{7}\{1/3\,(\ln 2 - \ln 1) + 1\}$$

= 0.351728.

ACTIVITY 6.3:

Suppose that X and Y have the jointprobability mass function as: p(x, y) = 0.1, 0.15, 0.2, 0.3, 0.1 and 0.15 for (x = 2, 4; y = 1, 3, 5), then find E {g(x, y)} if g(x, y) = xy^2. **[Ans. 35.2]**

Remark

✍ In calculating $E(X)$ over a two-dimensional space, one may use either the joint probability distribution of X and Y or the marginal distribution of X *as:*

$E[X] = \sum_x \sum_y xp(x,y) = \sum_x x\, p_x(x)$ if X is discrete random variable.

$E[X] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xf(x,y)\,dx\,dy = \int_{-\infty}^{\infty} xp_x(x)\,dx$ if X is continuous random variable, where

$p_x(x)$ is the marginal distribution of X. Similarly, we define

$E[Y] = \sum_x \sum_y yp(x,y) = \sum_y y\, p_y(y)$ if Yis discrete random variable.

$E[Y] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} yf(x,y)\,dx\,dy = \int_{-\infty}^{\infty} yp_y(y)\,dy$ if Y is continuous random variable,where

$p_y(y)$ is the marginal distribution of the random variable Y.

6.4Variance of a Random Variable

The expectation by itself is not an adequate measure of description of a distribution,and an additional measure is needed to be associated with the spreadof a distribution. Such a measure exists and is the variance of a random variable or of itsdistribution. Therefore, the variance (or the standard deviation) is a measure of the dispersion, or scatter, of the values of the randomvariable about the mean.

Definition 6.4:

Let X is a random variable. The variance of X, denoted by V(X) or Var(X) or δ_x^2, defined as:

$$V(X) = E\,(X - E(X))^2 = V(X) = E(X^2) - [E(X)]^2 \quad where \quad E(X^2) = \sum x_i^{\,2} f(x_i)$$

Note that, the positive square root of V(X) is called the standard deviation of X and denoted by σ_x.Unlike the variance, the standard deviation is measured in the same units as X (and $E(X)$) and serves as a yardstick of measuring deviations of $X$from $E(X)$.

Examples 6.11: Find the expected value and the variance of the r-v given in Example 3.2

Solution: $E(X) = \int_{-\infty}^{\infty} x.f(x)dx = \int_0^1 x.xdx + \int_1^2 x.(2-x)dx = \int_0^1 x^2dx + \int_1^2 (2x - x^2)dx$

$$= \left.\frac{x^3}{3}\right|_0^1 + \left(x^2 - \frac{x^3}{3}\right)\Bigg|_1^2 = \frac{1}{3} + \left(4 - \frac{8}{3}\right) - \left(1 - \frac{1}{3}\right) = \frac{1}{3} + \frac{4}{3} - \frac{2}{3} = 1.$$

$$E(X^2) = \int_{-\infty}^{\infty} x^2 . f(x)dx = \int_0^1 x^2 .xdx + \int_1^2 x^2(2-x)dx = \int_0^1 x^3 dx + \int_1^2 (2x^2 - x^3)dx$$

$$= \left.\frac{x^4}{4}\right|_0^1 + \left(\frac{2}{3}x^3 - \frac{x^4}{4}\right)\Bigg|_1^2 = \frac{1}{4} + \left(\frac{16}{3} - 4\right) - \left(\frac{2}{3} - \frac{1}{4}\right) = \frac{1}{4} + \frac{4}{3} - \frac{5}{12} = \frac{7}{6}.$$

$$\therefore V(X) = E(X^2) - [E(X)]^2 = \frac{7}{6} - 1^2 = \frac{1}{6}.$$

Remark

- ✍ More generally, for a random variable *X* taking on finitely many values $x_1, \ldots, x_n$ with respective probabilities $p(x_1), \ldots, p(x_n)$, the variance is: $Var(X) = \sum_{i=1}^{n} (x_i - E(X)^2 \, p(x_i)$ and represents the sum of the weighted squared distances of the points *xi*, $i = 1, \ldots, n$ from the center of location of the distribution, *E(X)*. Thus, the further from *E(X)* the *xi*'s are located, the larger the variance, and vice versa. The same interpretation holds for the case that *X* takes on (countable) infinite many values or is of the continuous type. Because of this characteristic property of the variance, the variance is referred to as a measure of *dispersion* of the underlying distribution.
- ✍ Let *X* and *Y* be random variables with joint probability distribution *p*(*x, y*) [or f(x, y)] and let H = g(x, y) be a real valued function of (X, Y), then the variance of the random variable (X,Y) and *g*(*X, Y*) are $V(XY) = E(X^2Y^2) - [E(XY)]^2$ and $E[\{g(X, Y)\}^2] - [E\{g(X, Y)\}]^2$, respectively.

6.4. Properties of Expectation and Variance

There are cases where our interest may not only be on the expected value of a r -v, but also on the expected value of a r -v related to X. In general, such relations are useful to explain the properties of the mean and the variance.

Property 1: If a is constant, then $E(aX) = aE(X)$.

Property 2: If b is constant, then $E(b) = b$.

Property 3: If a and b are constants, then $E(aX + b) = aE(X) + b$.

Property 4: Let X and Y are any two random variables. Then E(X + Y) = E(X) + E(Y). This can be generalized to *n* random variables, That is, if $X_1, X_2, X_3, \ldots, X_n$ are random variables then, $E(X_1 + X_2 + X_3 + \ldots + X_n) = E(X_1) + E(X_2) + E(X_3) + \ldots + E(X_n)$

Property 5: Let X and Y are any two random variables. If X and Y are independent. Then E(XY) = E(X)E(Y)

Property 5: Let (X, Y) is a two dimensional random variable with a joint probability distribution. Let $Z = H_1(X, Y)$ and $W = H_2(X, Y)$. Then E(Z + W) = E(Z) + E(W).

The proof is straight forward, using properties of summation and considers property 3 for this one:

$$E(aX+b) = \sum (ax+b).f(x) = \sum ax.f(x) + \sum b.f(x)$$

$$= a\sum x.f(x) + b\sum f(x) = aE(X) + b \text{, since} \sum f(x) = 1.$$

Now, if we let b = 0 or a = 0, from property 3, we get the following corollaries:

Property 1: For constant values a and b, $V(aX+b) = a^2V(X)$.

Proof: If Y= aX +b, then $E(Y) = aE(X) + b$ and $Y - E(Y) = a[X - E(X)]$.
Square and take expectations of both sides.

$E[Y - E(Y)]^2 = a^2E[X - E(X)]^2 \Rightarrow V(Y) = a^2V(X)$, i.e., $V(aX+b) = a^2V(X)$.

The following three properties are corollaries where b = 0, a = 1 or a = 0, respectively.

Property 2: Variance is not independent of change of scale, i.e. $V(aX) = a^2V(X)$.

Property 3: Variance is independent of change of origin, i.e., $V(X+b) = V(X)$.

Property 4: Variance of a constant is zero, i.e., $V(b) = 0$.

Property 5: Let $X_1, X_2, X_3, \ldots, X_n$ be *n* independent random variable, then $V(X_1 + X_2 + X_3 + \ldots + X_n) = V(X_1) + V(X_2) + V(X_3) + \ldots + V(X_n)$

Property 6: If (X, Y) be a two dimensional random variable, and if X and Y are independent then V(X + Y) = V(X) + V(Y) and V(X - Y) = V(X) + V(Y)

Examples 6.12: A continuous random variable X has probability density given by

$$f(x) = \begin{cases} 2e^{-2x} & x > 0 \\ 0 & x \le 0 \end{cases}$$ andfor a constant K.Find

(a) The variance of X (b) The standard deviation of X (c) Var(KX) (d) Var(K + X)

Solution: (a) V(X) = $\int x^2 f(x)dx$ - $[E(X)]^2$ = $\int 2x^2 e^{-2x}dx$ - $[\int 2xe^{-2x}dx]^2$ = $2(1/2)^2 - (1/2)^2$

= ¼ (b) SD(V(X)) = $\sqrt{1/4}$ = ½ (c) V(KX) = K^2V(X) = $\frac{K^2}{4}$ (d) V(K + X) = V(X) =1/4.

Example 6.13: Let X be a random variable with p.d.f. $f(x) = 3x^2$, for $0 < x < 1$.

(a) Calculate the Var (X).*(b)* If the random variable Y is defined by Y = 3X − 2, calculate the Var(Y).

Solution: (a) V(X) = $\int x^2 f(x)dx$ - $[E(X)]^2$ = $\int 3x^4 dx$ - $[\int 3x^3 dx]^2$ = $3/5 - [3/4]^2 = 3/80$

(b) V(3X – 2) = 9V(X) = 9x3/80 = 27/80.

6.5 Chebyshev's Inequality

In Section 6.4 we stated that the variance of a random variable tells us somethingabout the variability of the observations about the mean. If a random variablehas a small variance or standard deviation, we would expect most of the values tobe grouped around the mean. Therefore, the probability that the random variable assumes a value within a certain interval about the mean is greater than for asimilar random variable with a larger standard deviation. If we think of probabilityin terms of area, we would expect a continuous distribution with a large value ofσ to indicate a greater variability, and therefore we should expect the area to be more spread out. A distribution with a small standard deviation should have most of its area close to μ.We can argue the same way for a discrete distribution.

The Russian mathematician P. L. Chebyshev (1821–1894) discovered that thefraction of the area between any two values symmetric about the mean is relatedto the standard deviation. Since the area under a probability distribution curveor in a probability histogram adds to 1, the area between any two numbers is theprobability of the random variable assuming a value between these numbers.The following theorem, due to Chebyshev, gives a conservative estimate of theprobability that a random variable assumes a value within k standard deviationsof its mean for any real number k.

Theorem 6.1: (Chebyshev's Inequality)

Let X be random variable with E(X) = μ and variance $\sigma^2$and let k be any positive constant. Then the probability that any random variable X will assumea value within k standard deviations of the mean is at least $1 - \frac{1}{k^2}$. Thatis,$P(\mu - k\sigma < X < \mu + k\sigma) \geq 1 - \frac{1}{k^2}$.

Note that, Chebyshev's theorem holds for any distribution of observations, and for thisreason the results are usually weak. The value given by the theorem is a lowerbound only. That is, we know that the probability of a random variable fallingwithin two standard deviations of the mean can be *no less* than 3/4, but we neverknow how much more it might actually be. Only when the probability distributionis known can we determine exact probabilities. For this reason we call the theorema *distribution-free* result. The use of Chebyshev's theorem isrelegated to situations where the form of the distribution is unknown.

Examples 6.14: A random variable X has a mean μ = 8, a variance σ^2= 9, and an unknown probability distribution. Find

(a) $P(-4 < X < 20)$,

(b) $P(|X - 8| \geq 6)$.

Solution: (a) p(-4 < X < 20) = p{(-4-8)/3 < Z < (20-8)/3} = p(-4 < Z < 4) = 1

(b) p({|X-8|≥6} = p(14 < X or X > 14) = p(14 < X) + p(X > 14)

= p(-1.33 < Z) + p(Z > 2) = 0.5- p(0 < Z 1.33) + 0.5 – p(0 < Z <2)

= 1 – (0.4082 + 0.4772) = 0.1146

6.6 Covariance and Correlation Coefficient

6.6.1 Covariance

The covariance between two random variables is a measure of the nature of theassociation between the two. If large values of X often result in large values of Yor small values of X result in small values of Y , positive $X-\mu_X$will often result inpositive $Y-\mu_Y$and negative $X-\mu X$ will often result in negative $Y-\mu Y$. Thus, theproduct $(X-\mu X)(Y-\mu Y)$ will tend to be positive. On the other hand, if large Xvalues often result in small Y values, the product $(X-\mu X)(Y-\mu Y)$ will tend to benegative. The *sign* of the covariance indicates whether the relationship between twodependent random variables is positive or negative.

Definition 6.6:

The ***covariance***of two random variables X and Y is denoted by Cov(X, Y), is defined by

$$\text{Cov}(X, Y) = \sigma_{xy} = E[(X - E(X))(Y - E(Y)] = E(XY) - (EX)(EY)$$

N.B.: When X and Y are statistically independent, it can be shown that the covariance is zero.

$$\text{Cov}(X, Y) = E[(X - E(X))(Y - E(Y)] = E[(X - E(X))]E[(Y - E(Y)] = 0.$$

Thus if X and Y are independent, they are also uncorrelated. However, the reverse is not true as illustrated by the following example.

Examples 6.15: The pair of random variables (X, Y) takes the values (1, 0), (0, 1), (−1, 0), and (0,−1), each with probability ¼.

Solution: The marginal *p.m.f.`s* of X and Y are symmetric around 0, &**E**[X] = **E**[Y] = 0. Furthermore, for all possible value pairs of (x, y), either x or y is equal to 0, which implies that $XY = 0$ and **E**[XY] = 0. Therefore,

Cov(X, Y) = E[(X – E(X)(Y – E(Y)] = 0

6.6.2 Properties of Covariance

Property 1: Cov(X, Y) = Cov (Y, X)

Property 2: Cov (X, X) = Var(X)

Property 3: Cov(KX, Y) = K Cov(X, Y) for a constant K.

