PREFACE

With its grounding in the “guiding pillars of Access, Equity, Equality, Affordability and Accountability,” the New Education Policy (NEP 2020) envisions flexible curricular structures and creative combinations for studies across disciplines. Accordingly, the UGC has revised the CBCS with a new Curriculum and Credit Framework for Undergraduate Programmes (CCFUP) to further empower the flexible choice based credit system with a multidisciplinary approach and multiple/ lateral entry-exit options. It is held that this entire exercise shall leverage the potential of higher education in three-fold ways – learner’s personal enlightenment; her/his constructive public engagement; productive social contribution. Cumulatively therefore, all academic endeavours taken up under the NEP 2020 framework are aimed at synergising individual attainments towards the enhancement of our national goals.

In this epochal moment of a paradigmatic transformation in the higher education scenario, the role of an Open University is crucial, not just in terms of improving the Gross Enrolment Ratio (GER) but also in upholding the qualitative parameters. It is time to acknowledge that the implementation of the National Higher Education Qualifications Framework (NHEQF), National Credit Framework (NCrF) and its syncing with the National Skills Qualification Framework (NSQF) are best optimised in the arena of Open and Distance Learning that is truly seamless in its horizons. As one of the largest Open Universities in Eastern India that has been accredited with ‘A’ grade by NAAC in 2021, has ranked second among Open Universities in the NIRF in 2024, and attained the much required UGC 12B status, Netaji Subhas Open University is committed to both quantity and quality in its mission to spread higher education. It was therefore imperative upon us to embrace NEP 2020, bring in dynamic revisions to our Undergraduate syllabi, and formulate these Self Learning Materials anew. Our new offering is synchronised with the CCFUP in integrating domain specific knowledge with multidisciplinary fields, honing of skills that are relevant to each domain, enhancement of abilities, and of course deep-diving into Indian Knowledge Systems.

Self Learning Materials (SLM’s) are the mainstay of Student Support Services (SSS) of an Open University. It is with a futuristic thought that we now offer our learners the choice of print or e-slm’s. From our mandate of offering quality higher education in the mother tongue, and from the logistic viewpoint of balancing scholastic needs, we strive to bring out learning materials in Bengali and English. All our faculty members are constantly engaged in this academic exercise that combines subject specific academic research with educational pedagogy. We are privileged in that the expertise of academics across institutions on a national level also comes together to augment our own faculty strength in developing these learning materials. We look forward to proactive feedback from all stakeholders whose participatory zeal in the teaching-learning process based on these study materials will enable us to only get better. On the whole it has been a very challenging task, and I congratulate everyone in the preparation of these SLM’s.

I wish the venture all success.

Professor Indrajit Lahiri
Vice Chancellor

Netaji Subhas Open University

Four Year Undergraduate Degree Programme

Under National Higher Education Qualifications Framework (NHEQF) &
Curriculum and Credit Framework for Under Graduate Programmes

B. Sc. Mathematics (Hons.)

Programme Code : NMT

Course Type : Multi-Disciplinary Course (MDC)

Course Title : Statistical Techniques

Course Code : NMD-MT-01

Printed in accordance with the regulations of the
Distance Education Bureau of the University Grants Commission.

Netaji Subhas Open University

Four Year Undergraduate Degree Programme

Under National Higher Education Qualifications Framework (NHEQF) &
Curriculum and Credit Framework for Under Graduate Programmes

B. Sc. Mathematics (Hons.)

Programme Code : NMT

Course Type : Multi-Disciplinary Course (MDC)

Course Title : Statistical Techniques

Course Code : NMD-MT-01

: *Course Writer* :
Dr. Ashit Baran Aich
Director, Study Centre
Netaji Subhas Open University
[Ex-Reader, Statistics
Presidency College]

: *Course Editor* :
Prof. Ashis Chatterjee
Professor of Statistics &
Pro Vice-Chancellor
(Academic Affair)
University of Calcutta

: *Format Editor* :
Dr. Nemai Chand Dawn
NSOU

Notification

Ananya Mitra
Registrar (Additional Charge)

Netaji Subhas
Open University

UG-Mathematics
(NMT)

Course Title : Statistical Techniques
Course Code : NMD-MT-01

Unit 1 ❑ Probability

Structure

1.0 Objectives

The followings are discussed here:

- Concepts of Probability
- Classical Definition of Probability
- Axiomatic Definition of Probability
- Dependent and Independent Events
- Idea of Random Variables

1.1 Introduction

In this unit, we will study the theory of probability. The primary purpose of having mathematical theory of probability is to provide methematical models for experiments that may arise in different areas of human activity. Such models can then be used for prediction and dicision making. Use of probability theory makes the inferences scientifically valid.

