STA 211
Probability II

Ibadan Distance Learning Centre Series

STA 211
Probability II

By
Amahia G. N. Ph.D.
Department of Statistics
University of Ibadan

Published by
Distance Learning Centre
University of Ibadan

First published in 2007

ISBN 978-021-271-X

General Editor: Prof. Francis Egbokhare

Series Editor: O.I. Adeyemo and O.O. Akisanya

Typeset @ Distance Learning Centre, University of Ibadan

Printed by: Aderoyal Prints Communication 63, Oyo Road Opposite Wema Bank, Sango, Ibadan

Table of Contents

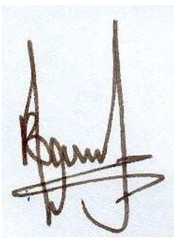
Vice-Chancellor's Message

I congratulate you on being part of the historic evolution of our Centre for External Studies into a Distance Learning Centre. The reinvigorated Centre, is building on a solid tradition of nearly twenty years of service to the Nigerian community in providing higher education to those who had hitherto been unable to benefit from it.

Distance Learning requires an environment in which learners themselves actively participate in constructing their own knowledge. They need to be able to access and interpret existing knowledge and in the process, become autonomous learners.

Consequently, our major goal is to provide full multi media mode of teaching/learning in which you will use not only print but also video, audio and electronic learning materials.

To this end, we have run two intensive workshops to produce a fresh batch of course materials in order to increase substantially the number of texts available to you. The authors made great efforts to include the latest information, knowledge and skills in the different disciplines and ensure that the materials are user-friendly. It is our hope that you will put them to the best use.



Professor Olufemi A. Bamiro, FNSE

Vice-Chancellor

Foreword

The University of Ibadan Distance Learning Programme has a vision of providing lifelong education for Nigerian citizens who for a variety of reasons have opted for the Distance Learning mode. In this way, it aims at democratizing education by ensuring access and equity.

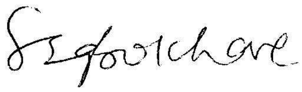
The U.I. experience in Distance Learning dates back to 1988 when the Centre for External Studies was established to cater mainly for upgrading the knowledge and skills of NCE teachers to a Bachelors degree in Education. Since then, it has gathered considerable experience in preparing and producing course materials for its programmes. The recent expansion of the programme to cover Agriculture and the need to review the existing materials have necessitated an accelerated process of course materials production. To this end, one major workshop was held in December 2006 which have resulted in a substantial increase in the number of course materials. The writing of the courses by a team of experts and rigorous peer review have ensured the maintenance of the University's high standards. The approach is not only to emphasize cognitive knowledge but also skills and humane values which are at the core of education, even in an ICT age.

The materials have had the input of experienced editors and illustrators who have ensured that they are accurate, current and learner friendly. They are specially written with distance learners in mind, since such people can often feel isolated from the community of learners. Adequate supplementary reading materials as well as other information sources are suggested in the course materials.

The Distance Learning Centre also envisages that regular students of tertiary institutions in Nigeria who are faced with a dearth of high quality textbooks will find these books very useful. We are therefore delighted to present these new titles to both our Distance Learning students and the University's regular students. We are confident that the books will be an invaluable resource to them.

We would like to thank all our authors, reviewers and production staff for the high quality of work.

Best wishes.



Professor Francis O. Egbokhare

Director

Preface To The Second Edition

The book has been thoroughly revised and enlarged. A number of changes have been made, particularly in lectures six and fourteen.

This edition retains one of the characteristic features of the book-enough exercises of all types and these exercises and the text form an integrated pattern. A special feature of these number of exercises in the lectures is that they have been specially constructed to illustrate the theory and are so designed that after doing them, the student should not only have a better grasp of the theory but should also know the motivation for the various steps. It is hoped that the book with these changes will be more useful to the students (and readers).

General Introduction

Probability is used to give a quantitative measure of the uncertainty associated with statements. The theory should be viewed as a conceptual structure whose conclusions rely on logic. The various concepts of probability are related to the physical world. We have clearly defined and developed these concepts like random variables, transformations, expected values, among others, one at a time, with sufficient elaboration. The topics which have been used to illustrate the theory have been so presented as to minimize peripheral descriptive materials and to concentrate on probabilistic content. It is expected that this approach would help the student to learn a variety of topics with ease.

Most students, brought up with a deterministic outlook find probability vague and difficult. In this course, the theory is developed axiomatically to remove the difficulties. The course will also afford you the opportunity of understanding the laws of probability, probability distributions and their applications, with emphasis placed on explanation, facility, and economy in the treatment. It is expected that this approach will give the student not only a working knowledge, but also an incentive for a deeper study of this fascinating subject.

This course is divided into three sections. In section one, we considered counting techniques. In order to handle the problem of counting points in complicated sample spaces, the counting techniques of permutations and combinations are used. Sections two and three deal with discrete and continuous probability distributions respectively.

Course Objectives

The objectives of the course include the following:

1. to understand the laws of probability;
2. to understand the concept of a probabilistic model for the mechanism generating a set of observed data, and the use of the model for calculating the probabilities of sample outcomes;
3. to equip you with the capacity to construct the sample space of various experimental outcomes with a view to helping you to develop the power of creative imagination; and
4. to offer you a sound foundation for statistical inference.

LECTURE ONE

Further Permutations and Combinations

Introduction

In many problems in probability, the number of points in the appropriate sample spaces is so great that efficient methods are needed to count them to arrive at the required probabilities. Such methods that are used in counting points in complicated sample spaces would be studied in this lecture. They include the techniques of permutations and combinations.

Objectives

At the end of this lecture, you should able to:

1. explain what the techniques of permutations and combinations are; and
2. discuss their functions in statistics.

Pre-Test

1. What do you understand as ‘permutation’ in statistics?
2. Briefly explain the technique of combination in Statistics

CONTENT

Permutations

In order to understand the counting technique of permutation, it is helpful to think in terms of objects which occur in groups. These groups may be characterized by type of object, the number belonging to each type, and the way in which the objects are arranged. Let us consider an example.

Example 1.1

Consider the letters a, b, c, d and e. There are five objects, one of each type. If we have the letters a, a, b, b and c, there are five objects; two of type a, two of b, and one of c. Furthermore, return to the first group of objects; abcde, bcdea, and cdeab differ in the order in which the five objects are arranged, but each of these groups contains the same number belonging to each type.

Example 1.2

Consider the number of different ways in which a, b, c may be written in a line. These are abc, acb, bac, bca, cab and cba. These are the six different permutations of three letters.

Example 1.3

Frequently, we are interested in finding the number of different permutations of n, different objects taking r (of them) at a time, (n >, r > o), both n and r are integers. The formation of these permutations is equivalent to writing in a line, in all possible orders, all the different groups of r letters, which may be chosen from n different letters. The number of such permutations is denoted ${}^{n}p_{r}$. The first letter may be chosen in n ways, since any one of the n may be chosen. This letter having been written down, we have to write alongside it, in all possible orders, all the different groups of (r – 1) letters which may be chosen from the remaining (n – 1) letters. This may be done in ${}^{n-1}P_{r-1}$ ways. It follows that,

$${}^{n}P_{r} = n.{}^{n-1}P_{r-1} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (1.1)$$

This result being true for all possible integral values of n and r, we have

$$\left.\begin{aligned} {}^{n-1}P_{r-1} &= (n-1).\ {}^{n-2}P_{r-2} \\ {}^{n-2}P_{r-2} &= (n-2).\ {}^{n-3}P_{r-3} \\ {}^{n-3}P_{r-3} &= (n-3).\ {}^{n-4}P_{r-4} \\ &\vdots \\ &\vdots \\ {}^{n-r+2}P_{2} &= (n-r+2).\ {}^{n-r+1}P_{1} \end{aligned}\right\} \ldots\ldots\ldots\ldots \quad (1.2)$$

Multiply equation (1.2) together and cancel the common factor to obtain

$${}^{n}P_{r} = n(n-1)(n-2)(n-3)\ldots\ldots(n-r+2).\ {}^{n-r+1}P_{1} \ldots\ldots\ldots\ldots (1.3)$$

Now ${}^{n-r+1}P_{1}$ in equation (1.3) is equal to (n – r + 1), since it denotes the number of ways of choosing one letter out of (n – r + 1) different letters. Hence from (1.3),

$${}^{n}P_{r} = n(n-1)(n-2)(n-3)\ldots\ldots(n-r+2)(n-r+1) \ldots\ldots\ldots\ldots (1.4)$$

Putting r = n in (1.4) we have

$${}^{n}P_{r} = n(n-1)(n-2)(n-3)\ldots\ldots 2.1 \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (1.5)$$

The expression on the right of equation (1.5) is the product of the first n positive integers. It is read "n factorial", with symbol n! or <u>In</u>. By definition 0! = 1.

Example 1.4

1.	5! = 5 x 4 x 3 x 2 x 1	=	120
	7! = 7 x 6 x 5 x … x 2 x 1	=	5,040
	10! = 10 x 9 x 8 x …. x 2 x 1	=	3,628,800

2. $^{12}P_6$ = 12 x 11 x 10 x 9 x 8 x 7 = 665,280

$^{20}P_8$ = 20 x 19 x 18 x 17 x 16 x 15 x 14 x 13 = 5,079,110,400

We can now generalize this procedure to obtain a convenient formula for the number of permutations of n different objects taken r at a time. Multiply equation (1.4) by $(n-r)!/(n-r)!$ to obtain

$$^{n}P_{r} = \frac{n(n-1)(n-2)(n-3)......(n-r+1)(n-r)!}{(n-r)!}$$

$$^{n}P_{r} = \frac{n!}{(n-r)!} \quad \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \quad (1.6)$$

Using equation (1.6) we have

1. $$^{12}P_{6} = \frac{12!}{6!} = \frac{12 \text{ x } 11 \text{ x } 10 \text{ x } 9 \text{ x } 8 \text{ x } 7 \text{ x } 6!}{6!}$$

= 12 x 11 x 10 x 9 x 8 x 7 = 665,280

2. $$^{20}P_{8} = \frac{20!}{12!} = 5{,}079{,}110{,}400$$

A special case of the formula in (1.6) for permutation occurs when all of the n objects are considered together.

That is,

$$^{n}P_{n} = \frac{n!}{(n-n)!} = \frac{n!}{0!} = \frac{n!}{1} = n!$$

Combinations

The order in which the objects are arranged is of importance in permutation; though it is not important in combination. Let $^{n}C_{r}$ denote the number of combinations of n different objects taken r at a time. To develop the formula for $^{n}C_{r}$, we need to consider the relationship between numbers of combinations and numbers of permutations for the same group of n objects taken r at a time.

A permutation is obtained by first selecting objects and then arranging them in some order, whereas a combination is obtained by performing only the first step. It follows that a permutation is obtained by taking every possible combination and arranging them in all possible ways. The total number of arrangement of r objects in r ways is r!. Thus

$^nP_r = {}^nC_r \cdot r!$

or

$$^nC_r = \frac{^nP_r}{r!} = \frac{n!}{r!(n-r)!} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (1.7)$$

Example 1.5

1. $^{12}C_6 = \frac{12!}{6!6!} = \frac{^{12}P_6}{6!}$

$= \frac{665280}{6!} = 924$

2. $^{20}C_8 = \frac{20!}{8!12!} = 125970$

Summary

In this lecture, we have discussed the concepts of permutation and combination. We illustrated how these two concepts are used in finding the number of points in appropriate sample spaces. Ability to arrive at the accurate number of points in sample spaces enhances the computation of required probabilities.

Post-Test

1. Determine the value of n in the equation $3\binom{n+1}{3} = 7\binom{n}{2}$

2. Express in factorial notation: (i) 6 x 5 x 4; (ii) (n + 2) (n + 1)n
   (iii) n(n – 1) …… (n – r + 1)

3. Evaluate (i) $\binom{n-1}{r-1} + \binom{n-1}{r}$; (ii) $\binom{n}{r-2} + 2\binom{n}{r-1} + \binom{n}{r}$.

LECTURE TWO

Permutations and Similar Elements

Introduction

In lecture one, our discussion of permutations and combinations pertained to groups of dissimilar element. In this lecture we now turn to the problem of determining the number of distinguishable arrangements that can be formed when some of the objects are identical.

Objectives

At the end of this lecture, you should be able to:

1. discuss permutations in relation to similar elements; and
2. compare and contrast the use of permutations with both similar and dissimilar elements.