Property 4:
$$Cov\left(\sum_{i=1}^{n} X_i, \sum_{j+1}^{m} Y_j\right) = \sum_{i=1}^{n}\sum_{j=1}^{m} Cov(X_i, Y_j)$$

Property 5: Var (X ± Y) = Var (X) + Var (Y) ± 2 Cov (X, Y)

6.6.3 Correlation Coefficient

Although the covariance between two random variables does provide informationregarding the nature of the relationship, the magnitude of σ_{XY} *does not indicateanything regarding the strength of the relationship*, since $\sigma_{XY}$is not scale-free.Its magnitude will depend on the units used to measure both X and Y. There is ascale-free version of the covariance called the **correlation coefficient** that is usedwidely in statistics.

Definition 6.7

Let X and Y be random variables with covariance *Cov(X, Y)*and standard deviations $\sigma_X$and σ_Y, respectively. The correlation coefficient (or coefficient of correlation)ρ of two random variables X and Y that have none zero variances is defined as:

$$\rho = \rho_{xy} = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}} = \frac{E(XY) - E(X)E(Y)}{\sqrt{Var(X)Var(Y)}} = \frac{\sigma_{xy}}{\sigma_x\sigma_y} = \frac{E\{[X - E(X)][Y - E(Y)]\}}{\sqrt{V(X)V(Y)}}.$$

It should be clear to the reader that $\rho_{XY}$is free of the units of X and Y. Thecorrelation coefficient satisfies the inequality $-1 \leq \rho_{XY} \leq 1$ and it assumes a value ofzero when $\sigma XY = 0$.

Examples 6.16: Let X and Y be random variables having joint probability density function

$$f(x,y) = \begin{cases} x+y & 0 \leq x \leq 1, 0 \leq y \leq 1 \\ 0 & elsewhere \end{cases}$$ then findCorr(X,Y)

Solution: Corr(X, Y) = $\dfrac{\text{E(XY) - E(X)E(Y)}}{\sqrt{Var(X)Var(Y)}}$ =((1/3) – (7/12)(7/12)}/ (264/3456) = 8.636364.

Remark

✍ If ρ_{xy} is the correlation coefficient between X and Y, and if V=AX +B and W = CY +D, where A, B, C and D are constants, then$\rho_{VW} = \dfrac{AC}{|AB|}\rho_{xy}$, where A and Bare different from 0.

6.7 Conditional Expectation

Recall that if X and Y are jointly discrete random variables, then the conditionalprobability distribution of X, given that $Y = y$, and Y given that X = x is defined, for all $y$such that$P\{Y = y\}$ >0, and for all x such that p(X =x) > 0 by: $p(x|y) = P\{X = x|Y = y\} = \frac{p(x,y)}{p_y(y)}$ and

$p(y/x) = P\{Y = y|X = x\} = \frac{p(x,y)}{p_x(x)}$.

Similarly, let us recall that if X and Y are jointly continuous with a joint probabilitydensity function $f(x, y)$, then the conditional probability density of X, given that$Y = y$, and Y given that X = x is defined, for all values of y *and* x such that $f_Y(y) > 0$, by $f(y/x) = \frac{f(x,y)}{f_x(x)}$and $f_X(x) > 0$, by $f(x|y) = \frac{f(x,y)}{f_y(y)}$.

Once a conditional distribution is at hand, an expectation can be defined as done inrelations to expectation of one dimensional random variable. However, a modified notation will beneeded to reveal the fact that the expectation is calculated with respect to a*conditional* pdf. The resulting expectation is the *conditional* expectation of one random variable, given the other random variable, as specified below.

Definition 6.7:

If X and Y have joint probability mass function $p(x, y)$, then the conditional expectation of X giventhat $Y = y$, and Y given that X =x for all values of y and x are:

$E(Y|X = x) = \sum_y yp(y/x)$ and $E(X|Y = y) = \sum_x xp(x/y)$.

If X and Y have joint probability density function $f(x, y)$, then the conditional expectation of X giventhat $Y = y$, and Y given that X =x for all values of y and x are:

E(Y| X = x) = $\int_{-\infty}^{\infty} y\, p(y|\, x)\, dy$ and E(X| Y= y) = $\int_{-\infty}^{\infty} xp(x|\, y)\, dx$

Property 1:E[E(X | Y)] = E(X) and E[E(Y | X)] = E(Y)

Property 2: Let X and Y be independent random variables. Then E(X | Y) = E(X) and E(Y | X) = E(Y)

Examples 6.17: Let X be Random variable taking on the values -2, -1, 1, 2 each with probability ¼ and define the random variable Y = X^2. Then find the conditional expectation of Y given X and X given Y?

Solution:

X	-2	-1	1	2	Y	1	4
P(x)	¼	1/4	1/4	1/4	P(y)	1/2	1/2

E(Y/X) = $\sum_{\forall y}$ p(y / x) = ½ *1 + ½ *4 = 5/2

E(X/Y) = $\sum_{\forall x}$ p(x / y) = $\pm 1 * 1/4 + \pm 2 * 1/4 = 0$

OR E{E(Y / X)} = E(Y) = ½ (1+4) = 3/2 and E{E(X / Y)} = E(X) = ¼ (1+2-1-2) = 0.

Example 6.23: Let X and Y be random variables having joint probability density function

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

Then find the conditional expectation of (a) Y given X= 0.5 (b) X given Y =0.25?

Solution: first: fx(x) = $\int f(x, y) dy$ = x + ½ and fy(y) = $\int f(x, y) dx$ = ½ + y

f(y / x) $= \frac{f(x,y)}{fx(x)} = \frac{x+y}{x+\frac{1}{2}}$ and f(x /y) = $\frac{x+y}{y+\frac{1}{2}}$

(a) E(Y / X) = $\int y f(\, y \,/\, x \; = 0.5) dy = \int y \, \left(\frac{1}{2} + y\right) dy$ = 7/12

(b) E(X / Y) = $\int x f(\, x \,/\, y \; = 0.25) dx = \int x \; 4/3 \left(\frac{1}{4} + x\right) dx$ = 11/18

Remark

✍ Just as we have defined the conditional expectation of X given the value of Y, we can also define the conditional variance of X given that $Y = y$: $\mathrm{Var}(X|Y) = E[(X - E[X|Y])^2]$. That is, $\mathrm{Var}(X|Y)$ is equal to the (conditional) expected square of the difference between X and its (conditional) mean when the value of Y is given. In other words, Var$(X|Y)$ is exactly analogous to the usual definition of variance, but now all expectations are conditional on the fact that Y is known.

✍ There is a very useful relationship between Var(X), the unconditional variance of X, and Var$(X|Y)$, the conditional variance of X given Y, that can often be applied to compute Var(X). To obtain this relationship, note first that, by the same reasoning that yields Var$(X) = E[X^2] - (E[X])^2$, we have Var$(X|Y) = E[X^2|Y] - (E[X|Y])^2$ so $E[\text{Var}(X|Y)] = E[E[X^2|Y]] - E[(E[X|Y])^2] = E[X^2] - E[(E[X|Y])^2]$.

SUMMARY

- If X is a discrete random variable, then the quantity $E[X] = \sum_{\forall x} xp(x)$ is called the *expected value* of X and a useful identity states that, for a function g, $E[g(X)] = \sum_{\forall g(x)} g(x)p(x)$ is expectation of functions of random variable.
- The expected value of a continuous random variable X is defined by $E[X] = \int_{-\infty}^{\infty} xf(x)dx$and a useful identity is that, for any function g, $E[g(X)] = \int_{-\infty}^{\infty} g(x)f(x)dx$.
- If X and Y have a joint probability mass function $p(x, y)$, then $E[X,Y] = \sum_{\forall x}\sum_{\forall y} xyp(x,y)$ and useful identity is that, for any function g, $E[g(X, Y)] = \sum_{\forall g(x,y0x} g(x,y)\, p(x,y)$. *If X and Y are a joint probability density function, then summation is replaced by integration.*
- The *variance* of a random variable X, denoted by Var(X), is defined by Var$(X) = E[X^2] - (E[X])^2$.
- The *covariance* between joint random variables X and Y is given by Cov$(X,Y)=E[XY]-E[X]E[Y]$.
- The correlation between joint random variables X and Y is defined by: Corr(X, Y) = ρ_{XY}

$$\frac{\text{E(XY) - E(X)E(Y)}}{\sqrt{Var(X)Var(Y)}}$$

- If X and Y are jointly discrete random variables, then the conditional expected value of X, given that $Y = y$, is defined by $E[X|Y = y] = \sum_{\forall x} xp(X = x\,/Y = y)$ and if X and Y are jointly continuous random variables, then $E[X|Y = y] = \int_{-\infty}^{\infty} xf(x\,/y)dx$

EXERCISE 6.1:

1. The joint probability density function of two random variables *X* and *Y* is given by

$$f(x, y) = \begin{cases} c(2x + y) & for\ 2 \leq x \leq 6,\ 0 \leq y \leq 5 \\ 0 & \text{otherwise} \end{cases}$$ For *c*= 1/210. Find

(a) *E*(*X*), (b) *E*(*Y*), (c) *E*(*XY*), (d) Cov(*X*, *Y*) (e) Cov(X $\pm$ Y) (f) Corr(X,Y)

2. The joint probability function of two discrete random variables *X* and *Y* is given by *f*(*x*, *y*) = *c*(2*x* + *y*), where *x* and *y* can assume all integers such that $0 \leq x \leq 2$, $0 \leq y \leq 3$, and *f* (*x*, *y*) = 0 otherwise. Then find

(a) *E*(*X*), (b) *E*(*Y*), (c) *E*(*XY*), (d) Cov(*X*, *Y*) (e) Cov(X $\pm$ Y) (f) Corr(X,Y)

3. The correlation coefficient of two random variables X and Y is -1/4 while their variances are 3 and 5. Find the covariance?

4.Let X and Y be continuous random variables with joint density function

$$f(x, y) = \begin{cases} e^{-(x+y)} & x \geq 0,\ \ y \geq 0 \\ 0 & otherwise \end{cases}$$, then find

(a) Var(X), *(b)* Var (Y), (c) σ_x, (d) σ_y, (e) σ_{xy}(f) ρ

CHAPTER 7

THE MOMENT AND MOMENT -GENERATING FUNCTIONS

Introduction

In this chapter, we concentrate on applications of moment-generating functions.The obvious purpose of the moment-generating function is in determining momentsof random variables. However, the most important contribution is to establishdistributions of functions of random variables.

Contents

- Definition of Moment and Moment-Generating functions
- Examples of moment generating functions
- Theorems on moment generating functions

Learning Outcomes

At the end of this chapter students will be able to:

- ✓ Understand the moment about the origin and the mean,
- ✓ Define the moment generating function,
- ✓ Apply the moment generating function for different distributions.

RESOURCES:

- 🕮 Statistics for Engineering and The Sciences, William M. & Terry S. (2007), 5th ed., pp. 157-160 & 200 - 215.
- 🕮 Probability & Statistics for Engineers & Scientists, Sharon L. Myers et al. (2012), 9th ed., pp. 111-138
- 🕮 A First Course in Probability, Sheldon Ross (2010). 8th ed., pp. 354-366.

7.1. Moment

If $g(X) = X^r$ for $r = 0, 1, 2, 3, \ldots$, the following definition yields an expected value called the r^{th} **moment about the origin** of the random variable X, which we denote by μ'_r.

Definition 7.1:

The r^{th} **moment about the origin** of the random variable X is given by:

$$\mu'_r = E(X^r) = \begin{cases} \sum X^r\, p(x), & if\ X\ is\ discrete \\ \int_{-\infty}^{\infty} X^r\, f(x)dx, & if\ X\ is\ continuous \end{cases}$$

Remark

- The r^{th} moment of a random variable X about the mean μ also called the k^{th} central moment, is defined as: $\mu_r = E[(X - \mu)^r]$ that is;

  $$\mu_r = \sum_x (x-\mu)^r p(x)$$, if X is discrete random variable

  $\mu_r = \int_{-\infty}^{\infty} (x - \mu)^r f(x)dx$, if X is continuous random variable, where r= 0, 1, 2, ... Since the first and second moments about the origin are given by $\mu'_1 = E(X)$ and $\mu'_2 = E(X^2)$, we can write the mean and variance of a random variable as $\mu = \mu'_1$ and $\sigma^2 = \mu'_2 - \mu'_1$.

Example 7.1: The random variable X can assume values 1 and -1 with probability ½ each. Then find the first and second moments of X about the origin and the mean and hence find the variance of X based on moment.

Solution: $\mu'_r = E(X^r) = \sum X^r\, p(x)$, then $\mu'_1 = E(X) = \sum x\, p(x) = (1\text{-}1)*1/2 = 0$

$\mu'_2 = E(X) = \sum X^2\, p(x) = (1+1)*1/2 = 1$

$\sigma^2 = \mu'_2 - \mu'_1 = 1 - 0 = 1$

N.B.: moment about the origin and the mean are the same as expected value of this random variable is zero.

7.2 The Moment -Generating Functions

Although the moments of a random variable can be determined directly from definition 7.1 an alternative procedure exists. This procedure requires us to utilize a **moment-generating function**.

Definition 7.2:

The moment generating function *M(t)* of the random variable *X* is defined for all realvalues of *t* by:

$M_x(t)=E(e^{tX})= \begin{cases} \sum e^{tX}\ p(x),\ if\ X\ is\ discrete\ with\ pmf\ of\ p(x) \\ \int_{-\infty}^{\infty} e^{tX}\ f(x)dx, if\ X\ is\ continuous\ with\ pdf\ of\ f(x) \end{cases}$

Property 1: M'(0) = E(X).