1.2 Basic concepts of Probability

The term 'probability' is frequently used in our day-to-day life, although the user may not be aware of its meaning. For instance, people would ask on the eve of a general election : What is the probability that Mr. A will win? But in so far as statistics is concerned, probability can only be attached to outcomes of those experiments which can be repeated any number of times under essentially identical conditions. Thus, in statements like "Mr. A's winning the election has probability 0.45", the word probability will only mean a numerical measure of the degree of belief attached to a statement. There is a school of subjective probability that is critical of the idea that probability can be calculated in an objective manner. In this chapter, we shall develop only the theory of mathematical probability.

Like any other scientist, a statistician also believes in the maxim that Nature obeys laws though these laws are not exact like those of, say, a physics. Even in the physicist world, the discoveries of the recent years (e.g., quantum mechanics, uncertainty principle, etc.) have emphasized uncertainty and indeterminacy. The laws of statistics are not deterministic, either due to the incompleteness of data or due to the inherent nature of the problem.

Take, for example, the uncertainties in the outcomes of tossing a coin, or in an individuals' behaviour in his consumption. The underlying systematic pattern, however, may not be apparent if only one toss or one individual is studied. But when one considers a large number of outcomes or aggregate behaviour of a large number of people, the regular pattern would come to light. The pattern is called 'statistical regularity'. The definition of probability is based on this concept of statistical regularity.

To get the idea of statistical regularity, let us consider a coin and toss it a large number of times. If after each toss we calculate the proportion of accumulated heads (or tails) upto that point of time, these proportions will be erratic initially. But as the number of tosses increases, these properties will stabilize and the resulting limiting value (ratio) will be called the probability of head, or that of tail as the case may be. If the coin is a perfect (unbiased) one, the limiting ratio will tend to 0.50.

This is called relative frequency approach to probability and has been explained diagrammatically as shown below :

Consider the outcomes of 200 tosses of a coin and the relative frequencies of heads as given in the following table :

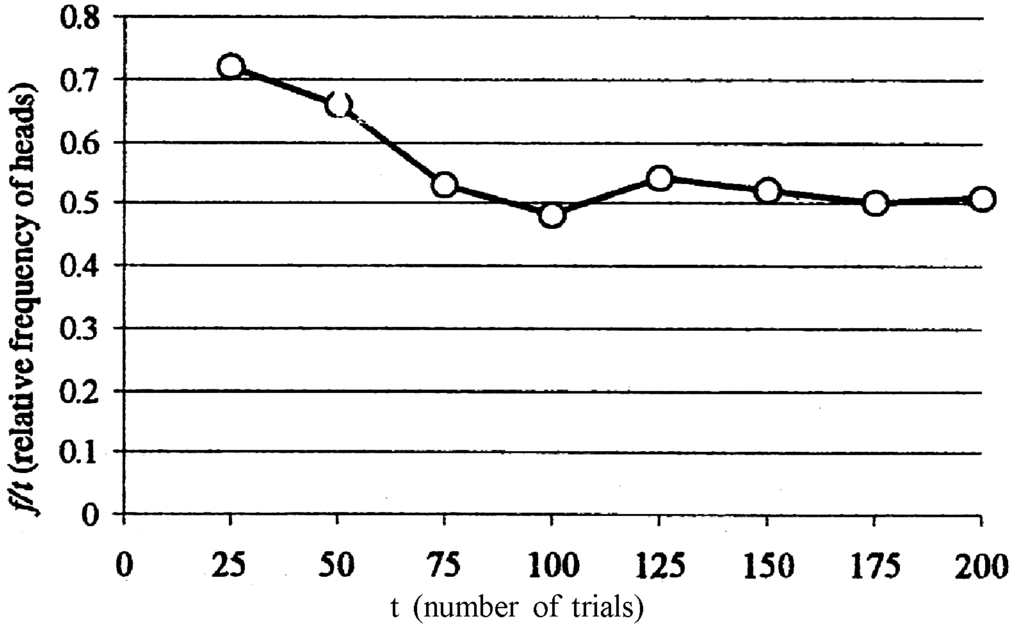


Fig. 6.1 Limitting behaviour of the relative frequency of heads as the total number of tosses increases, as given is table 6.1