Pre-Test

1. In how many ways can 9 people sit on a round table?
2. How many arrangements can be made with letters of LEATHERETTE?

CONTENT

We shall consider a simple illustration to aid in arriving at a formula for the number of different permutations that can be made of n objects, n_1 of type 1, n_2 of type 2, ….N_k of type k. Suppose there are four balls, two of which are black and two are white. If the balls are placed in a line, we would find six possible arrangements as follows:

BBWW, BWBW, BWWB, WWBB, WBWB, WBBW

= 6 arrangements …………………………………. (2.1)

If the balls had been of different colours, the number of possible permutations of the four balls taken four at a time would have been

$4P_4 = 4! = 24$ …………………………………. (2.2)

To determine the relationship between the results in equation (2.1), 6 and equation (2.2), 24, we consider one of the permutations of the two black balls and two white balls; namely,

BBWW

Suppose a number is printed on each of the balls to make it distinguishable from the other ball of the same colour. If the numbers 1 and 2 are printed on the balls, we can imagine the following arrangement:

B_1 B_2 W_1 W_2

The number of distinguishable arrangements that can be made of four balls, two of which are black and two are white, if the balls are placed in a line, is equal to:

$$6 \times 2! \times 2! = 24, \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (2.3)$$

since 2! arrangements can be made of the two black balls by permuting their subscripts while keeping the white balls unchanged, and 2! arrangements can be made of the two white balls keeping the black balls unchanged. It turns out that the relation between the 4! permutations when all balls were distinguishable and the six possible arrangements when there are two indistinguishable black and two indistinguishable white balls is

$$4! = 6 \times 2! \times 2! \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (2.4)$$

Therefore, if $P(n; n_1, n_2, \ldots, n_k)$ denotes the number of distinguishable arrangements that can be formed of n objects, taken n at a time, where n_1 are of type 1, n_2 of type 2,…, n_k of type k, and $n = n_1 + n_2 + \ldots + n_k$, we have the relationship,

$$P(n; n_1, n_2, ---, n_k) = \frac{n!}{n_1!, n_2!, ----, n_k!} \quad \ldots\ldots\ldots \quad (2.5)$$

A special case of the above result in (2.5) occurs when there are just two types of objects, as in our example of coloured balls. In this case, let x be of one type and (n – x) of the other type. Equation (2.5) gives us.

$$P(n, x, n-x) = \frac{n!}{x!(n-x)!} = \binom{n}{x} = {}^n\mathrm{C}_r$$

Thus, the number of different permutations that can be made of n objects, x of which are of one type and (n – x) of a second type, is equal to the number of combinations of n different objects, taken x at a time. It also follows that:

$$\binom{n}{x} = \frac{n!}{x!(n-x)!} = \binom{n}{n-x} \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (2.6)$$

The expression $\binom{n}{x}$ is often referred to as a binomial coefficient because of the way it appears in the binomial expansion discussed in lecture seven.

Example 2.1

Find the number of arrangements of the letters of the word "calculus" taken all at a time.

In the word, there are eight letters consisting of 2 c's, 1 a, 2 l's, 2 u's and 1 s.

The required number of arrangements is,

$$P(8;2,2,2,1,1) = \frac{8!}{2!\,2!\,2!\,1!\,1!} = \frac{40320}{8} = 5040$$

Example 2.2

A demographic survey questionnaire requires the respondent to answer each of the five successive questions with either a "yes" or "no". How many different possible responses are there? There are two possible responses for each question. Therefore, by the principle of multiplication there is 2 x 2 x 2 x ---- 2 = 2^5 = 32, responses.

Example 2.3

There are five students in a room. In how many ways can one or more of them leave the room?

Every student can be dealt with in two ways; he may leave the room or remain in the room. This gives, 2 x 2 x 2 x 2 x 2 = 2^5. This, however, includes the case in which all the students remain in the room, which is not permissible. Therefore, the required number = $2^5 - 1$ = 32 – 1 = 31 ways.

Summary

In this lecture we have examined the permutations of things which are not all different. We assume a situation in which we have n letters of which p are all alike (suppose they are all a's) and that the remaining (n – p) are different from a and from one another. The knowledge in this section helps up to determine the number of permutations of all the n letters.

Post-Test

1. In how many ways can 9 people sit on a round table?
2. A committee of 3 is to be chosen from 4 men and 3 women. If at least one man is to be included, how many possible selections are there?
3. How many arrangements can be made with the letters of LEATHERETTE?

LECTURER THREE

Probability Laws

Introduction

Probability is one of the fundamental tools of statistics. It had its formal beginnings with games of chance. A game of chance is one in which the outcome of a trial is uncertain. They include tossing a fair coin to determine the face that turns up and throwing a die to determine the number that turns up. It is recognized that though the outcome of any particular trial may be uncertain, there is a predictable long-term outcome. It is this long-term predictable regularity (called probability) that enables us to estimate probability. The term probability is used to give a quantitative measure to the uncertainty associated with our statements.

Objectives

At the end of this lecture, you should be able to:

1. clearly define the concept of probability; and
2. discuss the laws of probability.

Pre-Test

1. Two dices are thrown up, what is the probability of scoring either a double or a sum greater than 9?
2. compute (a) $P(A_1)$ (b) $P(A_{z)})$ (c) $P(A_1A_z)$

CONTENT

Views of Probability

If a random experiment can result in n mutually and equally likely trials (or outcomes) and if n_A of these trials have an attribute A, then the probability of A is n_A/n. This is called the classical view of probability.

Example 3.1

Suppose three fair coins are tossed once. The possible outcomes may be denoted by

{HHH, HHT, HTH, THH, HTT, THT, TTH, TTT}
w_1 w_2 w_3 w_4 w_5 w_6 w_7 w_8

If A is the event that at least two heads turn up,

$$A = \{w_1, \; w_2, \; w_3, \; w_4\}$$

$$P(A) = \frac{4}{8} = \frac{1}{2}$$

Example 3.2

Suppose that we roll a balanced die once. What is the probability of getting a five? The number of favourable outcomes is one since there is only one 5. The total number of possible outcomes is six; namely, (1, 2, 3, 4, 5, 6). Hence the probability of getting a five is $\frac{1}{6}$.

Another definition of probability is that of the relative frequency of the occurrence of an event in a large number of repetitions. If n is the number of trials and n(E) the number of occurrences of the event E, then the probability of E, denoted by P(E) is

$$P(E) = \lim_{n \to \infty} \frac{n(E)}{n}$$

Example 3.3

A and B take turns in throwing two fair dice, the first to throw a sum of 9 wins a price. If A has the first throw, compute the probability that A wins a price.

Solution

The sum of 9 can be made up in 4 ways: (3, 6). (4. 5), (6, 3), (5, 4) so that the probability of throwing the sum of 9 with two dice is 4/36. If A is to win, he should throw 9 in either the first, third, fifth, ……. throw.

The probability of A throwing 9 in the first throw $= \frac{1}{9}$

$$= \left(\frac{8}{9}\right)^2 . \frac{1}{9}, \quad \text{in the third throw}$$

$$= \left(\frac{8}{9}\right)^4 . \frac{1}{9}, \quad \text{in the fifth throw}$$

$$= \left(\frac{8}{9}\right)^{n-1} . \frac{1}{9}, \quad \text{in the n}^{\text{th}} \text{ throw}$$

Let S denote the sum of the probabilities of A.

$$S = \frac{1}{9} + \left(\frac{8}{9}\right)^2 . \frac{1}{9} + \left(\frac{8}{9}\right)^4 . \frac{1}{9} + ----- + \left(\frac{8}{9}\right)^{n-1} . \frac{1}{9}$$

$$= \frac{1}{9}\left(1 + r + r^2 + r^3 + ----- r^{n-1}\right)$$

where $r = \left(\frac{8}{9}\right)^2 . < 1$

This is a geometric series with a = 1 and common ratio = r. The sum to infinity is

$$S = \frac{1}{9}\left[\frac{a(1-r^n)}{1-r}\right] = \frac{1}{9}\left(\frac{a}{1-r}\right)$$

$$= \frac{1}{9}\left(\frac{1}{17/81}\right) = \frac{1}{9}\left(\frac{81}{17}\right)$$

$$= \frac{9}{17}$$

$\therefore$ Probability that A wins $= \frac{9}{17}$

You must have observed that there are situations which cannot conceivably fit into the framework of repeated outcomes under somewhat similar conditions. The third view of probability takes care of this drawback associated with the first and second views of probability. It is called the axiomatic or subjective view of probability, which is based on personal beliefs. For example, you may say that the probability that your friend, John would visit you is 70%.

In the axiomatic development of probability theory, probability is defined as a function defined on events (subsets of sample space, S), that is, it is a rule which associates to each event Aa certain real number P(A) which satisfies the following three axioms:

Axiom I: P(A) > O, i.e, the probability of every event is non-negative.

Axiom II: P(S) = 1, i.e, the probability of a certain event is unity.

Axiom III: If A_1, A_2, ----: are a countable number of sub-events of S such that

$A_1 \cap A_2 \cup_3$------ = S and $A_i \cap A_j = \phi$, $(i \neq j)$

then $P(A_i \cup A_j \cup_k \cup$ -----) = $P(A_i) + P(A_j) + P(A_k)$ + ----

i.e. the probability of a union of disjoint events is the sum of the probabilities of the events themselves.

Summary

The concept probability has been defined as a quantitative measure of the uncertainty associated with the statements we make. The quantitative measure can be computed by looking at the term probability from three perspectives; namely:

1. classical view of probability
2. relative frequency of the occurrence of an event in a large number of repetitions, and
3. axiomatic or subjective views of probability.

Problems in probability would usually fit into one of the frameworks above.

Post-Test

1. Two dices are thrown up. What is the probability of scoring either a double, or a sum greater than 9?
2. A bag contains 20 balls, 10 of which are red, 8 white and 2 blue. The balls are indistinguishable apart from colour. Two balls are drawn in succession, without replacement. What is the probability that they will both be red?
3. A basket contains two white balls W_1, and W_2 and three black balls B_1, B_2 and B_3 respectively. An experiment consists of selecting two balls at random without replacement. Let A_1 and A_2 denote the events: first ball is white and second ball is black respectively.

   Compute: (a) $P(A_1)$ (b) $P(A_2)$ (c) $P(A_1A_2)$

LECTURE FOUR

Conditional Probability, Independence and Bayes Rule

Introduction

We require that every possible outcome of a conceptual experiment under study can be enumerated. Here, an experiment is the process of taking measurement (or making an observation). Each conceivable outcome of the conceptual experiment under study is a sample point, and the totality of conceivable outcomes is the sample space. Sometimes, we restrict our attention to a subset of the sample space.

Objectives

At the end of this lecture, you should be able to:

1. explain what conditional probability is all about; and
2. discuss Independence and Bayes' Rule

Pre-Test

1. Two dices are rolled. What is the probability that the sum of the faces exceeds 8, given that one or more of the faces is a 6?
2. The content of three identical corns are:

   1 white, 2 red, 3 black balls

   3 white, 1 red, 3 black balls

   2 white, 3 red, 1 black balls

Conditional Probability

Given two events A and B, associated with a sample space, we want to define the conditional probability of event A given that event B has occurred. The probability of event A, given that another event B has occurred, is denoted by P(A/B) and is defined by

$$P(A/B) = \frac{P(AB)}{p(B)}, \text{ if } P(B) > O \quad \text{-------------------------} \ (4.1)$$

Example 4.1

In an experiment of tossing two fair coins once, compute the probability of two heads given a head on the first coin.

Solution

Sample space = Ω = {HH, HT, TH, TT}

Set A = Head on the first coin

B = Head on the second coin

We require

$$P(AB/A) = \frac{P(ABA)}{p(A)} = \frac{P(AB)}{p(A)}$$

$$= \frac{1/4}{1/2} = \frac{1}{2}$$

Independence of Two Events

If events A and B are independent, the conditional and unconditional probabilities are the same. That is, from equation 4.1, we find that

P(AB) = P(A) P(B/A), for conditional probability, and

P(AB) = P(A) P(B), if A and B are independent

Example 4.2

Suppose that we throw a fair die two times.

Let A = event that the first throw shows a 5

B = event that the second throw shows a 3

Clearly, A and B are independent events. Hence

$$P(AB) = P(A) . P(B) = \frac{1}{36}$$

Bayes Rule (or Theorem)

Bayes rule is based on conditional probability. We have:

$$P(A/B) = \frac{P(AB)}{P(B)} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (4.1)$$

also

$$P(B/A) = \frac{P(AB)}{P(A)} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (4.2)$$

$$\Rightarrow P(AB) = P(B/A) \ . \ P(A) \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (4.3)$$

Substituting from (4.3) in (4.1) we have

$$P(A/B) = \frac{P(B/A)P(A)}{P(B)} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (4.4)$$

This is called Bayes' rule or theorem

Let H_1, H_2, and D denote two hypothesis and the observed data, respectively. Suppose we substitute H_1 and H_2 in turn for A and D for B in (4.4).

Then we have

$$P(H_1/D) = \frac{P(D/H_1)P(H_1)}{P(D)} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (4.5)$$

$$P(H_2/D) = \frac{P(D/H_2)P(H_2)}{P(D)} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (4.5)$$

Hence

$$\frac{P(H_1/D)}{P(H_2/D)} = \frac{P(D/H_1)}{P(D/H_2)} \cdot \frac{P(H_1)}{P(H_2)} \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (4.6)$$

The L. H. S. of (4.6) is called posterior odds. The first term on the R. H. S. is called the likelihood ratio, and the second term is called the prior odds.

These odds play a significant role in the choice between probability models.

Example 4.3

The contents of three identical baskets B_i(i = 1, 2, 3) are:

B_1 : 4 red balls and 1 white ball

B_2 : 1 red ball and 4 white balls

B_3 : 2 red balls and 3 white balls.

A basket is selected at random and from it a ball is drawn. The ball drawn turns out to be red on inspection. What is the probability that it came from the first basket?

Solution

Let D = Data, the event of drawing a red ball.

Consider table 1 below.