Property 2: M''(0) = $E(X^2)$.

Property 3: In general, the n^{th} derivative of M(t) is given by: $M^n(t) = E(X^n e^{tX})$ implies $M^n(0) = E(X^n)$ for n ≥1.

Example 7.2: let X be a discrete random variable with pmf of p(x) = 1/3, for x = 0, 1, 2, then find the mgf, Mx(0), E(X) and V(X)

Solution: $Mx(t) = \sum e^{tx}\, p(x) = 1/3\ (1 + e^t + e^{2t})$

$Mx(0) = 1/3$

$M'(t) = d/dx(1/3\ (1 + e^t + e^{2t})) = 1/3\ (e^t + 2e^{2t})$

$M''(t) = d/dx(1/3\ (e^t + 2e^{2t})) = 1/3\ (e^t + 4e^{2t})$

M'(0) = 1/3 and M''(0) = 5/3

Then E(X) = M'(0) = 1/3 and $V(X) = M''(0) – [M'(0)]^2 = 5/3 – (1/3)^2 = 2/3$

Example 7.3: let X be a continuous random variable with pdf of $f(x) = e^{-x}$ x > 0, then find mgf and V(X).

Solution: $Mx(t) = E(e^{tx}) = \int e^{tx} f(x)dx = \int e^{(t-1)x}\, dx = \frac{1}{1-t}$

$M'(t) = \frac{1}{(1-t)^2}$ then M'(0) = 1 and $M''(t) = \frac{2}{(1-t)^3}$ then M''(0) = 2

⇨ $V(x) = M''(0) – [M'(0)]^2 = 2 - 1 = 1$

Example 7.4: let X has pdf of $f(x) = e^{-x}$ for x > 0 and let Y = 2x, then find mgf of Y and V(Y).

Solution: we have $Mx(t) = \frac{1}{1-t}$ from example 7.3 but here we should find My(t) as:

$My(t) = E(e^{ty}) = \int e^{ty} f(x)dx = \int e^{ty} e^{-x} dx = \int e^{(2t-1)x}\, dx = \frac{1}{1-2t}$

$V(Y) = M''(0) – [M'(0)]^2 = \frac{4}{(1-2t)^4} - [\frac{2}{(1-t)^2}]^2 = 4 – 2 = 2.$

ACTIVITY 7.1:

Find the moment generating function of a random variable X having density function given by:

$$f(x) = \begin{cases} 2e^{-2x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

Remarks:

- ✍ Two useful results concerning moment generating functions are, first, that the moment generating function uniquely determines the distribution function of the random variable and, second, that the moment generating function of the sum of independent random variables is equal to the product of their moment generating functions.
- ✍ It is also possible to define the joint moment generating function of two or more random variables. Let X and Y be two random variables with *m.g.f*`s, $M_x(t)$ and My(t), respectively. If $M_x(t) = M_y(t)$ for all value of t, then X and Y have the same probability distribution.That is, the moment generating function (*m.g.f*) is unique and completely determines the distribution function of the random variable X. Thus, two random variables having the same *m.g.f.* then would have the same distribution.In general, For any *n* random variables $X_1, . . . , X_n$, the joint moment generating function, $M(t_1, . . . , t_n)$, is defined, by $M(t_1, . . . , t_n) = E[e^{t_1X_1+ \cdot\cdot\cdot +t_nX_n}]$ for all real values of $t_1, . . . , t_n$.
- ✍ Suppose that X and Y are two independent random variables. Let Z = X + Y. Let $M_X(t)$, $M_Y(t)$ and $M_Z(t)$ be the *m.g.f*`s, of the random variable X, Y and Z respectively. Then for all value of $M_Z(t) = M_x(t)\, M_y(t)$. It is true for $X_1, …, X_n$ independent random variables as: $M(t_1, . . . , t_n) = M_{X_1}(t_1) \cdot \cdot \cdot M_{X_n}(t_n)$.
- ✍ If $M_X(t)$ is the moment generating function of the random variable *X* and *a* and *b* ($b \neq 0$) are constants, then the moment generating function of $(X + a) / b$ is $M_{(X + a)/b}(t) = e^{at/b} M_X(\frac{t}{b})$.

SUMMARY

- ☞ The moment generating function of the random variable X is defined by $M(t) = E[e^{tX}]$.
- ☞ The moments of *X* can be obtained by successively differentiating *M(t)* and then evaluating the resulting quantity at $t = 0$.
- ☞ The moment generating function uniquely determines the distribution function of the random variable and the sum of independent random variables is equal to the product of their moment generating functions.

CHECKLIST 7.1

Put a tick mark (√) for each of the following questions if you can solve the problems, and an X otherwise. Can you

1. Distinguish between the moment and the moment generating function of random variable? ☐
2. State the moment generating function of random variable? ☐
3. Compute the mean and variance of a random variable from its moment generating function? ☐

EXERCISE 7.1

1. Let X be a random variable with *p.d.*$f(x) = e^{-x}$, x > 0. Then
   (a) Find the *m.g.f.* $M_X(t)$ of a random variable X.
   (b) Find E(X), $E(X^2)$ and Var (X) by using $M_X(t)$.
   (c) If the random variable Y is defined by Y = 2 – 3X, determine $M_Y(t)$.
2. Find the moment generating function of a random variable Y = (X +1)/2from a density function of: $f(x) = \begin{cases} \frac{x}{2} & 0 \le x \le 2 \\ 0 & otherwise \end{cases}$
3. Find the first four moments of the random variable X having *pdf of:*

   $f(x) = \begin{cases} \frac{4x(9-x^2)}{81} & 0 \le x \le 3 \\ 0 & otherwise \end{cases}$ Then find the moment

   (a) About the origin, (b) About the mean, (c) The variance of X.

CHAPTER 8

COMMON DISCRETE PROBABILITY DISTRIBUTIONS AND THEIR PROPERTIES

Introduction

In this chapter we shall study some of the most common discrete probability distributions that figure most prominently in statistical theory and applications. We shall also study their parameters, that is, the quantities that are constants for particular distributions but that can take on different values for different members of families of the distribution of same kind. The most common parameters are the lower moments, mainly μ and the variance (sigma square).

Contents

Learning Outcomes

At the end of this chapter students will be able to:

- ✓ Define discrete probability distributions.
- ✓ Distinguish different discrete probability distributions
- ✓ Identify properties of different discrete probability distributions
- ✓ Identify the parameters of discrete probability distributions
- ✓ Solve Problems related to discrete probability distribution.
- ✓ Apply discrete probability distributions for real problems.

RESOURCES:

- 🕮 Statistics for Engineering and The Sciences, William M. & Terry S. (2007), 5th ed., pp. 211-217.
- 🕮 Probability & Statistics for Engineers & Scientists, Sharon L. Myers et al. (2012), 9th ed., pp. 143-164.
- 🕮 A First Course in Probability, Sheldon Ross (2010). 8th ed., pp. 263-266 & pp. 134-160.

8.1. Bernoulli Distribution

Bernoulli's trial is an experiment where there are only two possible outcomes, “success" or "failure". An experiment considered into a Bernoulli trial by defining one or more possible results which we are interested as ‘‘Success” and all other possible results as “Failure”. For instance, while rolling a fair die, a "success" may be defined as "getting even numbers on top" and odd numbers as "Failure".Generally, the sample space in a Bernoulli trial is S = {S, F}, S = Success, F = failure.

Therefore if an experiment has two possible outcomes “success” and “Failure”, their probabilities are $\theta$and $(1-\theta)$ respectively. Then the number of success (0 or 1) has a Bernoulli distribution.

Definition 8.1:

A random variable X has Bernoulli distribution and it referred to as a Bernoulli random variable if and only if its probability distribution given by: $f(x;\ \theta) = \theta^{x}(1-\theta)^{1-x},\ for\ \ x = 0, 1.$

Le X be a Bernoulli random variable having probability of success θ then the:

Property 1: ***Mean*** $E(X) = \sum_{x=0}^{1} x\, f(X=x) = \sum_{x=0}^{1} x\,\theta^{x}(1-\theta)^{1-x} = \theta$

Property 2:***Variance***$Var(X) = E(X - E(x))^2 = \theta\,(1-\theta)$

Property 3:***Moment Generating Function***: $M_x(t) = E(e^{tx}) = 1-\theta\,(1-e^{t})$ *for any t* $\Re$

Example 8.1:

- Tossing a coin and considering heads as success and tails as failure.
- Checking items from a production line: success = not defective, failure = defective.
- Phoning a call centre: success = operator free; failure = no operator free.
- Success of medical treatment.
- Student passes exam.
- A fair die is tossed. Let X =1 only if the first toss shows a “4” or “5”. Then $X \sim Be(1,\ \frac{1}{3})$

8.2. Binomial Distribution

In this sub-unit, we shall study one of the most popular discrete probability distributions, namely, the Binomial distribution. It simplifies many probability problems which, otherwise, might be very tedious and complicated while listing all the possible outcomes of an experiment.

Many real life experiments result from conducting a series of Bernoulli trails. Repeated trials play an important role in probability and statistics, especially when the number of trial is fixed, the parameter p (the probability of success) is same for each trial, and the trial are all independent. Several random variables are a rise in connection with repeated trials. The one we shall study here concerns the total number of success.

Examples of Binomial Experiments

- Tossing a coin 20 times to see how many tails occur.
- Asking 200 people whether they watch BBC news.
- Rolling a die 10 times to see if a 5 appears.

Derivation of the Binomial Distribution

Consider a set of n independent Bernoulli trials (n being finite) in which the probability of success in any trail is constant. Then, this gives rise to a binomial distribution.

To sum up these conditions, the binomial distribution requires that:

- An experiment repeated n times.
- Only two possible outcomes: success (S) or Failure (F).
- P(S) $=\theta$ (fixed at any trial).
- The n-trials are independent

Any experiment satisfying these four assumptions (or conditions) is said to have a binomial probability distribution.

Note: Since S and F are complementary events, $P(F)=1-P(S)=1-\theta=q$.

Let X be the number of successes in the n trials. Consider the event of getting k successes (X = k). Out of the n trials, if k aresuccesses, then the remaining (n-k) are failures, observed in any order, say, $S\ S\ F\ S\ F\ F\ S\ldots F\ S\ F$.

Since each trial is independent of the other, by the probability of independent events, $P(S\ S\ F\ S\ F\ F\ S\ldots F\ S\ F)=P(S).P(S).P(F).P(S).P(F).P(F).P(S)\ldots P(F).P(S).P(F)$

$$=\theta.\theta.q.\theta.q.q.\theta.\ldots q.\theta.q$$

$$= \underbrace{\theta.\theta.\theta\ldots\theta}_{k\ factors}.\ \underbrace{q.q.q\ldots q}_{n-k\ factors} = \theta^k q^{n-k}.$$

But k successes in n trials can occur in $\binom{n}{k}$ ways (recall that this is the number of possible selection of k out of n) and the probability of each of these ways is the same, viz., $\theta^k q^{n-k}$.

Hence, the probability of k successes in n trial in any order is $\binom{n}{k}\theta^k q^{n-k}$.

Such is what we call the binomial probability distribution, for the obvious reason that the probabilities of 0,1, 2, …, n successes, viz., $q^n, \binom{n}{1}\theta q^{n-1}, \binom{n}{2}\theta^2 q^{n-2}, \cdots, \theta^n$, are the successive terms in the binomial expansion of $(\theta+q)^n$.

N.B: Take n = 4 and k = 2. Then, there are $\binom{4}{2}=6$ possible outcomes (can you list them?).

Definition 8.2:

A random variable X has Binomial distribution and it referred to as a Binomial random variable if and only if its probability distribution given by: $f(x; n,\ \theta) = \binom{n}{x}\theta^x(1-\theta)^{n-x}$ *for x = 0, 1, . . . , n.*In generalbinomial distribution has the following characteristics:

Remark:

- The numbers given by ${}_nC_r$ are often called binomial coefficients, because they appear in the binomial expansion and have many interesting properties in connection with the binomial distribution.
- The two constants, n and p, are known as the **parameters** of the distribution, and the notation $X \sim$ B(n, p) shall be used to denote that the r-v X follows binomial distribution with parameters n and p. The above pmf is also denoted by $B(k; n, p)$.