Table 6.1 : Relative frequency of heads in 200 tosses of a coin

Number of tosses (t)	25	50	75	100	125	150	175	200
Accumulated number of heads (f)	18	33	40	48	67	78	87	102
Relative frequency f/t	0.72	0.66	0.53	0.48	0.54	0.52	0.50	0.51

Source : Statistics for Social Sciences by Gun & Aich, World Press, Kolkata-700 073.

1.2.1 Classical definition

We formalize the definition of probability by first considering 'sample space' associated with a 'random experiment'. By a random experiment we mean an experiment satisfying the following conditions :

(a) The experiment can be repeated any number of times under essentially similar physical conditions.

(b) The possible outcomes of the experiment are known, but the outcome of any particular trial is unpredictable.

The tossing of a coin, for example, constitutes such a random experiment.

Sample space—The set of all posssible outcomes of a random experiment is called the sample space while the elements of the sample sapce are called elementary events. A sample space can be finite or infinite depending on the number of elements in it.

Event—Any subset of the sample space is called an event. Clearly, sample space itself is also an event, called 'certain event'. Similarly, a null set, being also a subset of the sample space, is also an event which is called 'impossible event'.

Difference between null event (A) and P(A)=0 : Null event $\Rightarrow$P(A) = 0 P(A) = 0 $\neq$ Null event.

- **Exhaustive Events :** A set of events is said to be exhaustive, if they constitute the entire sample space.

- **Mutually Exclusive (Disjoint) events :** Two events are said to be mutually exclusive if they cannot occur together.

Exhaustive and mutually exclusive (disjoint) events :—

Example 1.1 A die is thrown once. What is the sample space in this case?

Ans. Note that if a die is thrown once, then any one of the digits 1, 2, 3, 4, 5, 6 may turn up. So, the set of all possible outcomes, namely,v S = {1, 2, 3, 4, 5, 6} is the sample space which is finite.

In this example, the subsets of S will be the events. Thus, for example, A = {1, 2}, B = {3, 4, 5}, etc. are events. It is to be noted that there are in total 2^6 subsets that can be formed out of S. In general, if the number of elementary events in S in n, then 2^n events can be formed out of S that include the impossible event ϕ (null set) and the certain event S.

Empirical (relative frequency) definition of probability :

If A be an event that may occur as a result of an experiment having S as its sample space, then the probability of occurrence of A, written P(A), is given by.

$$P(A)=\frac{\text{number of repetitions resulting in } A}{\text{total number of repetitions}}$$

$$=\frac{f_n(A)}{n}, \text{ (say).}$$

which tends to a fixed value as n tends to infinity as explained earlier. This is called the 'relative frequency approach' to probability.

Classical definition of probability : This definition used the terminology of set for events and sample space as given earlier. Thus, in this definition, $P(A)$ is given by

$P(A)=\frac{N(A)}{N(\text{S})}$, when $N(A)$ is the number of elementary events in A, and N(S) is that in S.

Clearly, $P\ (A) \geq 0$, as $N(A) \geq 0$

Also, $P\ (A) \leq 1$, as $N(A) \leq N(S)$.

Thus, $P(A)$ is a number lying between 0 and 1.

In this definition, it is assumed that all elementary events in S are 'equally likely' to occur.

This is a drawback (limitation) of the classical defintion as the term 'equally likely' cannot be explained without the concept of probability. In this sense, the classical definition of probability is circular in nature.

Another drawback of the classical definition is that it allows for probability only numbers that are in the form $\frac{m}{n}(m \leq n)$, thus rejecting irrational numbers in (0, 1). These are examples in geometric probability where P(A) can even take values such as $\frac{1}{\pi}$, where $\pi = 3{\cdot}14159$ (approximately)

Let us now define union, intersection difference and complement of events:

- Given two events A and B, their *union* is defined as an event $A \cup B$ which means occurrence of at least one of them.
- The *intersection* $A \cap B$ is the event of joint occurrence.
- *Difference* A - B is the event of occurrence of A only.
- *Complement of* A is the part of the sample space excluding A and written as A^c.