Table 1: Probability of Conditional Event

(1) **State of Nature** **Bi**	**(2)** **P(Bi)**	**(3)** **P(D/Bi)**	**(4)** **P(Bi) P(D/Bi)**	**(5)** **P(Bi/D)**
B_1(4R, 1W)	$1/3$	$4/5$	$4/15$	$4/7$
B_2(1R, 4W)	$1/3$	$1/5$	$1/15$	$1/7$
B_3(2R, 3W)	$1/3$	$2/5$	$2/15$	$2/7$
Total	**1**	**-**	$7/15$	**1**

The required probability is $P(B_1/D) = 4/7$

Summary

The conditional probability of an event A assuming B, denoted by P(A / B) is by definition the ratio

$$P(A/B) = \frac{P(AB)}{P(B)} \qquad (4.7)$$

where we assume that P(B) is not zero from (4.7)

$$\begin{aligned} P(A\,B) &= P(B)\,P(A / B) \\ &= P(A)\,P(B / A) \qquad (4.8) \end{aligned}$$

If instead of (4.8) we have

P(AB) = P(B) P(A) (4.9

is an indication that events A and B are independent.

Again from (4.8)

$$P(A/B) = \frac{P(B/A)P(A)}{P(B)} \qquad (4.10)$$

The result in (4.10) is known as the Bayes theorem. The terms, a priori and a posteriori, are often used for the probabilities P(A) and P(A / B) respectively.

Post-Test

1. Two dices are rolled. What is the probability that the sum of the faces exceeds 8, given that one or more of the faces is a 6?
2. The content of three identical corns are:

   1 white, 2 red, 3 black balls

   3 white, 1 red, 3 black balls

   2 white, 3 red, 1 black balls

A corn is selected at random and from it two balls are drawn at random without replacement. The two balls are one red and 1 white. What is the probability that they came from the second corn?

LECTURE FIVE

Random Variables, Expectations and Moments

Introduction

In lecture four, we discussed conditioned probability, independence and Bayes' rule. In the present lecture, we shall examine random variables, mathematical expectations and moments.

Suppose Ω is the sample space with sample point w. Sometime we are interested in the value X(w) associated with w, and not in w itself. X(w) may be observable while w is not.

Objectives

At the end of this lecture, you should be able to:

1. define random variables, mathematical expectations and moments; and
2. discuss their significance in statistical analysis.

Pre-Test

Find the mean and variance, if they exist, of each of the distributions>

1. $7\ (x) = 2x^{-3}, 1<x<\infty$
2. $7\ (x) = 6x(1-x), 0<x<1$

CONTENT

Example 5.1

Suppose a fair coin is tossed three times. The sample is given by

Ω =	{HHH,	HHT,	HTH,	THH,	HTT,	THT,	TTH,	TTT}
	w_1	w_2	w_3	w_4	w_5	w_6	w_7	w_8

We may be interested only in the number of times head turns up. Thus, with the outcome HHH = w_1 we will associate a number $X(w_1) = 3$, representing the number of heads in w_1. With the outcome HHT = w_2 we will associate $X(w_2) = 2$, and so on.

A function X on a space Ω to a space Ω' assigns to each point $w \in \Omega$ a unique point in Ω' denoted by X(w). X(w) is the image of the argument w under X. It is also called the value of X at w. X is also a mapping from Ω to Ω', denoted by $\Omega \to \Omega'$. w is mapped on to $X(w) = w' \in \Omega'$ by X. X also establishes correspondence relation between points in Ω with points in Ω'. $\Omega$is called the domain of X and Ω' is called the range. X is a random variable. It assigns numerical values to each point x defined on the sample space.

Mathematical Expectations

Mathematical expectations are very useful in solving problems involving distributions. Let x be a random variable having p.d.f. f(x) and let u(x) be a function of X such that $\int_{-\infty}^{\infty} u(x) f(x) dx$ exits, if X is continuous and $\sum_{X} u(x) f(x)$ exits, if X is a discrete type random variable. The integral or sum is called the mathematical expectation (or expected value of u(x) and it is denoted by E[u(x)].

That is

E[u(x)] = $\int_{-\infty}^{\infty} u(x) f(x) dx$, for continuous random variable …… (5.1)

= $\sum_{X} u(x) f(x)$, for discrete random variable. ……………. (5.2)

Example 5.2

Let X have the p.d.f

f(x) = 2(1 – x), 0 < x < 1

The expected value of X is given by

$$E(X) = \int_{0}^{1} x f(x) dx = \int_{0}^{1} 2x(1-x) dx$$

$$= 2\left[\frac{x^2}{2} - \frac{x^3}{3}\right] = \frac{1}{3}$$

Clearly, if X is a random variable, the rth moment of X, denoted by u_r^1, is defined as

u_r^1 = E(X^r) ………………………………………………….. (5.3)

if the expectation exists

Note that $u_1^1 = E(X) = \mu_X$, the mean of X.

Moment Generating Function (M. G. F)

Let X be a random variable with density f(x). the expected value of e^{tX} is defined to be the moment generating function (m.g.f) of X if the expected value exits for every value of t in some interval – h < t < h; h > 0. The m.g.f function, denoted by $m_X(t)$ or m(t) is given by

$$m(t) = E(e^{tX}) = \sum_{X} e^{tX} f(x) \quad \text{…………………………………..} \quad (5.5)$$

if X is discrete.

If m.g.f exits, then m(t) is continuously differentiable in some neighborhood of the origin. If we differentiate the m.g.f r times with respect to t, we have

$$\frac{d^r}{dt^r} m(t) = \int_{-\infty}^{\infty} x^r e^{xt} f(x) dx \quad \text{………………………………….} \quad (5.6)$$

and letting t ⟶ 0, we find

$$\frac{d^r}{dt^r} m(0) = E(X^r) = u_r^1 \quad \text{…………………………………} \quad (5.7)$$

where the rth derivative of m(t) is evaluated as t ⟶ 0.
The m.g.f. is unique and completely determines the distribution of the random variable. If two random variables have the same m.g.f., they have the same distribution.

Example 5.3
Let X be a random variable with p.d.f given by

$$f(x) = xe^{-x}, \quad 0 < x < \infty.$$

$$m(t) = E[e^{tx}] = \int_{0}^{\infty} e^{tx} xe^{-x} dx$$

$$= \int_{0}^{\infty} xe^{-x(1-t)} dx = \int_{0}^{\infty} \left(\frac{y}{1-t} \right) e^{-y} \frac{dy}{(1-t)}$$

where $y = x^{(1-t)}$

$$= \frac{1}{(1-t)^2}\int_0^\infty ye^{-y}dy$$

$m(t) = (1-t)^{-2}$, using integration by parts.

$m'(t) = \frac{dm(t)}{dt} = 2(1-t)^{-3}$ have m(0) = 2 = E(X)

$m''(t) = 6(1-t)^{-4}$ so that $m''(0) = 6 = E(X^2)$

Variance of X = $\sigma^2 = E(X^2) - [E(X)]^2$

= 6 – 4 = 2

Summary

A random variable X is a function (or a process), which assigns a numerical value X(.) to every outcome of an experiment. The resulting function must satisfy the following two conditions:

1. The set $\{X \leq x\}$ is an event for every x.
2. The probabilities of the events $\{X = \infty\}$ and $\{X = -\infty\}$ equal zero.

The distribution function of a random variable X is the function.

$$F(x) = P(X \leq x) \tag{5.8}$$

defined for every x from $-\infty$ to ∞.

The deviation $$f(x) = \frac{d}{dx}F(x) \tag{5.9}$$

of F(x) is called the density function of the random variable X. The expected values, moments, and moment generating functions are obtained using the density functions.

Post-Test

1. Find the mean and variance, if they exist, of each of the following distributions:

   a. f(x) = $2x^{-3}$, $1 < x < \infty$

   b. f(x) = $6x(1 - x)$, $0 < x < 1$

2. A random variable can assume only the values 1 and 3. If its mean is $^{8}/_{3}$, find the variable.

LECTURE SIX

Chebyshev's Inequality

Introduction

We have developed the tools by which expected values and variances of random variables were computed in the preceding lecture. In order to utilize these quantities it must be possible to evaluate at least approximately, the probabilities of specified differences between the estimates and the quantities we wish to estimate. The Chebyshev inequality deals with the probabilities of a random variable to lie between specified limits. These will be found useful especially in connection with limit theorems (in Lecture Fourteen) and non-parametric inference. The Chebyshev inequality is also a useful theoretical tool as well as a relation that connects the variance of a distribution with the intuitive notion of dispersion in a distribution. We shall first take up a lemma from which the inequalities will follow.

Objectives

At the end of this lecture, you should be able to:

1. define the theorem of Chebyshe's inequality, and
2. discuss the theorem accordingly.

6.2 Basic Lemma

Lemma 6.2.1.*Let b denote a positive constant and* $\omega(x)$ *a nonnegative function. Then*

$$P[\omega(X) \geq b] \leq \frac{1}{b} E[\omega(X)] \qquad 6.1$$

provided that the expectation exists.

Proof;

Let A denote the set

$$A = \{x | \omega(x) \geq b\}$$

Then if 0 < P(A) < 1,

$$E[\omega(X)] = E[\omega(X)|A]P(A) + E[\omega(X)|A^c]P(A)$$

$$E[\omega(X)] = E[\omega(X)|A]P(A) \geq bP(A) \qquad 6.2$$

The first inequality in (6.2) follows because $\omega(x) \geq 0$ and the second follow because $\omega(X) \geq b$ (X) for $x$on A. If $P(A) = 0$, the basic lemma is trivially true; and if $P(A) = 1$, the second term in the expression for $E[\omega(X)]$ is missing, so that again the desired result follows.

With $\omega(x) = (x - \mu)^2$ and $c = \sqrt{b}$ the basic lemma reduces to the Chebyshev inequality.

6.3 Chebyshev Inequality

For any constant $c > 0$,

$$P(|X - \mu| \geq C) \leq \frac{\sigma^2}{C^2} \qquad 6.3$$

It may be noted that the same inequality can be expressed in two useful forms:

$$P(|X - \mu| < C) \geq 1 - \frac{\sigma^2}{C^2} \qquad 6.4$$

$$P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2} \qquad 6.5$$

Note equation (6.5) follows from

$$P[(X - \mu)^2 \geq k^2\sigma^2] \leq \frac{E[(X - \mu)^2]}{k^2\sigma^2} \qquad 6.6$$

So that

$$P[|X - \mu| \geq k\sigma] \leq \frac{1}{k^2} \qquad 6.7$$

or

$$P[|X - \mu| < k\sigma] \geq 1 - \frac{1}{k^2} \qquad 6.8$$

Example 6.1 If X *is a random variable such that* $E(X) = 3$ and $E(X^2) = 13$, *compute the lower bound for the probability* $P(-2 < X < 8)$.

Solution:

$$Var(X) = E(X^2) - [E(X)]^2 = 13 - 9 = 4$$

From equation (6.8)

$$P[|X - 3| < 2k] \geq 1 - \frac{1}{k^2} \qquad 6.9$$

That is,

$$P[3 - 2k < X < 3 + 2k] \geq 1 - \frac{1}{k^2} \qquad 6.10$$

The given inequality is

$$P(-2 < X < 8) \qquad 6.11$$

Comparing equation (6.10) and (6.11)

$$3 - 2k = -2 \qquad 6.12$$

or

$$3 + 2k = 8 \qquad 6.13$$

so that $k = 2.5$ and $1 - \frac{1}{k^2} = 0.84.$

The lower bound for the probability

$$P(-2 < X < 8) = 0.84 \ \ or\ 84\%$$

Example 6.2: *A random X has the density function*

$$f(x) = e^{-x}, \qquad X > 0$$

Find the lower bound for the probability $P[|X - \mu| < 2]$

Solution:

The mean of X is given by

$$E(X^2) = \int_0^\infty xe^{-x}dx = \ [xe^{-x} + e^{-x}]_\infty^0 = 1$$

Using integration by parts.

$$E(X^2) = \int_0^\infty x^2e^{-x}dx = \left[e^{-x}\left(x^2 + 2(x+1)\right)\right]_\infty^0 = 4$$

$$Var(X) = \sigma^2 = 4 - 1 = 3$$

We have from (6.8),

$$P\left[|X - 1| < \sqrt{3k^2}\right] \ \geq\ 1 - \frac{1}{k^2} \qquad 6.14$$

Comparing (6.14) with $P[|X - 1| < 2]$, we find that

$$k\sqrt{k} = 2 \quad or\ 3k^2\ 4$$

i.e, $k^2 = \frac{4}{3}$ so that

$$1 - \frac{1}{k^2} = \frac{1}{4}$$

Example 6.3 *A symmetric die is thrown* 360 *times. Determine the lower bound for the probability of getting* 50 *to* 70 *ones.*

Solution:

Let m denote the total number of success. Then

$$m = X_1 + X_2 + \cdots + X_{360}$$

$$\rightarrow E(m) = np = 360\left(\frac{1}{6}\right) = 60$$

$$Var(m) = npq = 360\left(\frac{1}{6}\right)\left(\frac{5}{6}\right) = 50$$

Chebyshev's inequality gives

$$P\left[|m - 60| < k\sqrt{50}\right] \geq 1 - \frac{1}{k^2}$$

$$\rightarrow \left[60 - k\sqrt{50} < m < 60 + k\sqrt{50}\right] \geq 1 - \frac{1}{k^2}$$

Since we are interested in the probability between 50 and 70,

$$60 - k\sqrt{50} = 50 \; or$$

$$60 + k\sqrt{50} = 70 \; or$$

$$k = \sqrt{2}$$

Putting $k = \sqrt{2}$

$$P\left[60 - \sqrt{2}\sqrt{50} < m < 60 + \sqrt{2}\sqrt{50}\right] \geq 1 - \frac{1}{2}$$

$$i.e, P[50 < m < 70] \geq \frac{1}{2}$$

Summary

A measure of the concentration of a random variable X neat its mean μ is its variance σ^2. The probability that X is outside an arbitrary interval $(\mu - x, \mu + c)$ is negligible if the ratio $\frac{\sigma}{c}$ is sufficiently small. This fundamental result is known as the Chebyshev inequality. Chebyshev inequalities are used to compare the lower and upper bounds for the probabilities of the difference between any random variable and a preassigned value, usually the mean.