Le X be a Binomial distribution with n number of trials and probability of success θ then the:

Property 1: ***Mean*****:** $E(X) = \mu = \sum xp(x) = \sum_{x=1}^{n} x\binom{n}{x}\theta^x(1-\theta)^{n-x} = n\theta$

Property 2: ***Variance*****:** Var(X) = E(X – E(x))^2 = $n\theta\ (1-\theta)$

Property 3: ***Moment Generating Function*****:** $M_x(t) = E(e^{tx}) = \sum e^{tx}\, p(x) = (q + pe^t)^n,\ t \in \Re$

Remark

✍ the mean of the Binomial distribution is

$$E(X) = \sum_{x=0}^{n} x\,P(X = x)$$

$$= \sum_{x=0}^{n} x\ {}^{n}c_x\, p^x q^{n-x}$$

$$= \sum_{x=0}^{n} x\ {}^{n}c_x\, p^x q^{n-x}$$

$$= \sum_{x=0}^{n} x \frac{n!}{x!(n-x)!} p^x q^{n-x}$$

$$= \sum_{x=0}^{n} x \frac{n(n-1)!}{x(x-1)!(n-x)!} p\, p^{x-1} q^{n-x}$$

$$= np \sum_{x=1}^{n} \frac{(n-1)!}{(x-1)!(n-x)!} p^{x-1} q^{n-x}$$

$$= np \sum_{x=1}^{n} {}^{n-1}c_{x-1}\, p^{x-1} q^{n-x}$$

$$= np(q+p)^{n-1}$$

$$= np(1)^{n-1} \quad [\because q + p = 1]$$

$$= np$$

$\therefore$ The mean of the binomial distribution is np

✍ Variance of the Binomial distribution:

The variance of the Binomial distribution is

$$V(X) = E(X^2) - [E(X)]^2$$

$$= E(X^2) - (np)^2 \ldots\ldots\ldots\ldots\ldots\ldots..\ (1)\ [\because E(X) = np]$$

Now,

$$E(X^2) = \quad = \sum_{x=0}^{n} x^2\ {}^{n}c_x\, p^x q^{n-x}$$

$$= \sum_{x=0}^{n} [x(x-1) + x]\ {}^{n}c_x\, p^x q^{n-x}$$

$$= \sum_{x=0}^{n} x(x-1)\frac{n!}{x!(n-x)!}p^x q^{n-x} + \sum_{x=0}^{n} x\frac{n!}{x!(n-x)!}p^x q^{n-x}$$

$$= \sum_{x=0}^{n} x(x-1)\frac{n(n-1)(n-2)!}{x(x-1)(x-2)!(n-x)!}p^2\ p^{x-2}q^{n-x} + E(X)$$

$$= n(n-1)p^2\sum_{x=2}^{n}\frac{(n-2)!}{(x-2)!(n-x)!}p^{x-2}q^{n-x} + np$$

$$= n(n-1)p^2\sum_{x=2}^{n}{}^{n-2}c_{x-2}\ p^{x-2}q^{n-x} + np$$

$$= n(n-1)p^2(q+p)^{n-2} + np$$

$$= n(n-1)p^2\,(1)^{n-2} + np \quad [\because q+p=1]$$

$$= n(n-1)p^2 + np \text{ .... (2)}$$

Putting (2) in (1) we get

$$V(X) = n(n-1)p^2 + np - (np)^2$$

$$= np(np - p + 1 - np)$$

$$= np(1-p)$$

$$= npq$$

$\therefore$ The variance of the Binomial distribution is npq

Example 8.2: A machine that produces stampings for automobile engines is malfunctioning and producing 5%defectives. The defective and non-defective stampings proceed from the machine in a random manner. If the next five stampings are tested, find the probability that three of them are defective.

Solution: Let x equal the number of defectives in n = 5 trials. Then x is a binomial random variable with p, the probability that a single stamping will be defective, equal to 0.05, and q = 1- 0.05 = 1 – 0.05 = 0.95. The probability distribution for x is given by the expression:

$$P(X=3) = \binom{5}{3}0.05^3(1-0.05)^{5-3}$$

$$= \frac{5!}{3!(5-3)!}(0.05)^3(0.95)^2$$

$$= \frac{5x4x3x2x1}{3x2x1(2x1)}(0.05)^3(0.95)^2$$

Mean = np = 5x0.05 = 0.25 and variance = npq = 5x0.05x0.95 = 0.2375.

Example 8.3: Find the probability of getting five heads and seven tails in 12 flips of a balanced coin.

Solution: Given n = 12 trials. Let X be the number of heads.Then, p = Prob. of getting a head =1/2, and q = prob. of not getting a head=1/2. Therefore, the probability of getting k heads in a random trial of a coin 12 times is:

$$P(X=x)=\binom{12}{x}\left(\frac{1}{2}\right)^{x}\left(\frac{1}{2}\right)^{12-x}=\frac{\binom{12}{x}}{2^{12}}=\frac{\binom{12}{x}}{4096}$$. And for x =5, $P(X=5)=\frac{\binom{12}{5}}{4096}=\frac{792}{4096}=0.1934$.

Example 8.4: If the probability is 0.20 that a person traveling on a certain airplane flight will request a vegetarian lunch, what is the probability that three of 10 people traveling on this flight will request a vegetarian lunch?

Solution: Let X be the number of vegetarians. Given n = 10, p = 0.20, x = 3; then,

$$P(X=3)=\binom{10}{3}(0.2)^{3}(0.8)^{7}=0.201.$$

CHECKLIST 8.1:

Put a tick mark (√) for each of the following questions if you can solve the problems, and an X otherwise.

1. Can you state the assumptions underlying the binomial distribution? ☐
2. Can you write down the mathematical formula of the binomial distribution? ☐
3. Can you compute probabilities of events in a binomial distribution? ☐
4. Can you define and compute probabilities with hyper geometric rule? ☐

EXERCISE 8.1

1. The probability that a patient recovers from a rare blood disease is 0.4. If 100 people are known to have contracted this disease, ***what is the probability that less than 30 survive***?
2. A multiple-choice quiz has 200 questions each with 4 possible answers of which only 1 is the correct answer. What is the probability that sheer guess-work yields from 25 to 30 correct answers for 80 of the 200 problems about which the student has no knowledge?

3. A component has a 20% chance of being a dud. If five are selected from a large batch, what is the probability that more than one is a dud?

4. A company owns 400 laptops. Each laptop has an 8% probability of not working. You randomly select 20 laptops for your salespeople. (a) What is the likelihood that 5 will be broken?(b) What is the likelihood that they will all work?

5. A study indicates that 4% of American teenagers have tattoos. You randomly sample 30 teenagers. What is the likelihood that exactly 3 will have a tattoo?
6. An XYZ cell phone is made from 55 components. Each component has a .002 probability of being defective. What is the probability that an XYZ cell phone will not work perfectly?
7. The ABC Company manufactures toy robots. About 1 toy robot per 100 does not work. You purchase 35 ABC toy robots. What is the probability that exactly 4 do not work?
8. The LMB Company manufactures tires. They claim that only .007 of LMB tires are defective. What is the probability of finding 2 defective tires in a random sample of 50 LMB tires?
9. An HDTV is made from 100 components. Each component has a .005 probability of being defective. What is the probability that an HDTV will not work perfectly?

8.3. Poisson Distribution

The Poisson probability distribution, named for the French Mathematician S.D. Poisson (1781-1840), provides a model for the relative frequency of the number of "rare events" that occurs in a unit of time, area, volume, etc.

Examples of events whose relative frequency distribution can be Poisson probability distributions are:

The number of new jobs submitted to a computer in any one minute,

The number of fatal accidents per month in a manufacturing plant,

The number of customers arrived during a given period of time,

The number of bacteria per small volume of fluid,

The number of customers arrived during a given period of time.

The properties of Poisson random variables are the following.

- The experiment consists of counting the number of items X a particular event occurs during a given units,

- The probability that an event occurs in a given units is the same for all the units,
- The number of events that occur in one unit is independent of the number that occurs in other units.

Definition 8.3:

A random variable X has Poisson distribution with parameter λ and it referred to as a Poisson random variable if and only if its probability distribution given by: $p(x;\ \lambda) = \frac{\lambda^x e^{-\lambda}}{x!}$ *for x = 0, 1, 2, . . .*

Le X is a Poisson distribution with an average number of time an event occur (parameter)λ then:

Property 1: *Mean*: $E(X) = \mu = \sum_{x=0}^{\infty} x \frac{\lambda^x e^{-\lambda}}{x!} = \lambda$

Property 2: *Variance*: $Var(X) = E(X - E(x))^2 = \lambda$

Property 3: *Moment Generating Function*:

$M_x(t) = E(e^{tx}) = \sum_{x=0}^{\infty} e^{tx} \frac{\lambda^x e^{-\lambda}}{x!} = e^{\lambda(e^t - 1)}$ *for any t in* $\Re$

Remark:

- When n is large, the calculation of binomial probabilities will usually be tedious. In such cases, it can be approximated by the Poisson distribution. Let X be a binomial random variable with parameters n and p. Then, the Poisson distribution is the limiting case of the binomial distribution under the conditions: the number of trials, n is indefinitely large, i.e., $n \to \infty$; $P(S) = p \to O$ (Indefinitely small); and $np = \lambda$ (say), is constant. Then,

$\lim_{n\to\infty} P(X = x) = \lim_{n\to\infty} \binom{n}{x} p^x q^{n-x} = \frac{e^{-np}(np)^x}{x!}$. In general, the Poisson approximation will provide a excellent approximation if n ≥ 100 and p ≤ 0.05 or np ≥ 10.

Example 8.5: Suppose that customers enter a waiting line at random at a rate of 4 per minute. Assuming that the number entering the line during a given time interval has a Poisson distribution, find the probability that:

a) one customer enters during a given one-minute interval of time;

b) at least one customer enters during a given half-minute time interval.

Solution:a) Given $\lambda = 4$ per min, $P(x=1) = \frac{4^1 e^{-4}}{1!} = 4e^{-4} = 0.0733$.

b) Per half-minute, the expected number of customers is 2, which is a new parameter.

$P(X \geq 1) = 1 - P(X = 0)$, but $P(X = 0) = e^{-2} = 0.1353$.

$\therefore P(X \geq 1) = 1 - 0.1353 = 0.8647$.

ACTIVITY 8.1:

1. A certain kind of carpet has, on the average, five defects per 10 square meters. Assuming Poisson distribution, find the probability that a 15 square meter of the carpet will have at least 2 defects.
2. If X~ $P(\lambda)$, then show that $Z = \frac{X - \lambda}{\sqrt{\lambda}}$ has mean 0 and variance unity.

CHECKLIST

Put a tick mark (√) for each of the following questions if you can solve the problems, and an X otherwise.

1. Can you approximate the binomial distribution with Poisson? ☐
2. Can you state the conditions for these approximations? ☐
3. Can you write down the pmf of the Poisson distribution? ☐
4. Can you compute the probabilities related with the Poisson distribution? ☐

8.4 Geometric distribution

Geometricdistribution arises in a binomial experiment situation when trials are carried out independently (with constant probability*p*of an *Success*) until the *first* occurs. The random variable *X* denoting the number of required trials is a *geometrically* distributed with *parameter p*.

Often we will be interested in measuring the length of time before some event occurs, for example, the length of time a customer must wait in line until receiving service, or the length of time until a piece of equipment fails. For this application, we view each unit of time as Bernoulli trail and consider a series of trails identical to those described for the Binomial experiment. Unlike the Binomial experiment where X is the total number of successes, the random variable of interest here is X, the number of trails (time units) until the first success is observed.

Definition 8.4:

A random variable X has Geometric distribution with parameter P and it referred to as a Geometric random variable if and only if its probability distribution given by: $p(x;\, p) = p(1-p)^{x-1}, x =$ *0,1, 2, . . ., where p is probability of success and x is number of trials until the first success occurs.*

Le X is a Poisson distribution with an average number of time an event occur (parameter)λ then:

Property 1: ***Mean*:** $E(X) = \mu = \sum_{x=1}^{\infty} xp(1-p)^{x-1} = \frac{1}{p}$

Property 2: ***Variance*:** $Var(X) = E(X - E(x))^2 = \frac{1-p}{p^2} = \frac{q}{p^2}$

Property 3: ***Moment Generating Function:***

$M_x(t) = E(e^{tx}) = \sum_{x=1}^{\infty} e^{tx} p(1-p)^{x-1} = \frac{pe^t}{1-(1-p)e^t} = \frac{pe^t}{1-qe^t}$ *for any t,* $t < -log(q)$

Remark

- ✍ The mean and variance for a Poisson distribution are both λ.

$E(X) = \sum_{x=0}^{\infty} x \frac{\lambda^x}{x!} e^{-\lambda}$, (letting $y = x - 1$) $= e^{-\lambda}\lambda \sum_{x=1}^{\infty} \frac{\lambda^{x-1}}{(x-1)!}$

$$= e^{-\lambda}\lambda \sum_{y=0}^{\infty} \frac{\lambda^y}{y!} = e^{-\lambda}\lambda e^{\lambda} = \lambda$$

To calculate Var(X), we first calculate

$$E(X^2) = \sum_{x=0}^{\infty} x^2 \frac{\lambda^x}{x!} e^{-\lambda} = \lambda \sum_{x=1}^{\infty} \frac{(x-1+1)\lambda^{x-1}}{(x-1)!} e^{-\lambda} = \lambda \left[\sum_{x=1}^{\infty} \frac{(x-1)\lambda^{x-1}}{(x-1)!} e^{-\lambda} + \sum_{x=1}^{\infty} \frac{\lambda^{x-1}}{(x-1)!} e^{-\lambda} \right]$$

(rewriting x – 1 as y) then $E(X^2) = \lambda \left(\sum_{y=0}^{\infty} \frac{y\lambda^y}{y!} e^{-\lambda} + \sum_{y=0}^{\infty} \frac{\lambda^y}{y!} e^{-\lambda} \right) = \lambda\left(\lambda + e^{-\lambda}e^{\lambda}\right) = \lambda^2 + \lambda$, hence

$Var(X) = E(X^2) - (E(X))^2 = \lambda^2 + \lambda - \lambda = \lambda$.

Example 8.6: If the probability is 0.75 that an applicant for a driver's license will pass the road test on any given try. What is the probability that an applicant will finally pass the test on the fourth try?