1.2.2 Axiomatic definition

In this approach instead of defining probability explicitly, some conditions are postulated to be satisfied by probability.

Let $A_1, A_2, ..., A_k$ be the events belonging to the collection of events, namely, Q. This non-empty class of events Q is closed under finite unions and complementation, and is termed as the 'field' or 'algebra' of events.

In this approach to probability, probability is defined as a finite real-valued function P(.) defined on the field of events Q satisfying he following conditions.

(A) $P(A) \geq$ for any $A \in Q$,

(B) $P(S) = 1$, S being the sample space,

(C) $P\left(\bigcup_{i=1}^{k} A_i\right) = \sum_{i=1}^{k} P(A_i)$, where $A_i \in Q(i = 1, 2, ..., k)$ and are all disjoint.

The above result will have a natural extension to the case where k is infinity, and the corresponding Q will be termed as σ-field (or σ-algebra).

We now state and prove the addition and multiplication theorems of probability :

Theorem 1. Let A and B be two events which are not disjoint, i.e., they can occur together. Then the probability that either A, or B, or both will occur is given by

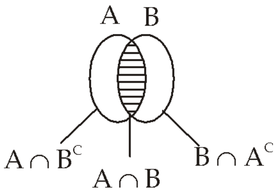
$$P(A \cup B) = P(A) + P(B) - P(A \cap B),$$

where $P\ (A \cap B)$ is the probability of their joint occurrence.

Proof : Here, A and B are not disjoint. We represent the set $A \cup B$ by $A \cup B = (A \cap B^C) \cup\ (A \cap B) \cup (B \cap A^C)$, where $A \cap B^C$, $A \cap B$, and $B \cap A^C$ are disjoint events.

Then by the axioms on probability,

we have,



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

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

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

$$= P(A) + P(B) - P\ (A \cap B)$$

$\because\ (A \cap B^C) \cup (A \cap B) = A,\ (B \cap A^C) \cup (A \cap B) = B)$

Thus, $\quad P(A \cup B) = P(A) + P(B) - P(A \cap B).$

Note : If A and B are disjoint (also called mutually exclusive events), we have $A \cap B = \phi$, null set.

In this case, $P\ (A \cap B) = P(\phi) = 0$, and $P\ (A \cup B) = P\ (A) + P(B)$

In general, let A_1, A_2, ..., A_n are events belonging to the field of events Q, say,

$$\text{then, } P\left(\sum_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i) - \sum_{i<j} P\left(A_i \cap A_j\right) + \sum_{i<j<k} P\left(A_i \cap A_j \cap A_k\right)$$

$$+\ (-1)^{n-1}\, P\left(A_1 \cap A_2 \cap .....A_n\right)$$

If the events A_i's are all disjoint, then

$$P\left(\overset{n}{\underset{i=1}{\cup}} A_i\right) = \sum_{i=1}^{n} P(A_i)$$

(Define conditional probability)

Theorem 2 : Let A and B two events belonging to the field of events Q, say, then the probability of their joint occurence is given by

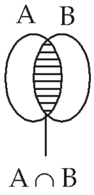
$P(A \cap B) = P(A)P(B/A) = P(B)P(A/B)$, when $P(B/A)$

$P(A/B)$ are conditional probabilities.

Here, it is assumed that $P(B/A)$, $P(A/B) > 0$

Proof : Let $N(A)$, $N(B)$, and $N(S)$ denote the number of elementary events in A, B, and S, where S is the sample space which is assumed to be finite.

Then, by the classical definition of probability,



we have $P(A \cap B) = \dfrac{N(A \cap B)}{N(S)}$

$$= \frac{N(A \cap B)}{N(A)} \cdot \frac{N(A)}{N(S)} = P(B/A) \cdot P(A),$$

where $P(B/A) = \dfrac{N(A \cap B)}{N(A)}$, is the

probability that B has occurred, given that event A has already occurred.

Similarly, $P(A \cap B) = P(B).P(A/B)$.

In general, if $A_1, A_2, \ldots\ldots, A_n$ are events belonging to the field of events Q, then

$P(A_1 \cap A_2 \cap A_3 \cap \ldots\ldots \cap A_n)$

$= P(A_1)\, P(A_2/A_1)\, P(A_3/A_1 \cap A_2) \ldots P(A_n/A_1 \cap A_2 \cap A_3 \cap \ldots . \cap A_{n-1})$.