Post-Test

1a. A symmetrical die is thrown 360 times. Determine the lower bound for the probability of getting 50 to 70 ones.

1b. Compute the required probability using the binomial distribution.

2. Two fair dice are tossed once. If X is the sum of the numbers showing up, compute the upper bound for the probability $P[|X - 7| \geq 3]$.

LECTURE SEVEN

The Bernoulli and the Binomial Probability Distributions

Introduction

The simplest experiment is one that may result in either of two mutually exclusive outcomes. Examples of such experiments include tossing a fair coin (head or tail), the outcome of a production process (good or defective), the sex of a baby to be born (male or female), etc. We will label the two outcomes of a Bernoulli trial success (s) and failure (f). The sample space is s = {s, f}.

Objectives

At the end of this lecture, you should be able to:

1. explain what Bernoulli; experiment is all about; and
2. discuss Binomial Probability distribution.

Pre-Test

1. Let X be b(2, p) and let y be b(4, p). If $p(X \geq 1) = \frac{5}{9}$, find $P(Y \geq 1)$.

2. The p.d.f of a random variable X is given by

$$f(x) = \binom{7}{x}\left(\frac{1}{2}\right)\left(\frac{1}{2}\right)^{7-x}, x = 0, 1, \ldots, 7.$$

Compute:

1. the m.g.f of X;
2. the mean and variance of X
3. $p(0 \leq X \leq 1)$.

CONTENT

The Bernoulli Experiment

Definition 7.1 A Bernoulli trial is an experiment with two mutually exclusive outcomes, success or failure.

Let S be the sample space for an experiment, and let ACS be any event with

$P(A) = p, \quad 0 < p < 1,$ and define

$$X(w) = \begin{cases} 1, & \text{if } w \in A \\ 0, & \text{if } w \in \overline{A} \end{cases}$$

The X is called the Bernoulli random variable with values 0 and 1 and parameter p. The probability distribution for X follows directly for S. Since X = 1 if and only if event A occurs. We have P(X = 1) = P(A) = P(A) = p, and since X = 0 if and only if event $\overline{A}$ occurs, it follows that P(X = 0) = P($\overline{A}$) = 1 – p = q.

We then have

$p_X(1) = p, \quad p_X(0) = q, \quad p + q = 1$

or

$p_X(x) = p^x q^{1-x}, \quad x = 0, 1, \quad p + q = 1$ …………… (7.1)

The mean of X is given by

$$E(X) = \sum_{X=0}^{1} x \; p(X = x)$$

$$= 0 . p(X = 0) + 1 . p(X = 1)$$

$$= 0 . q + 1 . p = p.$$

$$E(X^2) = \sum_{X=0}^{1} x^2 \; p(X = x)$$

$$= 0^2 . q + 1^2 p = p$$

$$Var(X) = p - p^2 = p(1 - p) = pq$$

The m.g.fof X is given by

$$m(t) = E[e^{tX}] = \sum_{X=0}^{1} e^{tX} \; p(X = x)$$

$$= e^0 . q + e^t . p = (q + pe^t)$$

$$= [(1 - p) + pe^t]$$ …………………………….. (7.2)

The Bernoulli random variable provides a convenient starting point for defining a Binomial random variable.

The Binomial Distribution

Definition 7.2: An experiment that consists of n (fixed) repeated independent Bernoulli trials, each with probability of success p, is called a binomial experiment with n trials and parameter p.

Definition 7.3: If X is a binomial random variable with parameters n and p, its probability distribution function is given by

$$f(x) = \binom{n}{x} p^x q^{n-x}, \qquad x = 0, 1, \ldots, n \qquad \ldots\ldots\ldots\ldots \quad (7.3)$$

The m.g.f of a binomial distribution is easily found. It is

$$m(t) = \sum_X e^{tx} f(x) = \sum_{x=0}^{n} e^{tx} \binom{n}{x} p^x (1-p)^{n-x}$$

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

$$= [(1-p) + pe^t]^n \qquad \ldots\ldots\ldots\ldots \quad (7.4)$$

for all real values of t. The mean and variance σ^2 of X may be computed from m(t). Since

$$m'(t) = n\left[(1-p) + pe^t\right]^{n-1}\left(pe^t\right)$$

and

$$m''(t) = n\left[(1-p) + pe^t\right]^{n-1}\left(pe^t\right) + n(n-1)\left[(1-p) + pe^t\right]^{n-2}(pe^t)^2,$$

it follows that

mean of X $= \mu = m'(0) = np$

and

Variance of X $= \sigma^2 = m''(0) - M^2$

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

$= \quad np(1 - p)$

Example 1

If the m.g.f of a random variable X is $\left(\frac{1}{3}+\frac{2}{3}e^t\right)^5$, find p(X = 2 or 3).

Solution

The p.d.fof X is

$$f(x) = \binom{5}{x}\left(\frac{2}{3}\right)^x\left(\frac{1}{3}\right)^{5-x}, \qquad x = 0,1,2,3,4,5$$

$$p(X = 2 \text{ or } 3) = \sum_{x=2}^{3} f(x)$$

$$= \sum_{x=2}^{3}\binom{5}{x}\left(\frac{2}{3}\right)^x\left(\frac{1}{3}\right)^{5-x}$$

$$= \binom{5}{2}\left(\frac{2}{3}\right)^2\left(\frac{1}{3}\right)^3 + \binom{5}{3}\left(\frac{2}{3}\right)^3\left(\frac{1}{3}\right)^2$$

$$= 10\left(\frac{4}{9}\right)\left(\frac{1}{27}\right)+10\left(\frac{8}{27}\right)\left(\frac{1}{9}\right)$$

$$= \frac{40}{9x27}+\frac{80}{9x27}=\frac{40}{81}$$

Example 2

If the m.g.f of a random variable X is $m(t) = \left(\frac{2}{3}+\frac{1}{3}e^t\right)^5$, find the mean and variance of X.

Solution

$$E(t) = m'(0) = \frac{5}{3}\left(\frac{2}{3} + \frac{1}{3}e^{0}\right)^{5} = \frac{5}{3}$$

$$\sigma^{2} \quad = \quad m''(0) - \mu^{2}$$

$$= \quad \frac{20}{9} + \frac{5}{3} - \frac{25}{9} \quad = \quad \frac{10}{9}$$

Summary

From our combinatorial analysis in lectures one and two, if a set has n elements, then the total number of its subsets consisting of k elements each equals $\binom{n}{k}$. We have used this result to find the probability of an event that occurs k times with constant probability p in n independent trials of an experiment. That is

$$p(k) = \binom{n}{k} p^{k}(1-p)^{n-k} \qquad (7.5)$$

When n = 1 and k = 0 or 1, we have

$$p'(k) = p^{k})1-p)^{1-k} \qquad (7.6)$$

Equations (7.5) and (7.6) are called the binomial and Bernoulli distributions respectively. We have used these distributions to derive the m.g.f. of a random variable X.

Post-Test

1. Let X be b(2, p) and let y be b(4, p). If p(X ≥ 1) = $\frac{5}{9}$, find P(Y ≥ 1).

2. The p.d.f of a random variable X is given by

$$f(x) = \binom{7}{x}\left(\frac{1}{2}\right)\left(\frac{1}{2}\right)^{7-x}, x = 0, 1, \ldots, 7.$$

Compute:

a. the m.g.f of X;
b. the mean and variance of X

c. $p(0 \le X \le 1)$.

LECTURE EIGHT

The Negative Binomial and the Geometric Distributions

Introduction

In the last lecture, we discussed Bernoulli and Binomial probability distribution. In this lecture, our focus will be on negative Binomial and Geometric distributions.

Objectives

At the end of this lecture, you should be able to:

1. discuss what negative binomial and geometric distribution are and;
2. compare and contrast the binomial probability distributions

Pre-Test

1. Explain the negative Binomial and the Geometric distributions
2. Discuss their relevance in statistical analysis.

CONTENT

Negative Binomial Distribution

You now have the necessary background for the study of negative binomial distribution after the treatment of binomial distribution. Consider a sequence of independent repetitions of a random experiment with constant probability p of success. Let the random variable Y denote the total number of failures in this sequence before the rth success. That is, (Y + r) is equal to the number of trials necessary to produce exactly r successes. Here r is a fixed positive integer. We are interested in determining the p.d.fof Y.

Let y be an element of {y |y = 0, 1, ….}. Then by the multiplication rule of probabilities, p(Y = y) = g(y) is equal to the product of the probability $\binom{y+r-1}{r-1}p^{r-1}q^{y}$ of obtaining exactly (r – 1) successes in the first (y + r – 1) trials and the probability p of a success on the (y + r)th trial. Hence, the p.d.f.g(y) of Y is given by

$$g(y) = \binom{y+r-1}{r-1}p^{r}q^{y}, \qquad y = 0, 1, \ldots\ldots$$

$$= \binom{y+r-1}{y} p^r q^y, \qquad y = 0, 1, \ldots\ldots$$

$$= 0, \text{ otherwise} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (8.1)$$

A distribution with a p.d.f. of the form g(y) in (8.1) is called a negative binomial distribution

The Mean and Variance of Y

The m.g.fof Y is given by

$$m(t) = E\left[e^{ty}\right] = \sum_{y=0}^{\infty} e^{ty} \binom{y+r-1}{y} p^r q^y$$

Note that $(1-x)^{-n} = \sum_{j=0}^{\infty} \binom{-n}{j} (-x)^j = \sum_{j=0}^{\infty} \binom{n+j-1}{j} x^j$ for $-1 < x < 1$

Hence

$$m(t) = \sum_{y=0}^{\infty} e^{ty} \binom{-r}{y} p^r (-q)^y$$

$$= \sum_{y=0}^{\infty} \binom{-r}{y} p^r (-qe^t)^y = p^r (1-qe^t)^{-r}$$

Now

$$m'(t) = rqe^t p^r (1-qe^t)^{-r-1}$$

$$E(Y) = \mu = m'(t)\big/_{t=0} = \frac{rq}{p}$$

$$E(Y^2) = m''(t) = rqe^t p^r (1-qet)^{-r-1} + (r+1)qe^t p^r (1-qe^t)^{-r-2} rqe^t$$

$$m''(t)\big/_{t=0} = \frac{rq}{p} + \frac{r^2q^2}{p^2} + \frac{rq^2}{p^2}$$

$$Var(Y) = m''(t)\big/_{t=0} - \frac{r^2q^2}{p^2} = \frac{rq^2}{p^2} + \frac{rq}{p} = \frac{rq}{p^2}$$

Geometric Distribution

For the negative binomial distribution, consider the particular case when r = 1, we get the probability distribution of Y, the number of failures preceding the first success. Thus, the distribution is given by

$$f(y) = pq^y, \quad y = 0, 1, 2, \ldots \qquad \ldots\ldots\ldots\ldots\ldots\ldots \quad (8.2)$$

$$= 0, \text{ otherwise}$$

Since the different terms in the probability distribution are the terms in a geometric series, the distribution is often called a geometric distribution.

The moments of this distribution may be obtained by substituting r = 1 in the corresponding moments of the negative binomial distribution.

We have

$$E(Y) = \mu = \frac{q}{p} \quad \text{and}$$

$$Var(Y) = \frac{q}{p^2} \quad \text{and}$$

$$m(t) = p(1 - qe^t)^{-1}$$

Example 8.1

If the m.g.f of a random variable X is $\frac{1}{9}\left[1-\frac{2}{3}e^t\right]^{-2}$, find the mean and variance of X.

Solution

$$m'(t) = \frac{1}{18}\left(\frac{2}{3}e^t\right)\left[1-\frac{2}{3}e^t\right]^{-3} + \left(\frac{3}{18}\right)\left(\frac{2}{3}e^t\right)^2\left[1-\frac{2}{3}e^t\right]^{-4}$$

Mean of X is

$$E(X) = \mu = m'(t)\Big/_{t=0} = 1$$

Variance of X is

$$\sigma^2 = m''(t)\big/_{t=0} - \mu^2$$

$= 4 - 1$

$= 3$

Summary

The negative binomial distribution

$$g(y) = \binom{y+r-1}{r-1} p^{rqy}, \qquad y = 0, 1, ... \qquad (8.3)$$

is modeled to compute the probability of y failures preceding the rth success. When interest is on just the first success, equation (8.3) becomes

$$g(y) = pq^y, \qquad y = 0, 1, \ldots\ldots \qquad (8.4)$$

Equation (8.4) is called the geometric distribution.