Solution: Assuming that trials are independent, we substitute x=4 and p=0.75 into the formula for the geometric distribution, to get: p(x) = $p(1-p)^{x-1}$ = $0.75(1-0.75)^{4-1}$ = $0.75(0.25)^3$ = 0.011719

Activity 8.2:

A manufacturer uses electrical fuses in an electronic system, the fuses are purchased in large lots and tested sequentially until the first defective fuse observed. Assume that the lot contains 10% defectives fuses.

(a) What is the probability that the first defective fuse will be one of the first five fuses tested?

(b) Find the mean and SD for X.

8.5 Negative Binomial Distribution

In connection with repeated Bernoulli trials, as we are sometimes interested in the number of the trial on which the r^{th} success occurs. For instance, the probability that the fifth person to hear a rumor will be the first one to believe it, or the probability that a burglar will be caught for the second time on his or her eight job.

Definition 8.6:

Suppose that independent trials, each having probability *p*, 0 <*p* <1, of being asuccess are performed until a total of *r* successes is accumulated. If we let *X* equal thenumber of trials required, then$(x,r,p) = \binom{n-1}{r-1} p^r(1-p)^{n-r}$, n = r, r+1, r+2,….

Remark:

- It follows because, in order for the r^{th} success to occur at the n^{th} trial, there must be $r-1$ successes in the first $n-1$ trials and the *n*th trial must be a success. The probability of the first event is: $\binom{n-1}{r-1} p^{r-1}(1-p)^{n-r}$and the probability of the second is *p*; thus, by independence, the above probability distribution is established.
- Any random variable *X* whose probability mass function is said to be a *negative binomial* random variable with parameters (*r*, *p*). Note that a geometric random variable is just a negative binomial with parameter (1, *p*).

Le X is a Negative Binomial distribution with its parameters k andθ then:

Property 1: ***Mean*****:** $E(X) = \mu = \sum_{x=k}^{\infty} x \binom{x-1}{k-1} \theta^k (1-\theta)^{x-k} = \frac{k}{\theta}$

Property 2:***Variance*****:** $Var(X) = E(X - E(x))^2 = \frac{k\,(1-\theta)}{\theta^2}$

Property 3: ***Moment Generating Function:***

$M_x(t) = E(e^{tx}) = \left(\frac{\theta e^t}{1-(1-\theta)e^t}\right)^k$ *for any t,* $t < -log(1-\theta)$

Example 8.8:

1. If the probability is 0.40 that a child exposed to a certain contagious disease will catch it, what is the probability that the tenth child exposed to the disease will be the third to catch it?

Solution: Substituting x=10, k=3, and p=0.40 into the formula for the NB distribution, we get,NB(10; 3, 0.40)= $\binom{9}{2}(0.40)^3(0.60)^7$ =0.0645.

ACTIVITY 8.3:

If the probability is 0.40 that a child exposed to acertain contagious disease will catch it. Whatis the probability that the tenth child exposed to the disease will be the third to catch?

CHECKLIST

Put a tick mark (√) for each of the following questions if you can solve the problems, and an X otherwise.

1. Can you state the conditions to apply the NB & Geometric distributions? ☐
2. Can you write down the mathematical formula of the NB & Geometric distributions? ☐
3. Can you compute probabilities of events in the NB & Geometric distributions? ☐

EXERCISE

1. If the probability is 0.75 that a person will believe a rumor about the transgressions of a certain politician, find the probabilities that

   a) the eighth person to hear the rumor will be the fifth to believe it;

b) the fifteenth person to hear the rumor will be the tenth to believe it.

2. If the probabilities of having a male or female child are both 0.50, find the probabilities that a family's
   a) fourth child is their first son;
   b) seventh child is their second daughter;
   c) tenth child is their fourth or fifth son?
3. An expert sharpshooter hits a target 95% of the time. Find the probability that she will miss the target for the second time on the fifteenth shot.
4. The probability that a burglar will get caught on any given "job" is 0.20. Find the probability that he will get caught for the first time on his fifth "job".
5. A die is cast until 6 appears. What is the probability that it must be cast more than five times?
6. An item is produced in large numbers. The machine is known to produce 5% defectives. A quality control inspector is examining the items by taking them at random. What is the probability that at least 4 items are to be examined in order to get 2 defectives?

8.6 Hypergeometric Distribution

We are interested in computing probabilities for thenumber of observations that fall into a particular category. But in the case of thebinomial distribution, independence among trials is required. As a result, if thatdistribution is applied to, say, sampling from a lot of items (deck of cards, batchof production items), the sampling must be done **with replacement** of each itemafter it is observed. On the other hand, the hypergeometric distribution does notrequire independence and is based on sampling done **without replacement**.

Applications for the hyper geometric distribution are found in many areas, withheavy use in acceptance sampling, electronic testing, and quality assurance. Obviously,in many of these fields, testing is done at the expense of the item beingtested. That is, the item is destroyed and hence cannot be replaced in the sample.Thus, sampling without replacement is necessary.

In general, we are interested in the probability of selecting x successes fromthe Mitems labeled successes and $n - x$ failures from the $N - M$items labeled failures when a random sample of size n is selected from N items. This is known as a **hypergeometric experiment**, that is, one that possesses the following two properties: A random sample of size n is selected without replacement from N items; and of the N items, Mmay be classified as successes and $N - M$are classified asfailures.The number X of successes of a hypergeometric experiment is called a **hypergeometricrandom variable**.

Definition 8.6:

The probability distribution of the hypergeometric random variable X, the numberof successes in a random sample of size n selected from N items of which M are labeled **success** and $N - M$labeled **failure**, is: $p(x; n, N, M) = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}$ *for* $x = 0, 1, 2, \ldots, n;\ x \le M, n - x \le N - M.$

The range of x can be determined by the three binomial coefficients in thedefinition, where x and $n-x$ are no more than $M$and $N - M$, respectively, and bothof them cannot be less than 0. Usually, when both M(the number of successes)and $N - M$(the number of failures) are larger than the sample size n, the range ofa hypergeometric random variable will be $x = 0, 1, \ldots, n$.

Let X be a Hypergeometric distribution with N items, selected sample size *n*among M labeled success then:

Property 1: ***Mean***: $E(X) = \mu = \sum_{x=0}^{min(k,n)} x \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}} = \frac{nM}{N}$

Property 2: ***Variance***: $Var(X) = E(X - E(x))^2 = \frac{nM\,(N-M)(N-n)}{N^2\,(N-1)}$

Property 3: ***Moment Generating Function:*** ***Not Given***

Remark

- ✍ When the number of samples in the lot is large, then the hypergeometeric probability mass function is approximated in to the probability mass function of a binomial random variable.

Example 8.10:Lots of 40 components each are deemed unacceptable if they contain 3 or more defectives. The procedure for sampling a lot is to select 5 components at random and to reject the lot if a defective is found. What is the probability that exactly 1 defective is found in the sample if there are 3 defectives in the entire lot?

***Solution*:** Using the hypergeometric distribution with n = 5, N = 40, M= 3, and x = 1, wefind the probability of obtaining 1 defective to be$p(1; 40, 5, 3) = \frac{\binom{3}{1}\binom{40-3}{5-1}}{\binom{40}{5}}$ = 0.3011.Once again, this plan is not desirable since it detects a bad lot (3 defectives) onlyabout 30% of the time.

Example 8.11: Two balls are selected at random and removed from a bag containing 5 blue and 3 green balls in succession.Find the pmf of blue balls.

Solution: If we letX: selection of blue balls (success), then given are a = 5 (blue balls), b = 3 (green balls), n = 2.Then, the probability of selecting blue balls is:

$P(X = x) = f(x) = \frac{{}_5C_x \times {}_3C_{2-x}}{{}_8C_2}$, x=0,1,2. So that, $f(0) = \frac{3}{28}$, $f(1) = \frac{15}{28}$, and $f(2) = \frac{10}{28}$.

ACTIVITY 8.4:

1. An urn contains 8 blue balls and 12 white balls. If five are drawn at random, without replacement. What is the probability that the sample will contain two blue and three white?
2. Among 16 applicants for a job, 10 have college degrees. If three of the applicants are randomly chosen for interviews, what are the probabilities that: (a) none has college degrees; (b) two have college degrees; (c) one has a college degree; (d) all three have college degrees?

8.7 Multinomial Distribution

An immediate generalization of binomial distribution arises when each trial has more than two possible outcomes. The probabilities of the respective outcomes are the same for each trial, and the trials are all independent. This would be the case, many types of experiments result in observations on a qualitative variable with more than two possible outcomes, for instance family income level as low, middle and high; when a person interviewed by an opinion poll are asked whether they are for a candidate, in favor or against her, or undecided or when samples of manufactured products are rated excellent, above average, average or inferior. To treat such kind of problem Multinomial distribution is very important. Such an experiment consists of n identical trials that are observation on n experiment units. Each trial must result in one and only one of k outcomes, the k classification categories.

The multinomial experiment should be:

- Consists n identical and independent trials
- There are k possible outcomes to each trial
- The probability of the k outcomes, denoted by $p_1, \ldots p_k$ remain the same trial to trial, where $p_1 + \ldots + p_k = 1$.
- The random variables of interest are the counts $X_1, X_2, X_3, \ldots, X_n$ in each of the k categories.

Definition 8.7:

A random variable $X_1, X_2, X_3, \ldots, X_n$ have Multinomial distribution and they are referred to as a Multinomial random variables if and only if their joint probability distribution given by: $f(x_1, x_2, \ldots, x_k;\, n, \theta_1, \theta_2, \ldots, \theta_k) = \binom{n}{x_1, x_2, \ldots, x_k} \theta_1^{x_1}, \theta_2^{x_2}, \ldots, \theta_k^{x_k}$ *for* $x_i = 0, 1, 2, 3, \ldots, n$ *for each i, where* $\sum_{i=1}^{k} x_i = n$ *and* $\sum_{i=1}^{k} \theta_i = 1$

Thus, the number of outcomes of the different kinds is random variables having the multinomial distributions of parameters, $n, \theta_1, \theta_2, \ldots, and \theta_k$.

Let X be a Multinomial random variable with its parameters.

Property 1: *Mean*: $E(X_1) = n\,\theta_1$, $E(X_2) = n\,\theta_2$ $E(X_3) = n\,\theta_3$ . . . $E(X_k) = n\,\theta_k$ or

$E(X_i) = n\theta_i \quad i = 1, 2, \ldots, k$

Property 2: *Variance*: $Var(X_i) = E(X_i - E(X_i))^2 = n\,\theta_i\,(1 - \theta_i)$ *for* $i = 1, 2, \ldots, k$

Property 3: *Moment Generating Function: Not Given*

ACTIVITY 8.5:

A certain city has three television stations. During prime time on Saturday nights, Channel 12 has 50 percent of the viewing audience, channel 10 has 30 percent of the viewing audience, and channel 3 has 20 percent of the viewing audience. Find the probability that among eight television viewers in that city, randomly choosen on a Saturday night, five will be watching Channel 12, two will be watching Channel 10, and one will be watching Channel 3?

SUMMARY

- ☞ The binomial pmf is given by: $P(X = x) = \binom{n}{x} p^x q^{n-x}, x = 0,1,2,\ldots,n$.
- ☞ For a binomial random variable, $E(X) = np$, and $V(X) = npq$.
- ☞ The pmf of a negative binomial distribution is given by: $NB(x;k,p) = \binom{x-1}{k-1} p^k q^{x-k}, \quad for \ \ x = k, \ k+1, \ k+2, \cdots$.
- ☞ The pmf of a geometric distribution is: G(x; p)=pq^{x-1},for x=k, k+1, k+2, ….
- ☞ When n is large and p is very small, the binomial is approximated by the Poisson distribution as: $P(X = x) = \frac{(np)^x e^{-np}}{x!}, \quad for \ x = 0,1,2,\ldots$.
- ☞ The Poisson distribution is used to model rare events and the pmf is given by:
- ☞ $P(X = x) = \frac{\lambda^x e^{-\lambda}}{x!}, x = 0,1,2,\ldots$, where λ is average number of successes.
- ☞ Both the Mean and Variance of a Poisson distribution equal to λ .

EXERCISES 8.1

1. Draw 6 cards from a deck without replacement. What is the probability of getting two hearts?
2. 42 balls are numbered 1 - 42. You select six numbers between 1 and 42. (The ones you write on your lotto card). 6 balls are selected at random. What is the probability that they contain: (a) 4 of yours? (b) 5 of yours?
3. Given that 51.3% of all newly born children are boys, then what is the probability that in a sample of 5 newly born children, exactly 3 are boys?
4. In a large collection of light bulbs we assume that 98% of these bulbs will not defective. If we select 10 bulbs from this collection, then what is the probability that 8 are not defective?

5. Of all the cars registered in Germany, 53% are German made. In a sample of 12 cars registered in Germany, what is the probability that 9 are foreign made?
6. *Consumer Reports* states that approximately 70% of all people who buy eyeglasses from a private doctor's office were highly satisfied. In a sample of 11 people buying eyeglasses from a private doctor, what is the probability that less than 10 are highly satisfied?
7. In the fabrication of steel beams, two types of flaws may occur: (1) the inclusion of a small quantity of foreign matter ("slag"); and (2) the existence of microscopic cracks. It has been found by careful laboratory investigation that for a certain size I-beam from a given foundry the mean distance between microscopic cracks is 40 feet along the beam, whereas the slag inclusions exist with an average rate of 4 per 100 feet of beam. Each of these types of flaw follows a Poisson process.
   (a) For a 20-foot I-beam of this size from this foundry, what is the chance of finding exactly 2 microscopic cracks in the beam?
   (b) For the same 20-foot beam, what is the chance of finding one or more slag inclusions?
   (c) If a 20-foot beam contained more than 2 flaws, it would be rejected. What is the probability that a 20-foot beam will be rejected?
   (d) Four 20-foot I-beams are supplied to a contractor by this foundry last year. Assume the flaw conditions between the four beams are statistically independent. What is the probability that only one of the beams had been rejected?