Statistically independent events :

Two events A, B are said to be independent statistically if $P(A \cap B) - P(A).\ P(B)$.

i.e., the probability of the joint occurrence is equal to the product of their individual occurrences.

Thus, if A_1, A_2,....., A_n are statistfically independent events, then

$$P\left(\overset{n}{\underset{i=1}{\cap}} A_i\right)=\prod_{i=1}^{n} P(A_i).$$

Example (a) : Let a die be thrown once. Define the events A and B as follows :

$A \equiv$ occurrence of points ≤ 4.

$B \equiv$ occurence of even points.

Then, $A = \{1, 2, 3, 4\}$,

$B = \{2, 4, 6\}$,

and S = sample space

$= \{1, 2, 3, 4, 5, 6\}$, $A \cap B = \{2, 4\}$

So, $P(A)=\frac{4}{6}=\frac{2}{3}$, $P(B)=\frac{3}{6}=\frac{1}{2}$

$$P(A/B) = \frac{2}{3},\ P(B/A) = \frac{2}{4}= \frac{1}{2}$$

Note that by using formulae,

$$P(A/B) = \frac{P(A\cap B)}{P(B)}=\frac{2}{6}\times\frac{6}{3}= \frac{2}{3}$$

and $P(B/A) = \frac{2}{6}\times\frac{6}{4}= \frac{1}{2}$.

Example (b) : Take four identical marbles. On the first, write symbols $A_1A_2A_3$. On each of the other three, write A_1, A_2 and A_3 respectively. Put the four marbles in an urn and draw one at random. Let E_i denote the event that the symbol A_i(i = 1, 2, 3) appears on the drawn marble.

Then, $P(E_1)=\frac{1}{2}=P(E_2)=P(E_3)$,

$$P(E_1 E_2) = \frac{1}{4} = P(E_1 E_3) = P(E_2 E_3),$$

also, $P(E_1 E_2 E_3) = \frac{1}{4}$

Thus, $P(E_1 E_2 E_3) \neq P(E_1)\, P(E_2)\, P(E_3)$,

So that the events E_1, E_2, E_3 are not independent.

But $P(E_1\ E_2) = \frac{1}{4} = P(E_1) \cdot P(E_2)$

Similarly, $P(E_1 E_3) = P(E_1)P(E_3)$, $P(E_2 E_3) = P(E_2)P(E_3)$,

showing that they are pairwise independent.

Mutual independence of events :

In general, $r(> 2)$ events $A_1, A_2, ..., A_r$ are said to be mutually independent if the following equations are satisfied :

$P(A_i \cap A_j) = P(A_i)\, P(A_j),\ 1 \leq i < j \leq r,$

$P(A_i \cap A_j \cap A_k) = P(A_i)\, P(A_j)\, P(A_k),\ 1\ \leq\ i < j < k \leq r$

$P(A_1\, A_2\, ...A_r) = P(A_1)\, P(A_2)\, ...\, P(A_r)$

Thus, there are in total $2^r - r - 1$ $(r > 2)$ such equations for r events.

Thus, for the mutual independence of r events, it is not enough that events are pairwise independent only. Rather, all the equations are to be satisfied.

Theorem 3. Let us consider a partition of the sample space S into the event B_1, B_2, ..., B_k such that

$$\bigcup_{i=1}^{k} B_i = S, \text{ and } B_i \cap B_j = \phi, \text{ null set, } i \neq j.$$

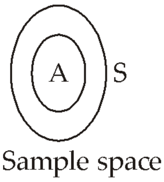
Then, for any event A, and assuming $P(B_i) > 0$, for such i, we have P(A)

$$= \sum_{i=1}^{k} P(B_i)\, \mathrm{P(A \mid B}_i).$$

Proof : Consider the events $A \cap B_1$, $A \cap B_2$, ..., $A \cap B_k$ for which

$$(A \cap B_i) \cap (A \cap B_j) = A \cap B_i \cap B_j$$

$$= A \cap \phi = \phi, \text{ null set.}$$



Sample space

then, $A = A \cap S$

$$= A \cap \left(\bigcup_{i=1}^{k} B_i \right)$$

$$= \bigcup_{i=1}^{k} (A \cap B_i)$$

so that $P(A) = P\left(\bigcup_{i=1}^{k} (A \cap B_i) \right)$

$$= \sum_{i=1}^{k} P(A \cap B_i), \text{ since } A \cap B_i\text{, s are disjoint or mutually exclusive.}$$

$$= \sum_{i=1}^{k} P(B_i) P(A / B_i) \text{ , by multiplication theorem.}$$

This theorem is called the theorem of total probability.