Post-Test

1. To investigate the effects of a certain drug on a given condition, a medical research team must first locate a person with the specified condition. Suppose 10% of the population has the given condition and the research team will interview people until they find k number of persons with the condition. Let X denote the number of people they must interview to locate k persons with the condition. Assume each person interviewed is a Bernoulli trial with probability of success p = 0.10 and that the trials are independent.

   Determine:

   a. The distribution of X.
   b. The mean and variance of X when k = 1.
   c. Comment on your values for (b).

2. John and James take turns in throwing two fair dices; the first to obtain a sum of 9 wins the sum of ₦1,000.00. If John has the first throw, compute:

   a. The probability that James wins the price at his first throw.
   b. The probability that John wins a price.
   c. James' expected value.

LECTURE NINE

The Poisson Distribution

Introduction

This lecture focuses on the discussion of the Poisson Distribution.

Objectives

At the end of this lecture, you should be able to:

1. discuss the Poisson distribution; and
2. state its significance in statistical analysis.

Pre-Test

1. What do you understand as the ‘Poisson distribution?
2. What is its significance in Statistical analysis?

CONTENT

Some of the applications of probability theory are concerned with modeling the time instants at which events occur. Examples of such events are telephone call arrival times at the switchboard of an establishment and customer arrival times at a department store.

Assume we would observe the phenomenon of interest for a period of time and the time instant at which we begin observing the phenomenon will be denoted by O. That is, the origin for our time scales. Furthermore, let us assume that we would observe the phenomenon of interest for a fixed time period of length t. ($t > O$). Clearly, the number of events that would occur in this fixed interval (O, t) is a random variable. Denote this random variable by X. Note that X is discrete since it is the number of events that would occur and its distribution would depend on the manner in which the events occur. Suppose in a sufficiently short length of time Δt, only 0 or 1 event can occur and the probability of exactly one event occurring in the short interval of length Δt is equal to $\lambda\Delta t$, where λ is a parameter. Finally, assume that any non-overlapping interval of length Δt are independent Bernoulli trials. These Bernoulli trials are the subdivisions of the interval of length t into n pieces, each, with probability of success equal to $\lambda\Delta t$.

i.e $\dfrac{t}{\Delta t} = n$

$$\Rightarrow \quad \frac{t}{n} \quad = \quad \Delta t$$

and p $= \quad \lambda\Delta t \quad = \quad \dfrac{\lambda \text{t}}{\text{n}}\left(q \quad = \quad 1 - \dfrac{\lambda \text{t}}{\text{n}}\right)$ ………………… (9.1)

It follows that X, the number of events in the interval of length t, is a binomial random variable. That is

$$f(k) \quad = \quad \binom{n}{k}\left(\frac{\lambda t}{n}\right)^k \left(1-\frac{\lambda t}{n}\right)^{n-k}$$

$$= \quad \frac{n!}{k!(n-k)!}\frac{(\lambda t)^k}{n^k}\left(1-\frac{\lambda t}{n}\right)^{-k}\left(1-\frac{\lambda t}{n}\right)^{n}$$

$$= \quad \frac{(\lambda t)^k}{k!}\left(1-\frac{\lambda t}{n}\right)^{h}\left(1-\frac{\lambda t}{n}\right)^{-k}\frac{n(n-1)(n-2).....(n-k+1)}{n^k}$$

$$\lim_{n \to \infty,} f(k) \left(1-\frac{\lambda t}{n}\right)^{n} \to e^{-\lambda t}, \left(1-\frac{\lambda t}{n}\right)^{-k} \to 1$$

$$\frac{n(n-1)(n-2).....(n-k+1)}{n^k} = 1\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)........\left(1-\frac{k+1}{n}\right) \to 1$$

Thus $f(k) \to \dfrac{(\lambda t)^k}{k!}e^{-\lambda t}$ …………………………………… (9.2)

for k = 0, 1, 2, …….

X is called a Poisson random variable with parameter λt = μ and f(k) is a Poisson p.d.f. i.e

$$f(x) = \frac{e^{-\mu}\mu^x}{x!}, \qquad x = 0,1,..........$$

The m.g.f of a Poisson distribution is given by

$$m(t) \quad = \quad \sum_{x=0}^{\infty} e^{tx} f(x) = \sum_{x=0}^{\infty} \frac{e^{tx} u^x e^{-u}}{x!}$$

$$= \quad e^{-u} \sum_{x=0}^{\infty} \frac{(ue^t)^x}{x!}$$

$$= \quad e^{-u} e^{ue^t} = e^{u(e^t - 1)}$$

for all real values of t

$$m'(t) = (ue^t) e^{u(e^t - 1)}$$

$$m''(t) = (ue^t)^2 e^{u(e^t - 1)} + (ue^t) e^{u(e^t - 1)}$$

$$m'(t) \big/_{t=0} \quad = \quad \mu$$

and

$$\sigma^2 \quad = \quad m''(t) \big/_{t=0} - \mu^2 \quad = \quad \mu^2 + \mu - \mu^2 \quad = \quad \mu$$

Thus the Poisson distribution has mean μ equal to the variance, σ^2,

Example 9.1

Suppose X has a Poisson distribution with $\mu = 2$. Then the p.d.f of X is

$$f(x) = \frac{2^x e^{-2}}{x!}, \qquad x \quad = \quad 0,\ 1,\ 2,\ ....$$

If we wish to compute $P(2 \leq X)$, we have

$$P(2 \leq X) \quad = \quad 1 - \{P(X = 0) + P(X = 1)\}$$

$$= \quad 1 - f(0) - f(1)$$

$$= \quad 1 - e^{-2} - 2e^{-2} \quad = \quad 1 - 3e^{-2}$$

$= \quad 0.594$

Example 9.2

If the m.g.f. of a random variable X is

$$m(t) \quad = \quad e^{4(e^t - 1)}$$

than X has a Poisson distribution with mean μ = 4. We have

$$f(x) = \frac{4^x e^{-4}}{x!}, \quad x = 0, 1, 2, \ldots..$$

$$P(X = 2) = \frac{4^2 e^{-4}}{2!}$$

$$= \quad 0.1465$$

Summary

A Poisson distributed random variable with parameter λt = μ takes the values 0, 1, ….. with probabilities.

$$f(x) = \frac{\mu^x e^{-u}}{x!}, \qquad x = 0, 1, \ldots\ldots \tag{9.3}$$

The Poisson distribution has mean equal to the variance. The moment generating function (m.g.f) is given by

$$m_{(t)} = e^{u\left(e^t - 1\right)} \tag{9.4}$$

Post – Test

1. A random variable has a Poisson distribution such that P(X = 1) = P(X = 2). Compute:
   a. The mean of X
   b. P (X = 3)

2. Let X have a Poisson distribution with $\mu = 100$. Use Chebyshev's inequality to determine a lower bound for $P(75 < X < 125)$.

LECTURE TEN

The Hypergeometric Distribution

Introduction

Our centre of focus in this lecture is on the Hypergeometric distribution.

Objectives

At the end of this lecture, you should be able to:

1. discuss the hypergeometric distribution; and
2. compare and contrast it with other forms of statistical distribution we have examined so far.

Pre-Test

1. Discuss the hypergeometric distribution.
2. Compare and contrast it with other forms of statistical distribution.

CONTENT

The binomial distribution was derived on the basis of n independent trials of an experiment. If the experiment consists of selecting individuals from a finite population of individuals, the trials will not be independent.

Random Sampling

Let N denote the size of a finite population from which a random sample of size n is drawn without replacement. Let the proportion of individuals in this finite population who possess a property of interest, say A, be denoted by p. Let X be a random variable corresponding to the number of individuals in the random sample who possess the property A. Our interest is to determine the p.d.f. of X. Observe that the expected number of individuals in the population who possess the property A is Np and those who do not possess the property would be equal to (N – Np.). It follows then that x individuals must come from Np and the remaining (n – x) individuals must come from (N – Np). Hence the desired density function is given by

$$f(x) = \frac{\binom{Np}{x}\binom{N-Np}{n-x}}{\binom{N}{n}}, x = 0, 1, \ldots, n. \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots.. \quad (10.1)$$

$=\quad 0,\ \text{otherwise}$

The Mean and the Variance of the Distribution

If X follows a hypergeometric distribution then

$$E(X) = \sum_{x=0}^{n} \frac{x\binom{K}{x}\binom{N-K}{n-x}}{\binom{N}{n}}, \text{ where } K = Np.$$

$$= n.\frac{K}{N}\sum_{x=0}^{n} \frac{\binom{K-1}{x-1}\binom{N-K}{n-x}}{\binom{N-1}{n-1}}$$

$$= n.\frac{K}{N}\sum_{y=0}^{n-1} \frac{\binom{K-1}{y}\binom{N-1-K+1}{n-1-y}}{\binom{N-1}{n-1}}, \qquad y = (x-1)$$

$$= n.\frac{K}{N}.\frac{\binom{N-1}{n-1}}{\binom{N-1}{n-1}} = \frac{nK}{N}$$

= np, since K = Np ………………………………. (10.2)

Using the relation $\sum_{i=0}^{m}\binom{a}{i}\binom{b}{m-i} = \binom{a+b}{m}$

$$E[x(x-1)] = \sum_{x=0}^{n} x(x-1)\frac{\binom{K}{x}\binom{N-K}{n-x}}{\binom{N}{n}}$$

$$= \quad n(n-1)\frac{K(K-1)}{N(N-1)}\sum_{x=2}^{n}\frac{\binom{K-2}{x-2}\binom{N-K}{n-x}}{\binom{N-2}{n-2}}$$

$$= \quad n(n-1)\frac{K(K-1)}{N(N-1)}\sum_{y=0}^{n-2}\frac{\binom{K-2}{y}\binom{N-2-K+2}{n-2-y}}{\binom{N-2}{n-2}}$$

$$= \quad \frac{n(n-1)K(K-1)}{N(N-1)}$$

Hence

$$Var(X) = E(X^2) - [E(X)]^2 = E[X(X-1)] + E(X) - [E(X)]^2$$

$$= \quad n(n-1)\frac{K(K-1)}{N(N-1)} + \frac{nK}{N} - \frac{n^2K^2}{N^2}$$

$$= \quad \frac{nK}{N}\left[(n-1)\frac{K-1}{N-1} + 1 - \frac{nK}{N}\right]$$

$$= \quad \frac{nK}{N}\left[\frac{(N-K)(N-n)}{N(N-1)}\right]$$

$$= \quad np\left[\frac{N(1-p)(N-n)}{N(N-1)}\right]$$

$$Var(X) = npq\left(\frac{N-n}{N-1}\right) \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (10.3)$$

Remark

The mean of the hypergeometric distribution coincides with the mean of the binomial distribution, and the variance of the hypergeometric distribution is (N – n)/(N – 1) times the variance of the binomial distribution. For large finite populations, the error arising from assuming that p is constant and the trials are independent, when the sampling population is very small and it may be ignored, in which case the binomial model is satisfactory. However, for populations which the sizes are small, a serious error will be introduced in using a binomial distribution. Therefore, it is necessary to apply a more appropriate distribution known as the hypergeometric distribution.

Example 10.1

A basket contains 6 white and 4 black balls and 3 balls are drawn without replacement. Compute the mean and the variance of the number of black balls that would be drawn.

Solution

Let X denote the number of black balls obtained

$$f(x) = \frac{\binom{4}{x}\binom{6}{3-x}}{\binom{10}{3}}, \quad x = 0, 1, 2, 3.$$

$$= 0, \text{ otherwise}$$

$$E(X) = \sum_{x=0}^{3} x f(x)$$

$$= \sum_{x=0}^{3} x \frac{\binom{4}{x}\binom{6}{3-x}}{\binom{10}{3}}$$

$$= \frac{1}{120}[0(20)+1(60)+2(36)+3(4)]$$

$$= \frac{1}{120}(144) = 1.20$$

$$E(X^2) = \sum_{x=0}^{3} x^2 f(x)$$

$$= \frac{1}{120}[0(20)+1^2(60)+2^2(36)+3^2(4)]$$

$$= \frac{1}{120}(240) = 2.0$$

$$\text{Var}(X) = E(X^2) - [E(X)]^2$$

$$= 2 - (1.2)^2$$

$$= 0.56$$

Example 10.2

A small rural community consists of 100 households, of whom 10 percent have malaria fever. Compute the probability of getting at most two households with malaria fever from a random sample of ten households from the community.

Solution

$$P(X \le 2) = \sum_{x=0}^{2} \frac{\binom{10}{x}\binom{90}{10-x}}{\binom{100}{10}}$$

$$= 0.94$$

Summary

The hypergeometric distribution is a probability model that assumes sampling without replacement. An error is introduced when constant probability is assumed when solving problems associated with the hypergeometric distribution. This error is small when the population size is relatively large.

Post-Test

1. A panel of 7 judges is to decide which of 2 final contestants (A and B) in a beauty contest will be declared the winner (based on a simple majority of the judges) Assume 4 of the judges will vote for A and the other 3 will vote for B. If 3 of the judges are randomly selected without replacement, what is the probability that a majority of them in the sample will favor A?
2. A box contains 10 white, 20 blue, 5 red, and 10 green balls. A sample of 10 balls is selected from the box at random, without replacement. If X is the number of white balls in the sample, determine:

   a. the probability distribution of X.
   b. the mean and variance of X
   c. $P(X > 8)$.