8. 8. The air quality in an industrial city may become substandard (poor) at times depending on the weather condition and the amount of factory production. Suppose the event of poor air quality occurs as a Poisson process with a mean rate of once per month. During each time period when the air quality becomes substandard, its pollutant concentration may reach a hazardous level with a 10% probability. Assume that the pollutant concentration between any two periods of poor air quality are statistically independent.

(a) What is the probability of at most 2 periods of poor air quality during the next 4-1/2 months?

(b) What is the probability that the air quality would ever reach hazardous level during the next three months?

CHAPTER 9

COMMON CONTINUOUS PROBABILITY DISTRIBUTIONS AND THEIR PROPERTIES

Introduction

In chapter one, we describes a large set of data by means of a relative frequency distribution. If the data represent measurements on a continuous random variable and if the amount of data is very large, we can reduce the width of the class intervals until the distribution appears to be smooth curve. Therefore, a probability density function is a theoretical model for this distribution and there are different types of continuous distribution are considered here in this section.

Contents

9.1 Uniform distribution

9.2 Normal distribution

9.3 Exponential distribution

9.4 Gama distribution

9.5 Chi - square distribution

9.6 t distribution

9.7 F distribution

Learning Outcomes

At the end of this chapter students will be able to:

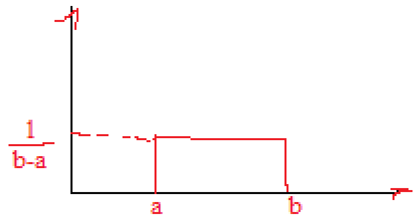
- ✓ Define discrete probability distributions.
- ✓ Distinguish different discrete probability distributions
- ✓ Identify properties of different discrete probability distributions
- ✓ Identify the parameters of discrete probability distributions
- ✓ Solve Problems related to discrete probability distribution.
- ✓ Apply discrete probability distributions for real problems.

RESOURCES:

- 🕮 Statistics for Engineering and The Sciences, William M. & Terry S. (2007), 5th ed., pp. 169-200.
- 🕮 Probability & Statistics for Engineers & Scientists, Sharon L. Myers et al. (2012), 9th ed., pp. 171-209 & 246-253.
- 🕮 A First Course in Probability, Sheldon Ross (2010). 8th ed., pp. 263-266 & pp. 186-211.

9.1 Uniform Distribution

One of the simplest continuous distributions in all of statistics is the **continuousuniform distribution**. This distribution is characterized by a density functionthat is "flat," and thus the probability is uniform in a closed interval, say [*a*, *b*].Suppose you were to randomly select a number X represented by a point in the interval $a \leq x \leq b$. The density function of X is represented graphically as follows.



Note that the density function forms a rectangle with base $b-a$and **constant height**$\frac{1}{b-a}$ to ensure that the area under the rectangle equals one. As a result, the uniform distribution is often called the **rectangular distribution**.

Definition 9.1:

A random variable of the shown in the above graph is called a uniform random variable. Therefore, the probability density function for a uniform random variable, X with the parameters of *a* and *b* is given by:

$$f(x) = \begin{cases} \frac{1}{b-a}, & a \leq x \leq b \\ 0, & elsewhere \end{cases}$$

Property 1: ***Mean***: $E(X) = \mu = \int_a^b x \frac{1}{b-a} dx = \frac{a+b}{2}$

Property 2: ***Variance***: $Var(X) = E(X - E(x))^2 = \frac{(b-a)^2}{12}$

Property 3: ***Moment Generating Function:***

$M_x(t) = E(e^{tx}) = \int_a^b e^{tx} \frac{1}{b-a} dx = \frac{e^{bt} - e^{at}}{t(b-a)}$ for $t \neq 0$

Example 9.1: The department of transportation has determined that the winning (low) bid X (in dollars) on a road construction contract has a uniform distribution with probability density function $f(x) = \frac{5}{8d}$, if $\frac{2d}{5} < x < 2d$, where **d** is the department of transportation estimate of the cost of job. (a) Find the mean and SD of X. (b) What fraction of the winning bids on road construction contracts are greater than the department of transportation estimate?

Solution: (a) E(X) = $\int_{2d/2}^{2d} x\,\frac{5}{8d}\,dx$ = *(2d- 2d/2)/2 = d/2*

$V(X) = E(X - E(x))^2 = \frac{(2d - 2d/2)^2}{12} = d^2/12$

(b) $p(X > d) = \int_{d}^{2d} \frac{5}{8d}\,dx = \frac{5}{8d}[x]_{d}^{2d} = \frac{5}{8d}(2d - d) = \frac{5}{8}$

Activity

Suppose the research department of a steel manufacturer believes that one of the company's rolling machines is producing sheets of steel of varying thickness. The thickness X is a random variable with values between 150 and 200 millimeters. Any sheets less than 160 millimeters thick must be scrapped, since they are unacceptable to buyers. (a) Calculate the mean and variance of X (b) Find the fraction of steel sheets produced by this machine that have to be scrapped.

9.2 Normal Distribution

The most important continuous probability distribution in the entire field of statisticsis the **normal distribution**.

Its graph, called the **normal curve**, is thebell-shaped curve which approximately describes many phenomenathat occur in nature, industry, and research. For example, physical measurementsin areas such as meteorological experiments, rainfall studies, and measurementsof manufactured parts are often more than adequately explained with a normaldistribution. In addition, errors in scientific measurements are extremely well approximatedby a normal distribution. In 1733, Abraham DeMoivre developed themathematical equation of the normal curve. It provided a basis from which muchof the theory of inductive statistics is founded. The normal distribution is oftenreferred to as the **Gaussian distribution**, in honor of Karl Friedrich Gauss(1777–1855), who also derived its equation from a study of errors in repeated measurementsof the same quantity.

A continuous random variable X having the bell-shaped distribution shown below is called a **normal random variable**. The mathematical equation for theprobability distribution of the normal variable depends on the two parameters $\mu$and σ, its mean and standard deviation, respectively. Hence, we denote the valuesof the density of X by $f(x; \mu, \sigma)$ or f(x)

The normal (or Gaussian) density function was proposed by C.F.Gauss (1777-1855) as a model for the relative frequency distribution of errors, such errors of measurement. Amazingly, this bell-shaped curve provides an adequate model for the relative frequency distributions of data collected from many different scientific areas.

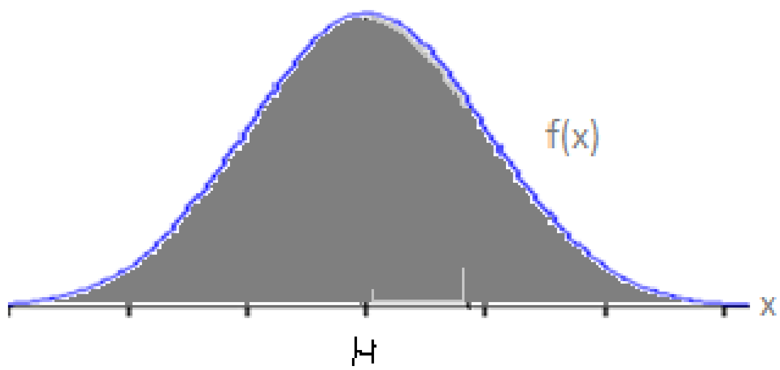
Definition 9.2:

A random variable X is normal or normally distributed with parameters μ and σ^2, (abbreviated N(μ, σ2)), if it is continuous with probability density function:

$$f(x) = \frac{1}{\sigma\sqrt{2\Pi}} e^{\frac{-1}{2}(\frac{x-\mu}{\sigma})^2} \qquad \infty < x < \infty; \sigma > 0 \text{ and } -\infty < \mu < \infty,$$ the parameters μ and σ^2 are the mean and the variance, respectively, of the normal random variable.

Properties of the Theoretical Normal Distribution

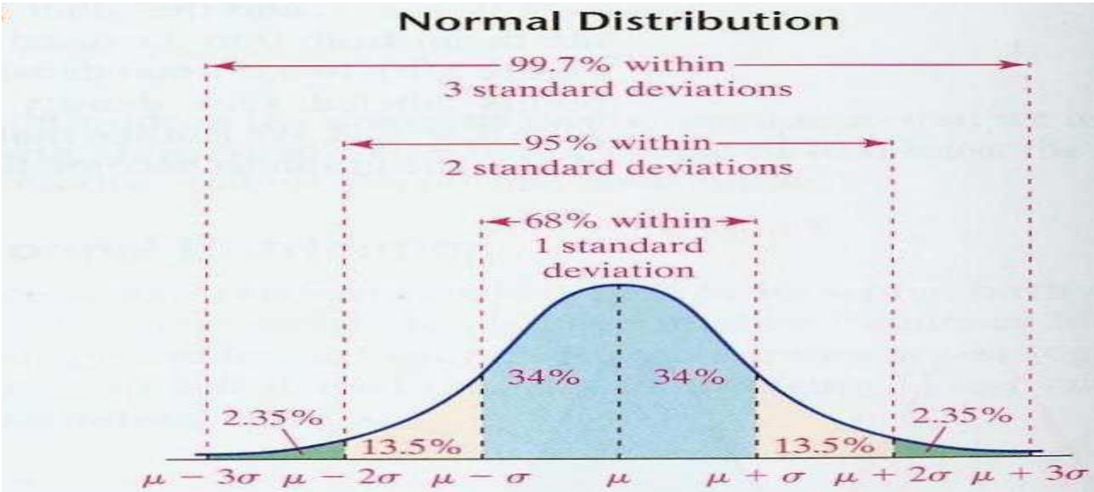
1. The curve is bell-shaped.



Normal probability curve.

2. The mean, median and mode are equal and located at the center of the distribution.
3. The curve is symmetrical about the mean and it is uni-modal.
4. The curve is continuous, i.e., for each X, there is a corresponding Y value.
5. It never touches the X axis.
6. The total area under the curve is 1 and half of it is **0.5000**
7. The areas under the curve that lie within one standard deviation, two and three standard deviations of the mean are approximately 0.68 (68%), 0.95 (95%) and 0.997 (99.7%) respectively.

Graphically, it can be shown as:



Remark

- Let X be a binomial random variable with parameters n and p. For large n, X has approximately a normal distribution with $\mu = np$ and $\sigma^2 = npq = np(1-p)$ and $P(X \le x) = \sum_{k=0}^{x} b(k,n,p) \approx$ area under normal curve to the left of $x + 0.5 = P(Z \le \frac{X+0.5-np}{\sqrt{npq}})$, *where* **+0.5** is called a **continuity correction** and the approximation will be good if np and $n(1-p)$ are greater than or equal to 5.

9.3 Standard Normal Distribution

If we want to compute the probability $P(a < X < b)$, we have to evaluate the area under the normal curve $f(x)$ on the interval (a, b). This means we need to integrate the function $f(x)$ defined above. Obviously, the integral is not easily evaluated. That is,

$P(a < X < b) = \frac{1}{\sigma\sqrt{2\pi}} \int_a^b e^{-\frac{(x-\mu)^2}{2\sigma^2}}\, dx$ cannot be integrated directly.

But this is easily evaluated using a table of probabilities prepared for a special kind of normal distribution, called the standard normal distribution (see Table A in the Appendix).

The following section is devoted to a discussion about the standard normal distribution and its application in the computation of probabilities.

If X is a normal random variable with the mean μ and variance σ then the variable $Z = \frac{X-\mu}{\sigma}$ is the standardized normal random variable. In particular, if μ = 0 and σ= 1, then the density function is called the standardized normal density and the graph of the standardized normal density distribution is similar to normal distribution.

Convert all normal random variables to standard normal in order to easily obtain the area under the curve with the help of the ***standard normal table***.

Definition 9.3:

Let X be a normal r-v with mean μ and standard deviation σ. Then we define the standard normal variable Z as: $Z = \frac{X-\mu}{\sigma}$. Then the pdf of Z is, thus, given by:

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}, -\infty < z < \infty.$$

Properties of the Standard Normal Curve (Z):

1. The highest point occurs at μ=0.
2. It is a bell-shaped curve that is symmetric about the mean, μ=0. One half of the curve is a mirror image of the other half, i.e., the area under the curve to the right of μ=0 is equal to the area under the curve to the left of μ=0 equals ½.
5. The total area under the curve equals one.
6. Empirical Rule:
   - Approximately 68% of the area under the curve is between -1 and +1.
   - Approximately 95% of the area under the curve is between -2 and +2.
   - Approximately 99.7% of the area under the curve is between -3 and +3.