Theorem 4. (Bayes' theorem)

Let B_1, B_2, ..., B_k be exhaustive and mutually exclusive events belong to the field of events Q such that $P(B_i) > 0$ for each i. Also, let A be another event in Q. Then

$$P(B_i / A) = \frac{P(B_i) P(A / B_i)}{\sum_{i=1}^{k} P(B_i) P(A / B_i)}, \text{ for } P(A) > 0,$$

$(i = 1, 2, 3, \text{ ....}, k)$

Proof : By the theorem of total probability, we have P (A) = $\sum_{i=1}^{k} P(B_i)P(A/B_i)$

Also, $P(A \cap B_1) = P(A)P\ (B_1 /A) = P(B_i)\ P\ (A/B_i)$

$$\therefore\ P(B_i / A) = \frac{P(B_i)P(A/B_i)}{P(A)}$$

$$= \frac{P(B_i)P(A/B_i)}{\sum\limits_{i=1}^{k} P(B_i)P(A/B_i)},\ i = 1,\ 2,\ .....,\ k.$$

Importance of Bayes' theorem : Bayes' theorem has a wide significance in statistical inference. In fact, a separate School of Statistics, called Bayesian School, is founded on this theorem. Here, B_1, B_2, ..., B_k may be looked upon as "causes" while A can be treated as 'effect'. So, here we discuss probability of a cause (or parameter) given the effect (or data). In this approach, parameters are treated as random variables having some probability distributions.

Example (c) : Let there be three urns numbered U_1, U_2, U_3 having marbles (2 white, 3 black), (3 white, 2 black) and (4 white, 5 black), respectively. An urn is chosen at random, and a marble is drawn from it. Let it be white what is the probability that U_2 was chosen?

Solution : Let A ≡ event that a white ball is drawn.

Then $P(A \mid \cup_1) = \frac{2}{5}$, $P(A \mid \cup_2) = \frac{3}{5}$, $P(A \mid \cup_3) = \frac{4}{9}$

Also, $P(U_1) = P(U_2) = P(U_3) = \frac{1}{3}$.

Then, by Bayes' theorem,

$$P(\cup_2 \mid A) = \frac{P(\cup_2)P(A \mid \cup_2)}{P(\cup_1)P(A \mid \cup_1) + P(\cup_2)P(A \mid \cup_2) + P(\cup_3)P(A \mid \cup_3)}$$

$$= \frac{\frac{1}{3}.\frac{3}{5}}{\frac{1}{3}\left(\frac{2}{5}+\frac{3}{5}+\frac{4}{9}\right)}$$

$$= \frac{27}{18+27+20}$$

$$= \frac{27}{65} = 0{\cdot}415$$

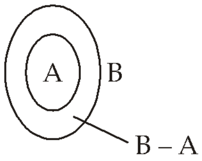
Note that $P(U_2 \mid A) > P(U_2)$ which means that the information or knowledge of the occurrence of A has improved the probability, i.e., reduced the margin of uncertainty about U_2.

Theorem 5 : The probability function $P(\cdot)$ is monotone, i.e., if A and B are events in the field of events Q and $A \subset B$, then $P(A) \leq P(B)$.

Proof : Let us units B as

$B = A \cup (B - A)$

So that A and $(B - A)$ are disjoint.



Then, $P(B) = P(A \cup (B - \text{A}))$

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

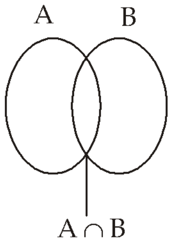
$\geq P(A)$, as $P(B - A) \geq 0$.

$\therefore P(A) \leq P(B)$.

Hence, $A \subset B \Rightarrow P(A) \leq P(B)$, showing that $P(\cdot)$ is monotone.

A simple proof of $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ can be obtained from the classical definition of probability, as shown below :

Let A, B be the two events in some field of events Q, and S be the sample space which is assumed to be finite. Also let $N(A)$, $N(B)$ and $N(S)$ be the number of elementary events in A, B and S, respectively.