LECTURE ELEVEN

The Uniform or Rectangular Distribution

Introduction

In this lecture, we shall focus our discussion on the uniform or rectangular distribution.

Objectives

At the end of this lecture, you should be able to:

1. clearly define the uniform or rectangular distribution; and
2. discuss its significance in statistics analysis.

Pre-Test

1. Explain what the uniform or rectangular distribution is all about in statistical analysis.
2. Compare and contrast the uniform or rectangular distribution with other forms of statistical distribution we have discussed so far.

CONTENT

Perhaps the simplest continuous random variable is one whose distribution is constant over some interval (a, b) and zero elsewhere. The distribution arises, for example, in tossing a die or in the study of rounding errors when measurements are recorded to a certain accuracy. Thus, if measurements of patient's body temperatures are recorded to the nearest degree, it would be assumed that the difference in degrees between the true temperature of the patient and the recorded temperature is some number between - 0.50 and + 0.50 and that the error is uniformly distributed throughout this interval.

Uniform Random Variable

Consider an experiment in which one chooses at random a point from the closed interval $[a, b]$ that is on the real line. Thus the sample space $\Omega = [a, b]$. Let the random variable X be the identity function defined on Ω. This means that the space of $X = \Omega$. Suppose from the nature of the experiment, the probability that the observed value for X falls in any interval of length Δt in $[a, b]$ is proportional to Δt and $f(x) > 0$ for

$A < X < b$. Since the probability that the observed value for X lies in the internal is proportional to the length of the interval, f(x) must be a constant, K (say), for $a < X < b$. That is,

$$f(x) = K, \quad a < X < b. \qquad \ldots\ldots\ldots\ldots\ldots\ldots \qquad (11.1)$$

But the integral of f(x) over the whole range for X is

$$\int_a^b K dx = K(b-a) \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots. \quad (11.2)$$

Since the expression in (11.2) must equal 1, it must be that

$$K = \frac{1}{b-a} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots. \quad (11.3)$$

Substituting for K in (11.1) from (11.3), we have

$$f(x) = \frac{1}{b-a}, \qquad a < X < 1$$

$$= 0, \text{ elsewhere} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots.. \quad (11.4)$$

The Mean and Variance of a Uniform Random Variable

If X is a uniform random variable, then

$$E(X) = \int_a^b \frac{x}{b-a} dx = \frac{b^2 - a^2}{2(b-a)} = \frac{a+b}{2}$$

We also have

$$E(X^2) = \int_a^b \frac{x^2}{b-a} dx = \frac{b^3 - a^3}{3(b-a)} = \frac{1}{3}\left(a^2 + ab + b^2\right)$$

It follows that

$$Var(X) = E\left(X^2\right) - [E(X)]^2 = \frac{1}{2}(b-a)^2$$

Also

$$m(t) = E\left[e^{tx}\right] = \int_a^b \frac{e^{tx}}{b-a} = \frac{e^{tb} - e^{ta}}{t(b-a)} \text{ for } t \neq 0$$

Example 11.1

If f(x) = 1, 0 < X < 1

Find

a. mean and variance of X
b. mean and variance of X^2

Solution

a. $E(X) = \int_0^1 xf(x)dx = \int_0^1 xdx$

$= \left[\frac{x^2}{2}\right]_0^1 = \frac{1}{2}$

$E(X^2) = \int_0^1 x^2 dx = \left[\frac{x^3}{3}\right]_0^1$

$= \frac{1}{3}$

$Var(X) = \frac{1}{3} - \frac{1}{4} = \frac{1}{12}$

b. $E(X^2) = \frac{1}{3}$

$E(X^4) = \int_0^1 x^4 dx = \left[\frac{x^5}{5}\right]_0^1 = \frac{1}{5}$

$Var(X) = \frac{1}{5} - \frac{1}{9} = \frac{4}{45}$

Example 11.2

X is uniform on- 1, 5) and Y is negative binomial with p = $^2/_3$.

Find the number of successes r, such that $\sigma_X^2 = \sigma_Y^2$

Solution

$Var(X) = \frac{(b-a)^2}{12} = \frac{36}{12} = 3$

$Var(y) = \frac{rq}{p^2} = \frac{r(1/3)}{(2/3)^2} = 3$

$r = 4$

Summary

The uniform or rectangular distribution provides the simplest probability model for continuous random variables. If X is a uniform random variable, then

$$f(x) = \frac{1}{b-a}, \qquad a < x < b$$

The mean of X is given by

$$E(X) = \frac{1}{2}(a+b) \qquad (11.5)$$

This is simply the arithmetic average of the limits while the variance of X is

$$Var(X) = \frac{1}{12}(b-a)^2 \qquad (11.6)$$

Both the mean and variance are independent of n and N.

Post-Test

1. If X has the p.d.f. $f(x) = \frac{1}{4}$, $-1 < X < 3$, zero elsewhere, find the mean and variance of X^2 and X^3.

2. X is uniform on (1, 4) and Y is Poisson with parameter λ. Estimate λ such that $\sigma_X^2 = \sigma_Y^2$.

LECTURE TWELVE

Normal Distribution

Introduction

This lecture focuses on the discussion of the normal distribution in statistical analysis.

Objectives

At the end of this lecture, you should be able to:

1. distinguish between the normal distribution in statistical analysis and other terms of statistical distribution we have discussed; and
2. discuss the significance of the normal distribution in statistical analysis.

Pre-Test

1. What is the use of normal distribution in statistical analysis?
2. Differentiate between the normal distribution and other forms of statistical distribution.

CONTENT

The normal distribution has a unique position in probability theory, and can be used as an approximation to other distributions. In practice, normal theory can frequently be applied, with small risk of serious error, when substantially non-normal distributions correspond more closely to observed values. This allows u to take advantage of the elegant nature and extensive supporting numerical tables of normal theory.

Most arguments for the use of the normal distribution are based on forms of central limit theorems. These theorems state conditions under which the distribution of standardized sums of random variables tends to a unit normal distribution as the number of variables in the sum increases. That is, with conditions sufficient to ensure an asymptotic unit normal distribution.

Definitions

A random variable X is defined to be normally distributed if it has the p.d.f given by

$$f(x) = f(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)} \text{ .... } (12.1)$$

where the parameters μ and σ satisfy $-\infty<\mu<\infty$ and $\sigma > 0$

The p.d.fof $Z = (X - \mu)/\sigma$ is

$$f(Z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}Z^2} \text{ .... } (12.2)$$

which does not depend on the parameters μ and σ. This is called the standard form of normal distribution. The random variable Z is called a standard, or unit, normal variable.

Since

$$P(X \le x) = P[Z \le (x-\mu)/\sigma] \quad \ldots\ldots\ldots \quad (12.3)$$

such probability can be evaluated from tables of the cumulative distribution function of Z, which is,

$$\Phi(Z) = P[Z \le z] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{Z^2}{2}dz} \quad \ldots\ldots\ldots \quad (12.4)$$

The quantiles of the distribution are defined by $\Phi(Z_\alpha) = \alpha$ so that $Z_{1-\alpha}$ is the upper 100α percent point, and $Z_\alpha (= -Z_{1-\alpha})$ is the lower 100α percent point of the distribution.

The normal distribution is symmetrical about $X = \mu$. The p.d.f. has points of inflexion at $X = \mu \pm \sigma$. The distribution is unimodal with mode at $X = \mu$, which is also the median of the distribution.

The modal value of the p.d.f. is $1/\sqrt{2\pi}$ = 0.3979. The mean deviation of X is $\sigma\sqrt{2/\pi}$ = 0.798σ. For all normal distributions.

$$\frac{\text{mean deviation}}{\text{standard deviation}} = \sqrt{\frac{2}{\pi}} = 0.798 \quad \ldots\ldots\ldots \quad (12.5)$$

Approximations

The most common use of the normal distribution is as an approximation; either normality is ascribed to a distribution in the construction of a model, or a known distribution is replaced by a normal distribution with the same expected value and standard deviation. An example of such replacement is the normal approximation to the binomial.

Area under the Normal Curve

Suppose we replace the area under the normal curve by A; then

$$A = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2 dx} \quad \ldots\ldots\ldots \quad (12.6)$$

and on making the substitution Z = (x - μ)/σ, we find that

$$A = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-Z\frac{2}{2}dZ} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (12.7$$

To evaluate the integral A in (12.7) we note that A > 0 and that A^2 may be written

$$A^2 = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{1}{2}Z^2} dz \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{1}{2}y^2} dy$$

$$= \frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-\frac{1}{2}\left(Z^2+y^2\right)} dzdy$$

This integral can be evaluated by changing to polar coordinates by the substitution:
Y = r sin θ and Z = r cos θ, and the integral becomes

$$A^2 = \frac{1}{2\pi}\int_{0}^{\infty}\int_{0}^{2\pi} re^{-\frac{r^2}{2}} d\theta dr$$

$$= \int_{0}^{\infty} e^{-\frac{r^2}{2}} rdr = 1$$

The Moment Generating Function

$$m(t) = E\left[e^{tX}\right] = e^{t\mu} E\left[e^{t(X-\mu)}\right]$$

$$= e^{t\mu}\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}} e^{t(x-\nu)} e^{-\left(\frac{1}{2\sigma^2}\right)(x-\mu)^2} dx$$

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

If we complete the square inside the bracket, it becomes

$$(x - \mu)^2 - 2\sigma^2 t(x - \mu) = (x - \mu)^2 - 2\sigma^2 t(x - \mu) + \sigma^4 t^2 - \sigma^4 t^2$$

$= \quad (x - \mu - \sigma^2 t)^2 - \sigma^4 t^2$

and we have

$$m(t) = \left(e^{tu+\frac{1}{2}\sigma^2 t^2}\right)\left(\frac{1}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{\infty} e^{-\frac{1}{2\sigma^2}(x-\mu-\sigma^2 t)} dx\right)$$

The second function above is a normal distribution with mean $\mu + \sigma^2 t$ and variance σ_1^2

$$m(t) = e^{ut+\frac{1}{2}\sigma^2 t^2}$$

Now

$$m'(0) = \left(u+\sigma^2 t\right)e^{ut+\frac{1}{2}\sigma^2 t^2} \Big/ t = 0$$

$$= \mu$$

$$m''(0) = \left[\sigma^2 e^{ut+\frac{1}{2}\sigma^2 t^2} + \left(u+\sigma^2 t\right)^2 e^{ut+\frac{1}{2}\sigma^2 t^2}\right] \Big/ t = 0$$

$$= \sigma^2 + u^2$$

$$\text{Var}(X) = m''(0) - \mu^2$$

$$= \sigma^2$$

Remark 1

If $X \sim N(\mu, \sigma^2)$, then

$$P(a < X < b) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)$$

Remark 2

$$\Phi(x) = 1 - \Phi(-x)$$

Remark 3

If the random variable X is N(μ, σ^2), $\sigma^2 > 0$, then

$$(X-\mu)^2/\sigma^2 \text{ is } \chi^2_{(1)}$$

Example 12.1

If X has the m.g.f m(t) = e^{2t+32t^2} then X has a normal distribution with mean μ = 2 and σ^2 = 64.

Example 12.2

Let X be n(μ, σ^2), Then

P(μ - 2σ < X < μ + 2σ)

$$= N\left(\frac{\mu+2\sigma-\mu}{\sigma}\right) - N\left(\frac{\mu-2\sigma-\mu}{\sigma}\right)$$

$$= \Phi(2) - \Phi(-2)$$

= 0.954

Summary

We have introduced you to the normal random variables, one of the most important and most commonly encountered continuous random variables. In this lecture, we discussed probability distribution and we have shown how the probability can be used.

Post-Test

1. Let X be n(u, σ^2) so that P(X < 89) = 0.90 and P(X < 94) = 0.95. Find μ and σ^2.
2. If e^{3t+8t^2} is the m.g.f of the random variable X, Find P(- 1 < X < 9).
3. Let X b n(5, 10). Find P[0.04 < $(X-5)^2$ < 38.4].
4. If X is n(2, 2). Find a number K such that
   a. P(X > - K) = 0.20
   b. P(X > K) = 0.10

LECTURE THIRTEEN

The Joint, Marginal and the Conditional Distributions

Introduction

In this lecture, we shall examine the joint, marginal and conditional distributions.

Objectives

At the end of this lecture, you should be able to:

1. explain the joint, marginal and conditional distributions; and
2. discuss their use in statistical analysis.

Pre-Test

1. Explain the joint, marginal and conditional distributions.
2. Compare and contrast them with the previous forms of statistical distribution examined so far.

CONTENT

We are often interested not just in one random variable but in the relationship among several random variables. A distribution function is called a joint p.d.f when more than one random variable is involved.

Marginal and Conditional Distributions

Let $f(x_1, x_2)$ be the p.d.f of two random variables X_1 and X_2. Thus $f(x_1, x_2)$ is the joint p.d.f of the random variables X_1 and X_2. Consider the event $a< X_1< b$,

$a < b$. This event can occur when and only when the event

$a < X_1< b, -\infty< X_2<\infty$ occurs.