Steps to find area under the standard normal distribution curve

i. Draw the picture

ii. Shade the desired area /region

i. If the area/region is:

- ✓ between 0 and any Z value, then look up the Z value in the table,
- ✓ in any tail, then look up the Z value to get the area and subtract the area from 0.5000,
- ✓ between two Z values on the same side of the mean, then look up both Z values from the table and subtract the smaller area from the larger,
- ✓ between two Z values on opposite sides of the mean, then look up both Z values and add the areas,
- ✓ less than any Z value to the right of the mean, then look up the Z value from the table to get the area and add 0.5000 to the area,

- ✓ greater than any Z value to the left of the mean, then look up the Z value and add 0.5000 to the area,
- ✓ in any two tails, then look up the Z values from the table, subtract the areas from 0.5000 and add the answers.

Note that finding the area under the curve is the same as finding the probability of choosing any Z value at random.

Example 9.3: Find the probabilities that a r-v having the standard N.D will take on a value

a) Less than 1.72; b) Less than -0.88;
c) Between 1.30 and 1.75; d) Between -0.25 and 0.45.

Solution: Making use of Table A,

a) $P(Z < 1.72) = P(Z < 0) + P(0 < Z < 1.72) = 0.5 + 0.4573 = 0.9573$.

b) $P(Z < -0.88) = P(Z > 0.88) = 0.5 - P(0 < Z < 0.88) = 0.5 - 0.3106 = 0.1894$.

c) $P(1.30 < Z < 1.75) = P(0 < Z < 1.75) - P(0 < Z < 1.30) = 0.4599 - 0.4032 = 0.0567$.

d) $P(-0.25 < Z < 0.45) = P(-0.25 < Z < 0) + P(0 < Z < 0.45)$.

$= P(0 < Z < 0.25) + P(0 < Z < 0.45) = 0.0987 + 0.1736 = 0.2723$.

ACTIVITY 9.1:

Find the area under the standard normal distribution curve between Z =0 and Z = 2.31 **[Ans: 0.4896]**, Z =0 and Z = -179 **[Ans: 0.4633]**, Z = 2.01 and Z = 2.34 **[Ans: 0.0126],** Z = -1.35 and Z = - 0.71 **[Ans: 0.1504],** Z = 1.21 and Z = -2.41 **[Ans: 0.8789],** to the right of Z = 1.54 **[Ans: 0.0618],** to the left of Z = 1.75 **[Ans: 0.9599]**

Example 9.4: Find a) $Z_{0.01}$; b) $Z_{0.05}$

Solution: a) $Z_{0.01}$ Corresponds to an entry of 0.5 - 0.01 = 0.4900. In Table A, look for the value closest to 0.4900, which is 0.4901, and the Z value for this is Z= 2.33. Thus, $Z_{0.01} = 2.33$.

b) Again, $Z_{0.05}$ is obtained as 0.5 - 0.05 = 0.4500, which lies exactly between 0.4495 and 0.4505, corresponding to Z = 1.64 and Z= 1.65. Hence, using interpolation, $Z_{0.05} = 1.645$.

Remark

- ✍ The curve of any continuous probability distribution or density function is constructed so that the area under the curve bounded by the two ordinates *a*= x_1 and b= *b*equals the probability that the random variable X assumes a value between $a = x_1$ and $x = b$. Thus, for the normal curve:

$$P(a < X < b) = P\left(\frac{a-\mu}{\sigma} < \frac{X-\mu}{\sigma} < \frac{b-\mu}{\sigma}\right) = P(z_1 < Z < z_2), say.$$

Now, we need only to get the readings from the Z- table corresponding to z_1 and z_2 to get the required probabilities, as we have done in the preceding example.

✍ If X is a binomial random variable with mean $\mu = np$ and variance $\sigma^2 = npq$, then the limiting form of the distribution of $Z = \frac{X-np}{\sqrt{npq}}$, as $n \to \infty$, is the standard normal distribution $n(z; 0, 1)$.

Example 9.5: If the scores for an IQ test have a mean of 100 and a standard deviation of 15, find the probability that IQ scores will fall below 112.

Solution: IQ ~ N(100, 225)

$$P(Y<112) = P[\frac{Y-\mu}{\sigma} < \frac{112-100}{15}]$$

$$= P[Z<.800] = 0.500 + P(0<Z<.800) = 0.500 + 0.2881 = 0.7881$$

Example 9.6: Suppose that X~ N (165, 9), where X = the breaking strength of cotton fabric. A sample is defective if X<162. Find the probability that a randomly chosen fabric will be defective.

Solution: Given that $\mu = 165$ and $\sigma^2 = 9$,

$$P(X < 162) = P\left(\frac{X-\mu}{\sigma} < \frac{162-\mu}{\sigma}\right) = P\left(Z < \frac{162-165}{3}\right)$$

$= P(Z < -1) = 0.5 - P(-1 < Z < 0)$ (Since $P(Z < 0) = 0.5$)

$= 0.5 - P(0 < Z < 1)$ (By symmetry)

$= 0.5 - 0.3413 = 0.1587$ (Table value for Z = 1)

ACTIVITY 9.2:

1. The average IQ score of students in a school for gifted children is 165 with a standard deviation of 27. A random sample of size 36 students is taken. What is the probability that:

(b) The sample mean score will be greater than 170;

(c) The sample mean score will be less than 158;

(d) The sample mean score will be between 155 and 160;

(e) The samples mean score is less than 170 or more than 175?

(Answer: a,0.5-0.3643=0.1357,b,0.5-.4332=.0668, c,0.4821-0.3643=.1178,d,1+0.3643-0.4821=0.8822)

2. The saving accounts maintained at an international bank have a mean birr μ and standard deviation of birr 1200. If a sample of 100 saving accounts is selected at random, what is the probability that the sample mean will be greater than the population mean by more than birr 120?

CHECKLIST

Put a tick mark (√) for each of the following questions if you can solve the problems, and an X otherwise:

1. Can you write down the pdf of the Normal Distribution (N.D)? ☐
2. Can you state and verify the properties of the normal curve? ☐
3. Can you define the standard N.D with its properties? ☐
4. Can you compute probabilities of a normal r-v? ☐
5. Can you state the conditions to approximate the binominal by the N.D? ☐
6. Can you approximate the binomial by the N.D under these conditions? ☐

EXERCISE

1. Find the value of Z if the area between -Z and Z is a) 0.4038; b) 0.8812; c) 0.3410.

3. The reduction of a person's oxygen consumption during periods of deep meditation may be looked up on as a r-v having the N.D with $\mu = 38.6$ cc per minute and $\sigma = 6.5$ cc per minute. Find the probabilities that during such a period a person's oxygen consumption will be reduced by (a) at least 33.4 cc per minute; (b) at most 34.7 cc per minute

4. A random variable X has a N.D with $\sigma = 10$. If $P(X < 82.5) = 0.8212$, find P(X> 58.3).

5. The yearly number of major earthquakes, over the world, is a r-v having approximately the N.D with $\mu = 20.8$ and $\sigma = 4.5$. Approximate the probabilities that in any given year there will be a) exactly 9; b) at most 19; c) at least 19 major earthquakes.

6. In a distribution exactly normal, 7% of the items are below 35, and 89% are under 63. Find the parameters of the distribution.

7. In each of the following cases, check whether or not the conditions for the normal approximation to the binomial distribution are satisfied. a) n = 200, p = 0.01; b) n = 150, p = 0.97; c) n = 100, p = 1/8; d) n = 65, p = 0.10.

8. A sample of 100 items is taken at random from a batch known to contain 40% defectives. Approximate the probabilities that the sample contain a)at least 44; b)exactly 44 defectives
9. A normal distribution has mean μ =62.5. Find σ if 20% of the area under the curve lies to the right of 79.2.
10. The mean grade on a final examination was 72 with standard deviation of 9. If the top 10% of the students are to receive A's, what is the minimum grade a student must get in order to receive an A?

9.4 Exponential Distribution

Exponential distribution is an important density function that employed as a model for the relative frequency distribution of the length of time between random arrivals at a service counter when the probability of a costumer arrival in any one unit of time is equal to the probability of arrival during any other. It is also used as a model for the length of life of industrial equipment or products when the probability that an "old" component will operate at least t additional time units, given it is now functioning, is the same as the probability that a "new" component will operate at least t time units. Equipment subject to periodic maintenance and parts replacement often exhibits this property of "never growing old".

The exponential distribution is related to the Poisson probability distribution. In fact, it can be shown that if the number of arrivals at a service counter follows a Poisson probability distribution with the mean number of arrivals per unit of time equal to $\frac{1}{\beta}$.

Definition 9.4:

The continuous random variable *X* has an **exponential distribution**, withparameter β, if its density function is given by: f(x) = $\frac{e^{-x/\beta}}{\beta}$, x ≥ 0, $\beta \geq 0$.

Property 1: *Mean*: $E(X) = \mu = \int_0^\infty x \frac{e^{-x/\beta}}{\beta} dx = \beta$

Property 2: *Variance*: $Var(X) = E(X - E(x))^2 = \int_0^\infty x^2 \frac{e^{-x/2}}{2} dx - \beta^2 = \beta^2$

Property 3: *Moment Generating Function:*

$M_x(t) = E(e^{tx}) = \int_0^\infty e^{tx} \frac{e^{-x/\beta}}{\beta} dx = \frac{1}{(1-\beta t)}$ for $t < \frac{1}{\beta}$

Remark

- ✍ A key property possessed only by exponential random variables is that they are *memoryless*, in the sense that, for positive *s* and *t*, $P\{X > s + t | X > t\} = P\{X > s\}$. If *X* represents the life of an item, then the memoryless property states that, for any *t*, the remaining life of a *t*-year-old item has the same probability distribution as the life of a new item. Thus, one need not remember the age of an item to know its distribution of remaining life.

Example 9.7: Let X be an exponential random variable with pdf of : f(x) = $\frac{e^{-x/2}}{2}$, $x \geq 0$ then finf the mean and variance of the random variavle X.

Solution: is $E(X) = \mu = \int_0^\infty x \frac{e^{-x/2}}{2} dx = 2$ and $Var(X) = E(X - E(x))^2 = 4$.

Example 9.8: The probability density of X is $f(x) = \begin{cases} 3e^{-3x} for x > 0 \\ 0 \quad elsewhere \end{cases}$ then what is the mean and variance of this pdf?

Solution: this distribution is an exponential and the mean and variance it is obtain in the manner as: E(X) = $\int_0^\infty x\, 3e^{-3x} dx = 1/3$ and V(X) = $\int_0^\infty x^2\, 3e^{-3x} dx - (1/3)^2 = 1/9$.

ACTIVITY 9.3:

Assume X has an exponential distribution with parameter of $\lambda > 0$ and pdf f(x)= $a e^{-\lambda x}$, x > 0. Then find the value of a as f(x) is pdf and identify the value of x if $P(X \leq x) = ½$.

9.5 Chi-square Distribution

The chi-squared distribution plays a vital role in statistical inference. It hasconsiderable applications in both methodology and theory. The chi-squared distribution is animportant component of statistical hypothesis testing and estimation.Topics dealing with sampling distributions, analysis of variance, and nonparametricstatistics involve extensive use of the chi-squared distribution.

Characteristics of the χ^2 Distribution

- χ^2 values cannot be negative since.
- The χ^2 distribution is non-symmetric.
- For large values of n(usually greater than 30), the χ^2 distribution may beapproximated by the normal.
- The chi-square distribution contains only one parameter called the **degrees of freedom**
- The degrees of freedom when working with a single population variance is ν= n-1.

Definition 9.5:

The continuous random variable X has a **chi-squared distribution**, with v**degrees of freedom**, if its density function is given by:

$f(x; v) = \begin{cases} \frac{1}{2^{v/2}\Gamma(v/2)} x^{\frac{v}{2}-1} e^{\frac{-x}{2}}, x > 0 \\ 0 \qquad , elsewhere \end{cases}$, where v is a the degree of freedom in such a way that it is positive integer.

Property 1: ***Mean***: $E(X) = \mu = \int_0^\infty x \frac{1}{2^{v/2}\Gamma(v/2)} x^{\frac{v}{2}-1} e^{\frac{-x}{2}} dx = \boldsymbol{v}$

Property 2: ***Variance***: $Var(X) = E(X - E(x))^2 = \int_0^\infty x^2 \frac{1}{2^{v/2}\Gamma(v/2)} x^{\frac{v}{2}-1} e^{\frac{-x}{2}} dx - v^2 = \boldsymbol{2v}$

Property 3: ***Moment Generating Function:***

$M_x(t) = E(e^{tx}) = \int_0^\infty e^{tx} \frac{1}{2^{v/2}\Gamma(v/2)} x^{\frac{v}{2}-1} e^{\frac{-x}{2}} dx \quad = (\mathbf{1 - 2t})^{-\frac{v}{2}}$

Remarks

- If S^2 is the variance of a random sample of size n from a normal population with mean μ *and* var*iance* σ^2, then the random variable $\frac{(n-1)S^2}{\sigma^2}$ has χ^2 distribution with (n-1) degrees of freedom, and this is used in tests of variance.
- Consider n independent random variables with the standard normal distribution; call these variables Z_i, i = 1, 2, . . . , n. Then the statistic $X^2 = \sum_{i=1}^{n} Z_i^2$ is also a random variable having a chi-square distribution.