Then, $P(A \cup B) = \frac{N(A \cup B)}{N(S)}$,

$$= \frac{N(A) + N(B) - N(A \cap B)}{N(S)}$$

$N(A \cap B)$ being the number of elementary events in the set $A \cap B$.

Thus, $P(A \cup B) = \frac{N(A)}{N(S)} + \frac{N(B)}{N(S)} - \frac{N(A \cap B)}{N(S)}$

$$= P(A) + P(B) - P(A \cap B)$$

Corollary : Put $B = A^C$, then $A \cup B = A \cup A^C = S$, $A \cap B = A \cap A^C = \phi$.

This gives $1 = P(A \cup A^C)$

$$= P(A) + P(A^C) = 0$$

$\therefore P(A^C) = 1 - P(A)$.

Example (d) : Suppose n objects numbered 1, 2, 3, ..., n are distributed at random among n cells also numbered 1, 2, 3,...., n. What is then the probability that there will be no matching of the numbered objects with the cells?

Solution : We say that there is a matching if the ith object goes to the ith cell. Let A_i denote the event that ith object goes to the ith cell (i = 1, 2, ..., n), and A_i^c denotes the complementary event of no matching . We went to obtain the probability $P\left(\bigcap_{i=1}^{n} A_i^C\right)$.

By De Morgan's rule, we have

$$\bigcap_{i=1}^{n} A_i^C = \left(\bigcup_{i=1}^{n} A_i\right)^C$$

$$\therefore\ P\left(\bigcap_{i=1}^{n} A_i^C\right) = P\left(\bigcup_{i-1}^{n} A_i\right)^C = 1-P\left(\bigcup_{i-1}^{n} A_i\right)$$

$$= 1-\left[\sum_{i=1}^{n} P(A_i) - \sum_{i<j}\sum P\left(A_i \cap A_j\right) + \sum_{i\ <j\ <k}\sum\sum P\left(A_i \cap A_j \cap A_k\right)\right.$$

$$\left. +\dots\dots\dots + (-1)^{n-1} P\left(A_1 \cap A_2 \cap \dots\dots\dots A_n\right)\right] \quad \dots\dots \text{(i)}$$

Note that, $P\left(A_j\right) = \dfrac{\underline{|n-1}}{\underline{|n}} = \dfrac{1}{n}$, $P(A_i \cap A_j) = \dfrac{\underline{|n-2}}{\underline{|n}} = \dfrac{1}{n(n-1)}$,

and, finally, $P(A_1 \cap A_2 \cap \dots\dots \cap A_n) = \dfrac{1}{\underline{|n}}$

So, from (1),

$$P\left(\bigcap_{i=1}^{n} A_i^c\right) = 1-\frac{n}{n}+\binom{n}{2}.\frac{1}{n(n-1)}-\binom{n}{3}\frac{1}{n(n-1)(n-2)}\dots+(-1)^n\frac{1}{\underline{|n}}$$

$$= 1-\frac{1}{\underline{|1}}+\frac{1}{\underline{|2}}-\frac{1}{\underline{|3}}+\dots\dots+(-1)^n\frac{1}{\underline{|n}}$$

For large value of n , the above probability can be approximated by $P\left(\bigcap_{i=1}^{n} A_i^c\right) \simeq e^{-1}$.

1.2.3 Random variable

In many situations, the outcomes of a random experiment are in the form of qualitative characters. For example, in tossing of a coin once the sample space is $S = \{H, T\}$, where '*H*' and '*T*' stand respectively for a head and tail. Similarly, if we want to record the sex of each new born child in the maternity ward of a city hospital on a particular day, then also the sample space cannot be described numerically, rather as S ={Male, Female}. It would be mathematically simpler if we could quantify the sample space. This is achieved by defining a real-valued function on the sample space. This function is called a random variable ($r.v$).

Thus, for example, in a coin-tossing experiment, when the coin is tossed once, we define on S a function X (.) such that

$X(H) = 1, X(T) = 0.$

In mathematical terms, we write $X : S \to \{1, 0\}$, i.e., X is a mapping of S into $\{1, 0\}$.

Other examples of a random variable are the number of heads obtained in tossing a coin trice, the number of boys or girls in families of a locality in the city, and so on.