Example 13.1

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

$$f(x_1, x_2) = \frac{1}{21}(x_1 + x_2), \quad x_1 = 1, 2, 3; \quad x_2 = 1, 2,$$

Then $P(X_1 = 3) = f(3, 1) + f(3, 2) = {}^{3}/_{7}$

$P(X_1 = 2) = f(2, 1) + f(2, 2) = {}^{1}/_{3}$

$P(X_1 = 1) = f(1, 1) + f(1, 2) = {}^5/_{21}$

$P(X_2 = 1) = f(1, 1) + f(2, 1) + f(3, 1) = {}^3/_7$

$P(X_2 = 2) = f(1, 2) + f(2, 2) + f(3, 2) = {}^4/_7$

The above calculations are displayed in the table below

Table showing joint p.d.fof X_1 and X_2

x_2 \ x_1	**1**	**2**	**$f(x_2)$**
1	2/21	3/21	5/21
2	3/21	4/21	7/21
3	4/21	5/21	9/21
$f(x_1)$	9/21	16/21	1

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

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

Similarly

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

Thus $P(X_1 = 1) = f(x_1, 1) = {}^5/_{21}$

$P(X_2 = 1) = f(x_2, 1) = {}^9/_{21}$

From the preceding results it follows that $f(x_1/x_2)$ is the distribution of X_1 given $X_2 = x_2$, so that the joint density can be written as the product of the marginal and conditional densities. That is

$$f(x_1, x_2) = f(x_1)\, f(x_2/x_1)$$
$$= f(x_2)\, f(x_1/x_2)$$

If $f(x_1, x_2) = f(x_1)f(x_2)$ for all X_1 and X_2, then X_1 and X_2 are said to be independent. It follows that

$$f(x_2/x_1) = \frac{f(x_1, x_2)}{f(x_1)}, \, f(x_1) > 0$$

Similarly

$$f(x_1/x_2) = \frac{f(x_1, x_2)}{f(x_2)}, \, f(x_2) > 0$$

and

$$\int_{-\infty}^{\infty} f(x_2/x_1) dx_2 = \int_{-\infty}^{\infty} \frac{f(x_1, x_2)}{f(x_1)} dx_2$$

$$= \frac{1}{f(x_1)} \int_{-\infty}^{\infty} f(x_1, x_2) dx_2$$

$$= \frac{1}{f(x_1)} f(x_1) = 1$$

This shows that $f(x_2/x_1)$ has the properties of a p.d.f. of one continuous type random variable. It is called the conditional p.d.f. of the continuous random variable X_2, given the continuous, random variable X_1 has the value x_1.

Example 13.2

Let X_1 and X_2 have the joint p.d.f.

$$f(x_1, x_2) = 2, \quad 0 < x_1 < x_2 < 1$$
$$= 0 \text{ elsewhere}$$

The marginal p.d.f. of X_1 is

$$f(x_1) = \int_{x_1}^{1} 2dx_2 = 2[x]_{x_1}^{1}$$
$$= 2(1 - x_1), \quad 0 < x_1 < 1$$

Similarly that of X_2 is

$$f(x_2) = \int_{0}^{x_2} 2dx_1 = 2[x]_{0}^{x_2}$$
$$= 2x_2, \quad 0 < x_2 < 1$$

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

$$f(x_1/x_2) = \frac{f(x_1, x_2)}{f(x_2)}$$

$$= \frac{2}{2x_2} = \frac{1}{x_2}, \quad 0 < x_1 < x_2; \; 0 < x_2 < 1$$

$$= 0, \text{ elsewhere}$$

The conditional mean of X_1, given $X_2 = x_2$ is

$$E[X_1/X_2 = x_2] = \int_{-\infty}^{\infty} x_1 f(x_1/x_2) dx_1$$

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

$$= \left(\frac{1}{x_2}\right) \int_0^{x_2} x_1 dx_1 = \left(\frac{1}{x_2}\right) \left[\frac{x_1^2}{2}\right]_0^{x_2}$$

$$= \left(\frac{1}{x_2}\right) \left(\frac{x_1^2}{2}\right)$$

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

and the conditional variance of X_1,, given $X_2 = x_2$ is

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

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

$$= 0 \text{ elsewhere}$$

Summary

The probability models we considered in the previous lectures involved only one random variable. In this lecture, we have expanded the idea to include the joint distribution of two random variables. We have showed how the individual probability distributions (marginals) of the random variables can be derived from the joint probability distribution.

Post-Test

1. Let X_1 and X_2 have the joint p.d.f.

   $f(x_1, x_2) = x_1 + x_2, \; 0 < x_1 < 1, \; 0 < x_2 < 1$

   Find the conditional mean and variance of X_2, given $X_1 = x_1, \; 0 < x_1 < 1$.

2. Let the random variables X and Y have the joint p.d.f

   $f(x, y) = x + y, \; 0 < x < 1; \; 0 < y < 1.$

   Compute;

   a. the mean of Y and X
   b. the variance of Y and X, and
   c. the covariance of X and Y.

LECTURE FOURTEEN

Distribution of Functions of Random Variables

Objectives

At the end of this lecture, you should be able to:

1. discuss 'distribution of functions of random variables'; and
2. explain its relevance in statistical analysis.

Pre-Test

1. What do you understand as 'distribution of functions of random variables'?
2. What is the relevance of this in statistical analysis?

Introduction

4.1 Univariate Distribution

The procedure for obtaining a distribution that eliminates the necessity for estimating unknown parameters is based on a change of variable technique. The simplest of these arises when one wishes to apply normal distribution theory to a problem but discovers that the random variable X involved does not possess a normal distribution. We recall that the distribution function of a random variable$Y$, denoted by$G(y)$, satisfies the relations

$$G(t) = P\{Y \leq t\} = P\{\omega(X) \leq t\} \qquad 14.1$$

When t is any desired value. The inequality $\omega(X) \leq \_$ t can be expressed as an inequality on X. The relationship between y and x is such that there is a unique value of x to each value of y. Let the value of x corresponding to the value of t for y be denoted by r. Consequently, since $\omega(x) \leq t$ if, and only if, $x \leq t$, it follows that

$$P\{\omega(X) \leq t\} = P\{X \leq r\} = \int_{-\infty}^{r} f(x)dx \qquad 14.2$$

Thus from (14.1)

$$G(t) = \int_{-\infty}^{r} f(x)dx \qquad 14.3$$

In view of the fact that r is a function of t, it follows from the calculus formula for differentiating an integral with respect to its upper limit that

$$\frac{dG(t)}{dt} = \frac{dG(t)}{dr} \cdot \frac{dr}{dt} = f(r)\frac{dr}{dt} \qquad 14.4$$

This formula is valid at any point r where $f(r)$ is continuous. It will be assumed that $f(x)$is a continuous function of x. Since t and r were any pair of corresponding values of y and x, respectively, and were introduced to keep from confusing upper limit variables with dummy variables of integration, this relationship may be written

$$\frac{dG(y)}{dy} = f(x)\frac{dx}{dy} \qquad 14.5$$

But in view of the relationship between a distribution function and its density function, the left side of (14.5) is the density of Y . Hence the desired formula is

$$g(y) = f(x)\frac{dy}{dx} \qquad 14.6$$

This derivation for a change of variable $Y = \omega(X)$ in which $\omega(x)$ is a decreasing function will give a negative value. Since $\frac{dx}{dy}$ will be negative in this case, a formula that is valid is given by

$$g(y) = f(x)\left|\frac{dx}{dy}\right| \qquad 14.7$$

$\left|\frac{dx}{dy}\right| = \omega'(y) = J$,(Say) is referred to as the Jacobian of the inverse transformation of
x = ω (y).

Before formula (14.7) can be applied, it is necessary to replace x in f(x) by its value in terms of y, which means that it is necessary to solve the relation $y = h(x)$ for x in terms of y. One can calculate $\omega'(y)$ from this inverse relationship, or else calculate
$\omega'(y)$from the original relationship $y = h(x)$ and take its reciprocal.

Example 14.1 *Given the distribution*

$$f(x) = e^{-x}, \quad x > 0$$

Find the density function of the variable$Y = X^2$

Solution:

Since $y = x^2$ is an increasing function of x when$x > 0$, the result in equation (14.17) can be applied.

Hence $x = \omega(y) = y^{\frac{1}{2}}$ and $\omega'(y) = \frac{1}{2\sqrt{y}.}$

That is, $\omega'(y) = \frac{1}{2}y^{-\frac{1}{2}}, y > 0$

$$g(y) = f[\omega(y)]|\omega'(y)|$$
$$= e^{-y^{\frac{1}{2}}}\left|\frac{1}{2\sqrt{y}}\right|, y > 0$$
$$= \frac{1}{2}e^{-\sqrt{y}y^{\frac{1}{2}}}$$
$$= \frac{1}{2}y^{-\frac{1}{2}}e^{-y^{\frac{1}{2}}}, y > 0$$
$$= 0, otherwise$$

Example 14.2 let *X be a random variable having p.d.f.*

$$f(x) = 2x, 0 < X < 1$$

Find the density of $8X^3$

Solution:

The inverse function of $Y = 8X^3$is

$$x = \frac{1}{2}y^{\frac{1}{3}=\omega(y)} \qquad and$$
$$\omega'(y) = \frac{1}{6}y^{-\frac{2}{3}}, \qquad 0 < y < 8$$

The require density of Y is given by

$$g(y) = 2[\omega(y)]|\omega'(y)|$$

$$= 2\left(\frac{1}{2}y^{\frac{1}{3}}\right)\left|\frac{1}{6}y^{-\frac{2}{3}}\right|$$
$$= \frac{1}{6}y^{-\frac{1}{3}}, \quad 0 < y < \infty$$
$$0, \ otherwise$$

Let X have the uniform density over the interval $(0,1)$ with distribution
$$f(x) = 1, \quad 0 < x < 1$$
Find the density function of $y = -2\log_e X$.
Solution:
We have $Y = -2\log_e x$or $Y = \log_e \frac{1}{x^2}$,
That is, $x = \omega(y) = e^{-\frac{y}{2}}$
$$|\omega'(y)| = \frac{1}{2}e^{-\frac{y}{2}}, \qquad 0 < y < \infty$$
The distribution of Y is given by
$$g(y) = [\omega(y)]|\omega'(y)|$$
Now
$$f[\omega(y)] = 1$$
Hence
$$g(y) = 1 \cdot \frac{1}{2}e^{-\frac{y}{2}}, \qquad 0 < y < \infty$$
$$= 0, \qquad otherwise$$
14.2 Bivariate Transformation
The transformations of the variates x and y for bivariate distributions follow the ordinary laws for the transformation of differentials.
For example, if
$$dF = f(x,y)dxdv,$$
$$x = x(u,v),$$
$$y = y(u,v),$$
where$u$ and v are functions of x and y.
We have
$$dF = f[x(u,v), y(u,v)]|J|dudv$$
where$J$ is the Jacobian of the transformation given by
$$J = \frac{\partial(x,y)}{\partial(u,v)} = \begin{vmatrix} \frac{\partial x}{\partial y} & \frac{\partial y}{\partial u} \\ \frac{\partial x}{\partial v} & \frac{\partial y}{\partial v} \end{vmatrix}$$
Example 14.4: *Let* X_1 *and* X_2 *denote a random sample of size two from the distribution*
$$f(x) = e^{-x}, \qquad 0 < x < \infty$$
If $Y_1 = X_1 + X_2$and $Y_2 \frac{X_1}{X_1+X_2}$, show that $Y_1 and\ Y_2$ are stochastically independent.
Solution:
The joint p.d.f. of $X_1$and X_2 is
$$g(x_{1,}x_2) = f(x_1)f(x_2) = e^{-x_1-x_2}, 0 < x_1 < \infty; 0 < x_2 < \infty$$

$$= 0, otherwise$$

Now

$$y_1 = u_1(x_1, x_2) = x_1 + x_2$$
$$y_2 = u_2(x_1, x_2) = \frac{x_1}{x_1 + x_2}$$

These transformations can be written as; $x_1 = y_1 y_2$ and $x_2 = y_1(1 - y_2)$ so, that

$$J = \begin{vmatrix} y_2 & y_1 \\ 1 - y_2 & -y_1 \end{vmatrix}$$

Where $\{0 < y_1 < \infty;\ 0 < y_2 < 0\}$ in the $y_1 y_2$ plane. The joint p.d.f of Y_1 and Y_2 is given by

$$g(y_1, y_2) = y_1 e^{-y_1 y_2 - y_1(1 - y_2)}$$
$$= y_1 e^{-y_1}, \qquad 0 < y_1 < \infty;\ 0 < y_2 < 0$$
$$= 0, \qquad otherwise$$

The marginal distribution of Y_1 is

$$g(y_1) = \int_0^1 g(y_1, y_2) dy_2$$
$$= \int_0^1 y_1 e^{-y_1} dy_2$$
$$y_1 e^{-y_1} \int_0^1 1\ dy_2 = y_1 e^{-y_1} .1$$
$$= y_1 e^{-y_1} 0 < y_1 < \infty;$$
$$= 0, \qquad otherwise$$

The marginal distribution of Y_2 is

$$g(y_2) = \int_0^\infty g(y_1, y_2) dy_1$$
$$= \int_0^\infty y_1 e^{-y_1} dy_1$$

Noting that $\Gamma(\alpha) = \int_0^\infty y^{\alpha-1} e^{-y} dy$,it follows that

$$g(y_2) = \Gamma(2) = 1$$

Therefore,

$$g(y_2) = 1, \qquad 0 < y_2 < 1$$

$$= 0, \qquad otherwise$$

Since $g(y_1, y_2) = g(y_1)g(y_2)$ it follows Y_1 and $Y_2$are stochastically independent.