9.6 Student's t-Distribution

The probability distribution of *T* was first published in 1908 in a paper writtenby W. S. Gosset. At the time, Gosset was employed by an Irish brewery thatprohibited publication of research by members of its staff. To circumvent this restriction,he published his work secretly under the name "***Student***." Consequently,the distribution of *T* is usually called the Student *t*-distribution or simply the *t*-distribution.In deriving the equation of this distribution, Gosset assumed that thesamples were selected from a normal population. Although this would seem to be avery restrictive assumption, it can be shown that non-normal populations possessingnearly bell-shaped distributions will still provide values of T that approximatethe t-distribution very closely.In developing the sampling distribution of T,we shall assume that our random sample was selected from a normal population.

Theorem 9.1

Let $X_1, X_2, \ldots, X_n$be independent random variables that are all normal withmean μ and standard deviation σ. Let$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ and $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$, then the random variable $T = \frac{\bar{X}-\mu}{S/\sqrt{n}}$has a t-distribution with $v = n - 1$ degreesof freedom.From this theorem we have the following definition

Definition 9.6:

Let Z be a standard normal random variable and V a chi-squared random variablewith v degrees of freedom. If Z and V are independent, then the distribution ofthe random variable

T, where$T=\frac{Z}{\sqrt{V/v}}$,is given by the probability density function: $f(t) = \frac{\Gamma\left(\frac{v+1}{2}\right)}{\Gamma\left(\frac{v}{2}\right)\sqrt{\pi v}}\left(1+\frac{t^2}{v}\right)^{-\left(\frac{v+1}{2}\right)}$, $-\infty < t < \infty$is known as t-distribution with vdegreesof freedom.

Some characteristics of t-distribution

- It is symmetric about its mean
- It has a standard deviation and variance GREATERthan 1.
- There are actually many t distributions, one for each degree of freedom
- As the sample size increases, the t distribution approaches the normal distribution.
- It is bell shaped.
- The t-scores can be negative or positive, but the probabilities are always positive.

Property 1: ***Mean***: $E(X) = \mu = \int_{-\infty}^{\infty} x \frac{\Gamma\left(\frac{v+1}{2}\right)}{\Gamma\left(\frac{v}{2}\right)\sqrt{\pi v}}\left(1+\frac{t^2}{v}\right)^{-\left(\frac{v+1}{2}\right)} dx = 0$

Property 2: ***Variance***: $Var(X) = E(X - E(x))^2 = E(X^2) - [E(X)]^2 = \frac{v+1}{v-1}$, $v > 1$.

Property 3: ***Moment Generating Function:Does Not Exist***

Remark

- If the sample size is large enough, say $n \geq 30$, the distribution of T does not differ considerably from the standard normal. However, for n <30, it is useful to deal with the exact distribution of T. However, the t distribution has “fatter” tails than the normal.

- ✍ In view of its importance, the t distribution has been tabulated extensively. The t-Table at the end of this module contains values of $t_{\alpha,\nu}$, for α = 0.10, 0.05, 0.025, 0.01, 0.005, and ν = 1, 2, 3, …, 29 degrees of freedom; where $t_{\alpha,\nu}$ is such that the area to its right under the curve of the t distribution with ν degrees of freedom is equal to α.

Example 9.10: For a t-distribution with ν =19 *df*, find t_α values leaving an area of

a) 0.05 to the right; c) 0.10 to the left;

b) 0.975 to the right; d) half of α =0.01 on either side.

Solution; referring to Table B with ν =19 *df*, we have

a) $t_{0.05}$ =1.729; c) $t_{0.90} = -t_{0.10}$ =-1.328.

b) $t_{0.975} = -t_{0.025}$ =-2.093; d) $t_{\frac{\alpha}{2}} = t_{0.005} =$ =2.861; & $t_{0.095} = -2.861$.

Example 9.11: In 16 one-hour test runs, the gasoline consumption of an engine averaged 16.4 gallons with a standard deviation of 2.1 gallons. In order to test the claim that the average gasoline consumption of this engine is 12.0 gallons per hour, calculate the t value and $t_{\alpha,\nu}$, for α =0.05.

Solution: Substituting n=16, μ =12.0, $\overline{X}$ =16.4, and S=2.1 in the formula, we get

t= $\dfrac{\overline{X}-\mu}{S/\sqrt{n}} = \dfrac{16.4-12.0}{2.1/\sqrt{16}}$ =8.38; and the table value for ν =n-1=15 is $t_{0.05,15}$ =1.753.

9.7 F-Distribution

We have motivated the *t*-distribution in part by its application to problems in whichthere is comparative sampling (i.e., a comparison between two sample means).While it is of interest to let sample information shed lighton two population means, it is often the case that a comparison of variability isequally important, if not more so. The *F*-distribution finds enormous applicationin comparing sample variances. Applications of the *F*-distribution are found inproblems involving two or more samples.The statistic *F* is defined to be the ratio of two independent chi-squared randomvariables, each divided by its number of degrees of freedom. Hence, we can write*F* $=\frac{U/v_1}{V/v_2}$,where *U* and *V* are independent random variables having chi-squared distributionswith v_1 and v_2 degrees of freedom, respectively. We shall now state the samplingdistribution of *F*.

Definition 9.7:

Let U and V be two independent random variables having chi-squared distributionswith $v_1$and $v_2$degrees of freedom, respectively. Then the distribution of therandom variable $F = \frac{U/v_1}{V/v_2}$is given by the density function: $f(f) = \frac{\Gamma\left(\frac{v_1+v_2}{2}\right)\left(\frac{v_1}{v_2}\right)^{v_1/2}}{\Gamma\left(\frac{v_1}{2}\right)\Gamma\left(\frac{v_2}{2}\right)} \frac{f^{\left(\frac{v_1}{2}\right)-1}}{\left(1+\frac{v_1 f}{v_2}\right)^{\left(\frac{v_1+v_2}{2}\right)}}$, f > 0 is known as the ***F*-distribution** with $v_1$and $v_2$degrees of freedom.

Some characteristics of F-distribution

Since F is formed by chi-square, many of the characteristics in chi-square are also possessed by the F distribution.

- The F-values are all non-negative.
- The distribution is non-symmetric.
- There are two independent degrees of freedom, one for the numerator (v_1), and one for the denominator (v_2).
- A different table is needed for each combination of degrees of freedom.

Property 1: ***Mean***: $E(X) = \mu = \int_0^\infty x \frac{\Gamma\left(\frac{v_1+v_2}{2}\right)\left(\frac{v_1}{v_2}\right)^{v_1/2}}{\Gamma\left(\frac{v_1}{2}\right)\Gamma\left(\frac{v_2}{2}\right)} \frac{f^{\left(\frac{v_1}{2}\right)-1}}{\left(1+\frac{v_1 f}{v_2}\right)^{\left(\frac{v_1+v_2}{2}\right)}} dx = \frac{v_1}{v_1-2}$

Property 2: ***Variance***: $Var(X) = E(X^2) - [E(X)]^2 = \frac{2v_1^2(v_2+v_1-2)}{v_2(v_1-2)^2(v_1-4)}$, v_1***>4.***

Property 3: ***Moment Generating Function: Does Not Exist***

Remark

- The curve of the F-distribution depends not only on the two parameters $v_1$and $v_2$but also on the order in which we state them. Therefore, writing $f_\alpha(v_1, v_2)$ for $f_\alpha$with $v_1$and $v_2$degrees of freedom, we obtain $f_{1-\alpha}(v_1, v_2) = \frac{1}{f_\alpha(v_2, v_1)}$. Thus, for instance, the f-value with 6 and 10 degrees of freedom, leaving an area of 0.95 to the right, is $f0.05(6, 10) = \frac{1}{0.05(10,\ 6}$ = 0.246.
- If $S_1^2$and $S_2^2$are the variances of independent random samples of size $n_1$and $n_2$taken from normal populations with variances $\sigma_1^2 1$ and σ_2^2, respectively, then$F = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} = \frac{\sigma_2^2 S_1^2}{\sigma_1^2 S_2^2}$ has an

F- distribution with $v_1 = n_1 - 1$ and $v_2 = n_2 - 1$ degrees of freedom. Or suppose that a random variable X has a χ^2 distribution with n_1 degrees of freedom and a random variable Y has χ^2 distribution with n_2 degrees of freedom. Suppose also that these two chi-square variables are independent. Then the ratio of the two divided by their respective degrees of freedom is the F –Distribution. Thus, F-distribution is formed by the ratio of two independent chi-square variables divided by their respective degrees of freedom, i.e., $F_{(v_1, v_2)} = \frac{\chi_1^2(v_1)/v_1}{\chi_2^2(v_2)/v_2}$

SUMMARY

- A random variable X is said to be *uniform* over the interval (a, b) if its probability density function is given by: f(x) = $\frac{1}{b-a}$, $a \leq x \leq b$ and 0 elsewhere. Its expected value and variance are: $\frac{a+b}{2}$ and $\frac{(b-a)^2}{12}$, respectively.
- A random variable X is said to be *normal* with parameters μ and σ2 if its probabilitydensity function is given by: $f(x) = \frac{1}{\sigma\sqrt{2\Pi}} e^{\frac{-1}{2}(\frac{x-\mu}{\sigma})^2}, \infty < x < \infty; \sigma > 0 \text{ and } -\infty < \mu < \infty$, the parameters μ and σ^2 are its expected value and variance.

☞ If X is normal with mean μ and variance σ^2, then Z, defined by $Z = \frac{X-\mu}{\sigma}$ is normal with mean 0 and variance 1. Such a random variable is said to be a *standard* normal random variable.

☞ A random variable whose probability density function is of the form $f(x) = \frac{e^{-x/\beta}}{\beta}$, $x \geq 0$, $\beta \geq 0$ is said to be an *exponential* random variable with parameter λ. Its expected value and variance are, respectively, β and β^2.

☞ The continuous random variable X has a chi-squared distribution, with v degrees of freedom, if its density function is given by: $f(x;\ v) = \frac{1}{2^{v/2}\,\Gamma(v/2)} x^{\frac{v}{2}-1} e^{\frac{-x}{2}}, x > 0$ and 0 elsewhere. its mean and variance are v and $2v$.

☞ The random variable having probability density function of $f(t) = \frac{\Gamma\left(\frac{v+1}{2}\right)}{\Gamma\left(\frac{v}{2}\right)\sqrt{\pi v}}\left(1 + \frac{t^2}{v}\right)^{-\left(\frac{v+1}{2}\right)}$, $-\infty<t<\infty$ is known as t-distribution with v degrees of freedom. Its mean and variance are 0 and $\frac{\boldsymbol{v+1}}{\boldsymbol{v-1}}$, $\boldsymbol{v}>1$.

☞ The random variable having probability density function of $f(f) = \frac{\Gamma\left(\frac{v_1+v_2}{2}\right)\left(\frac{v_1}{v_2}\right)^{v_1/2}}{\Gamma\left(\frac{v_1}{2}\right)\Gamma\left(\frac{v_2}{2}\right)} \frac{f^{\left(\frac{v_1}{2}\right)-1}}{\left(1+\frac{v_1 f}{v_2}\right)^{\left(\frac{v_1+v_2}{2}\right)}}$, $f > 0$ is known as the F-distributionwith $v_1$and $v_2$degrees of freedom. Its mean and variance are $\frac{v_1}{v_1-2}$ and $\frac{2v_1^2(v_2+v_1-2)}{v_2(v_1-2)^2(v_1-4)}$, $v_1 > 4$.

EXERCISE 9.1

1. A bank manager has determined from experience that the time required for a security guard to make his rounds in a bank building is a random variable having anapproximately normal distribution with mean = 18.0 minutes and standard deviation =3.2 minutes. What are the probabilities that a security guard will complete his rounds of the bank building in: (a) less than 15 minutes (b) 15 to 20 minutes (c) more than 20 minutes
2. The owner of an automobile towing service company knows that the number of towingservice calls the company makes each day is a random variable having approximately a normal distribution with the mean 36.2 and the standard deviation 5.1. What are the probabilities that in any given day the company will make: (a) exactly 30 towing service calls (b) at most 30 towing service calls

3. A television station claims that its late evening news program regularly has 35 percent of the total viewing audience. If this claim is correct, what is the probability that among 500 late evening viewers, more than 200 will be watching the station's newsprogram?
4. The random variable X is normally distributed with mean 80 and standard deviation 12. Then (a) what is the probability that a value of X chosen at random will be between 65 and 95? (b) what is the probability that a value of X chosen at random will be less than 74?
5. The random variable X is normally distributed with mean 65 and standard deviation 15. Find x_o, such that $P(x \geq x_o) = .6738$.
6. The scores on a placement test have a mound-shaped distribution with mean 400 and standard deviation 45. (a) what percentage of people taking this exam will have scores of 310 or greater? (b) what percentage of the people taking this test will have scores between 445 and 490?
7. The amount of delay time for a given flight is exponentially distributed with a mean of 0.5 hour. Ten passengers on this flight need to take a subsequent connecting flight. The scheduled connection time is either 1 or 2 hours depending on the final destination. Suppose 3 and 7 passengers are associated with these connection times, respectively. (a) Suppose John is one of the 10 passengers needing a connection. What is the probability that he will miss his connection? (b) Suppose he met Mike on the plane and Mike also needs to make a connection. However, Mike is going to another destination and thus has a different connection time from John's. What is the probability that both John and Mike will miss their connections? (c) A friend of John's, named Mary, happens to live close to the airport where John makes his connection. She would like to take this opportunity to meet John at the airport. Suppose she has already waited for 30 minutes beyond John's scheduled arrival time. What is the probability that John will miss his connection so that they could have a leisurely dinner together? Assume John's scheduled connection time is 1 hour in this part (c).