14.3 Limit Theorems

The theory of large samples in statistical inference makes us of certain results concerning the limiting behaviour of sequences of random variables and probability distributions.

Consider an infinite sequence of random variables defined on a probability space: $Y_1, Y_2, \cdots$ (This space may be taken to be the space of infinite sequence of real numbers, in which case the $Y's$ are coordinate functions). It is desirable to define what might be meant by the limit of the sequence. But random variables are functions, and there are several ways in which a given sequence of functions might be considered to approach a limit function.

The sequence $\{Y_n\}$ is said to converge to Y in distribution if and only if each point λ where $F_Y(\lambda)$ is continuous.

$$\lim_{n\to\infty} F_{Y_n}(X) = F_Y(\lambda)$$

This implies that for large but finite n, the probability $P(Y_n \leq \lambda)$ can be approximated by the probability $F_Y(\lambda)$ which may be considerably simpler to derive.

The sequence $\{Y_n\}$ is said to converge to Y in the mean, or in quadratic mean, or mean square, if and only if the average squared difference tends to zero:

$$\lim_{n\to\infty} E[(Y_n - Y)^2] = 0$$

The sequence Y_n is said to converge to Y in probability, if and only if for any $\varepsilon > 0$,

$$\lim_{n\to\infty} P(|Y_n - Y| \geq \varepsilon) = 0$$

The sequence $Y_n$converges to Y almost surely, or with probability 1, if and only if

$$P\left(\lim_{n\to\infty} Y_n = Y\right) = 1$$

14.4 Laws of Large Numbers

A functional notion in the formulation of a probability model is that the probability of an event is intended to embody the observed phenomenon of long-run stability in the relative frequency of occurrence of the event in a sequence of trials of the experiment. It is then of interest to determine whether stability is a mathematical consequence of the axioms in the model that has been developed. That is so is a result referred to as a law of large numbers. Ina particular case of trials in which an event A occurs with probability p and does not occur with probability $(1-p)$, the law of large numbers would assert that

$$\lim_{n\to\infty} \frac{Y}{n} = p$$

where$Y$ denotes the number of times in n trials that A occurs (frequency of the event A). If the trials are independent experiments, the relative frequency of A does tend toward the probability of A in the mean, in probability, and in almost surely.

14.5 The Central Limit Theorem

It is a remarkable fact that the random variable Y_n defined as the arithmetic mean of a sequence of *n*-independent replicas (in distribution) of a random variable X, has a distribution whose shape tends to a limiting shape that is independent of the distribution of X, so long as X has a finite variance, as *n* becomes infinite.

T study this limiting shape, it is necessary to modify Y_n so that the limiting distribution is not singular-the limiting distribution function of Y_n itself is a step function with a single step at $E(X)$. On the other hand, the variable

$$nY_n = S_n = X_1 + X_2 + \cdots + X_n$$

has both a mean $(n\mu)$ and a variance ($n\sigma^2$, where $\sigma^2 = var(X)$) that infinite with n. Thus one loses track of the shape of the distribution because it flattens out and moves off to infinity.

A device that keeps the distribution from shrinking or expanding excessively is that of standardization.

$$Z_n = \frac{Y_n - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{S_n - n\mu}{\sqrt{n\sigma^2}},$$

$E(Z_n) = 0$and$Var(Z_n) = 1$ since $Var(Y_n) = \frac{\sigma^2}{n}$

The random variable Z_n has a distribution whose shape can be examined as n becomes infinite since Z_n has fixed mean and variance.

The standardized Z_n can be written

$$Z_n = \sum_{i=1}^{n} U_i \qquad where \quad U_i = \frac{X_i - \mu}{\sqrt{n\sigma^2}}$$

Since $E(X_i - \mu) = 0$and $Var(X_i - \mu) = \sigma^2$, It follows that

$$\Phi_{X_i-\mu}(t) = 1 - \frac{\sigma^2 t^2}{2} + 0(t^2),$$

so that

$$\Phi_{U_i}(t) = \Phi_{x_i-\mu}\left(\frac{1}{\sqrt{n\sigma^2}}\right) = 1 - \frac{t^2}{2n} + 0\left(\frac{t^2}{n}\right)$$

The characteristic function of Z_n is then the nth power.

$$\Phi_{Z_n}(t) = \left[\Phi_{U_i}(t)\right]^n$$
$$= \left[1 - \frac{\sigma^2 t^2}{2} + 0(t^2)\right]^n$$
$$= e^{-\frac{t^2}{n}}$$

the above discussion can be summarized in the following theorem.

Theorem 14.5.1 Central Limit Theorem

let$X_1, X_2, \ldots$ *be a sequence of identically distributed random variables with mean* μ *and variance* σ^2 *(both finite), any finite number of which are independent.*

Let $S_n = X_1 + X_2 + \cdots + X_n$*Then for each* Z,

$$\lim_{n\to\infty} P\left(\frac{S_n - n\mu}{\sqrt{n\sigma^2}} \le Z\right) = \Phi(z) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{z} exp\left(-\frac{1}{2}u^2\right) du$$

Since the limit and can be made arbitrarily close to the limit by taking n large enough, it follows that the distribution function of the standardized sum can be approximated with the aid of the standard normal curve. In particular,

setting$z = \frac{y-n\mu}{\sqrt{n\sigma^2}}$ one has

$$P(X_1 + X_2 + \cdots + X_n \le y) \cong \Phi\left(\frac{y - n\mu}{\sqrt{n\sigma^2}}\right)$$

Summary

In this lecture we laid the foundation for the study of sampling distributions. Given two random variables Y and X if $y - \omega(y)$ is an increasing or decreasing function and $f(x)$ is the density function of X then $g(y)$, the density function of Y, is given by

$$g(y) = f(x)\left|\frac{dx}{dy}\right|$$

in which x is to be replaced by its value in terms of y by means of the relation $y = \omega(y)$

Knowledge of sampling distributions help to explain why

a. under certain conditions some statistics possess probability distributions that can be approximated by the normal curve, and

b. sample statistics and their probability distributions are used to make inferences about sampled populations

Post Test

1. Let X have the p.d.f. $f(x) = \left(\frac{x}{3}\right)^2$, $0 < x < 3$. Find the p.d.f. of $Y = X^3$
2. If the p.d.f. of X is $f(x) = 2xe^{-x^2}$ $0 < x < \infty$ Find the p.d.f. of $Y = X^3$
3. Let X_1, X_2 be a random sample of size 2 from the distribution having p.d.f

   $f(x) = e^{-x}$, $0 < x < \infty$ Find the joint p.d.f. of Y_1 and Y_2 where $Y_1 = X_1 + X_2$ and $Y_2 = \frac{X_1}{(X_1+X_2)}$

4 Give $f(x) = e^{-x}$, $x > 0$, find the density of the variable

a. $Y = \frac{1}{X}$ and
b. $Y \log_e X$

LECTURE FIFTEEN

The Gamma and the Chi-Square Distributions

Introduction

In this lecture, we present a simple derivation of the gamma and the chi-square distributions and their applications. In applied work, gamma distributions give useful representations of many physical situations. They have been used to make realistic adjustments to exponential distributions in representing lifetimes in “life-testing” situations. For example, it is frequently the probability model for waiting time until “death”.

Objective

At the end of this lecture, you should be able to:

1. use the Gamma and the Chi-square distributions in calculating probabilities; and
2. explain the significance of the two distribution models in statistical analysis.

Pre-Test

1. State the significance of the Gamma and the Chi-square distribution models in statistical analysis.
2. discuss the application of the two distribution models in statistical analysis.

CONTENT

The Gamma Density

The density

$$\Gamma(\alpha) = \int_0^\infty y^{\alpha-1} e^{-y} dy \quad \text{--} \quad (15.1)$$

is sometimes called an incomplete gamma function. If α> 1, an integration by parts shows that

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

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

$$\Gamma(\alpha) = (\alpha - 1)(\alpha - 2)(\alpha - 3) \ldots\ldots (4)(3)(2)(1)\Gamma(1)$$

$$= (\alpha - 1)! \quad \text{--} \quad (15.2)$$

Since $\Gamma(1) = 1$, it follows that we must take $0! = 1$.

In (15.1), let us consider a change of variable by writing

$y = \frac{x}{\beta} > \beta > 0$, so that

$$dy = \left(\frac{1}{\beta}\right) dx$$

Then

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

or equivalently,

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

Since $\alpha > 0$, $\beta > 0$, and $\Gamma(\alpha) > 0$, we have

$$f(x) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-\frac{x}{\beta}}, \quad 0 < x < \infty$$

$$= 0, \quad \text{otherwise} \quad \text{---------------------------------------} \quad (15.3)$$

A random variable X has a gamma distribution if its probability density function is of forum (15.3).

Moments and Other Properties

The m.g.f of the gamma distribution is given by

$$m(t) = \int_0^\infty e^{tx} \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-\frac{x}{\beta}} dx \text{ -- } (15.4)$$

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

If we set $y = \frac{x}{\beta}(1-\beta t)$, $x = \frac{\beta y}{(1-\beta t)}$

$t < \frac{1}{\beta}$, we have from (15.4)

$$m(t) = \left(\frac{1}{1-\beta t}\right)^\alpha \int_0^\infty \frac{1}{\Gamma(\alpha)} y^{\alpha-1} e^{-y} dy$$

$$= \frac{1}{(1-\beta t)^\alpha}, \quad t < \frac{1}{\beta} \text{ -----------------------------------}(15.5)$$

Now, $m'(t) = (-\alpha)(-\beta)(1-\beta t)^{-\alpha-1}$

and $m''(t) = (-\alpha)(-\alpha-1)(-\beta)^2(1-\beta t)^{-\alpha-2}$

The mean of the gamma density is

$\mu = m'(0) = \alpha\beta$

and variance

$\sigma^2 = m''(0) - \mu^2 = \alpha(\alpha + 1)\beta^2 - \alpha^2\beta^2 = \alpha\beta^2$

Waiting time

Suppose the waiting time (W) of an event has a gamma p.d.f. with $\alpha = k$ and $\beta = {}^1/_\lambda$. Accordingly $E(W) = \mu = \alpha\beta = K({}^1/_\lambda)$. So that the expected waiting time can be obtained for k changes. For $k = 1$, for example $E(W) = {}^1/_\lambda$.

The Chi-Square Distribution

Consider the special case of the gamma distribution in (15.3) in which $\alpha = \frac{r}{2}$, where r is a positive integer, and $\beta = 2$. That is

$$f(x) = \frac{1}{r\left(\frac{r}{2}\right)2^{r/2}} x^{\frac{r}{2}-1} e^{-\frac{x}{2}}, \quad 0 < x < \infty$$

= 0, otherwise -- (15.6)

A random variable X has a chi-square distribution if its probability density function is of form (15.6). The mean and variance of the chi-square distribution are

$$\mu = \alpha\beta = (^r/_2)(2) = r \quad \text{and} \quad \sigma^2 = (^r/_2)(2)^2 = 2r.$$

Example 15.1

If X has the p.d.f $f(x) = \frac{1}{4} x e^{-\frac{x}{2}}, \quad 0 < x < \infty$

$= 0, \quad$ otherwise.

Compute:

1 the mean and variance of X
2 them.g.f. of X.

Solution

1. X is $\chi^2(4)$. Hence $\mu = 4$ and $\sigma^2 = 8$
2. $m(t) = (1 - \beta t)^{\alpha} = (1 - \gamma t)^{-2}$, $t < ½$

Example 15.2

Let X be $\chi^2(10)$. Compute $P(3.25 \leq X \leq 20.50)$.

Solution

$P(3.25 \leq X \leq 20.50) = P(X \leq 20.50) – P(X \leq 3.25)$

$= 0.975 – 0.025$

$= 0.95$

Summary

In this lecture we have introduced you to the Gamma and, of course the Chi-square distribution, one of the important and frequently encountered sampling distributions. We derived the probability distribution of the Chi-square random variable and we also demonstrated how the probability distribution could be used in computing probabilities.

Post-Test

1. If $(1.2t)^{-6}$, $t < ½$, is the m.g.f. of the random variable X, find: (a) the mean and variance of X. (b) $P(X < 5.23)$.
2. If X is $\chi^2(5)$, estimate the constants K and d so that $P(K < X Cd) = 0.95$ and $P(X < K) = 20.025$.
3. Let X have the uniform distribution with p.d.f. $f(x) = 1$, $0 < x < 1$ and zero otherwise
   a. Find the distribution function of $Y = 2\log_e^X$.
   b. What is the p.d.f of Y.
   c. Compute the mean and variance of Y.

References

Cramer, H. 1955.The Elements of Probability Theory.John Wiley.

Feller, W. 1957.*An Introduction to Probability Theory and Its Applications*.Vol. 1.John Wiley.

Hogg, R. V. and Craig, A. T. 1970.*Introduction to Mathematical Statistics*. New York: Macmillan Publishing Co.

Larson, H. J. 1982. *Introduction to Statistical Theory and Statistical Inference*. New York: John Wiley.

Mood, A. M., Graybill, F. A. and Boes, D. C. 1974.*Introduction to the Theory of Statistics*.McGraw-Hill.

Parzen, E. 1972.*Modern Probability Theory and its Applications*. John Willey.