The difference between a random variable and a usual mathematical variable is that while the occurrence of a random variable depends on chance, the latter type of variable will either occur or will not occur. It is to be noted that random variable plays the most important role in statistical analysis.

1.3 Empirical and Theoretical Distribution

1.3.1 Concepts

Let $X_1, X_2, ..., X_n$ be a random sample from some distribution. Then, calculate the ratio

$$S_n(x) = \frac{\text{number of } X_i \text{ which are} \leq x}{n}$$

where x is a pre-assigned fixed number.

Note that $S_n(x)$ simply gives the proportion of observations in the sample which are less than or equal to x. It is a random variable taking values $\frac{0}{n}, \frac{1}{n}, ..., 1$. This $S_n(x)$ is termed as the empirical distribution function.

The function $S_n(x)$ satisfies the following conditions :

(1) $S_n(x)$ is non-decreasing, and continuous from the left,

(2) $S_n(-\infty) = 0$, $S_n(+\infty) = 1$

(3) $S_n(x)$ is a step-function with discontinuities at n-points.

Theoretical distribution function :

The theoretical distribution function $F(x)$ of a random variable X is defined as

$F(x) = P(X \leq x)$, when x is a pre-assigned

constant,

$F(x)$ satisfies the following conditions—

(1) $F(-\infty) = 0$, $\mathrm{F}(\infty) = 1$.

(2) F is non-decreasing

(3) F is continuous at least from the left.

(4) The set of points of discontinuity of F is at most countable.

1.3.2 Probability mass function and probability density function

Probability mass function (p.m.f) :

Let X be a discrete random variable taking values in a set S say, of non-negative integers. Then, the probability

$\phi(k) = P(X = k), k \in S$

is the pobability that X takes a particular value k. Then $p(k)$ with be the probability mass function (p.m.f.) of X.

Clearly, $p(k)$ satisfies the following conditions—

(1) $p(k) \geq 0$, (2) $\sum_{k \in S} p(k) = 1$.

Probability density function (p.d.f) :

In the continuous case, the probability distribution assigns a probability to every interval in which x may lie. Thus, the probability that X lies in an interval centred at x of length dx (where 'dx' is very small), is given by

$$P\left[x-\frac{1}{2}dx \le X \le x+\frac{1}{2}dx\right] = f(x)dx$$

The function $f(x)$ will be termed as the probability density function (p.d.f) of the continuous random variable X.

The following are some important properties of $f(x)$:

(1) $F(x) = \int\limits_{-\infty}^{x} f(t)dt$, when F (x) is the distribution function (also called 'cumulative distribution function') of X.

(2) $f(x) = F'(x) = \frac{dF(x)}{dx}$, provided $F'(x)$ exists.

In other words, given $F(x)$ we can calculate $f(x)$ by differentiation, and given $f(x)$ we can obtain $F(x)$ by integration as given above.

1.3.3 Mathematical expectation and variance

The quantities 'expectation' and 'variance' of a random variable X measure turns important characteristics of its probability distribution. While the first measures the 'centre' of the distribution, the second measures the 'spread' of the distribution about the centrally located value.

Suppose X is discrete with pmf (x). Then the expectation of X, written $E(x)$, is given by

$E(x) = \sum\limits_{x \in s} xp(x)$, when S is the set of all possible values of X, provided the sum exists.

The variance of X, written $V(x)$, is given by

$V(x) = E(X-\mu)^2$, when $\mu = E(\mathrm{x})$

$= \sum\limits_{x \in s} (x-\mu)^2 f(x)$, which is essentially positive

In case X is continuous, we have

$$E(X) = \int_{-\infty}^{\infty} xf(x)dx \text{ (provided it exists)}$$

and $$V(X) = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx \geq 0 .$$

On simplification, $V(X) = E\left(X^2\right) - (E(X))^2$,

which implies $E\left(X^2\right) \geq (E(X))^2$

Two results are worth noting :

(1) $E(a + bX) = a + bE(X)$,

(2) $V(a + bX) = b^2V(X)$, a & b being constant.

Example 1.2.1 Suppose a fair coin is tossed twice. Let X denote the number of heads obtained. Write down the probability distribution of X. Also, calculate $E(X)$ and $V(X)$.

Solution : Since the coin is tossed twice in succession, the sample space is $S = \{HH, HT, TH, TT\}$

Define X such that $X(HH) = 2$, $X(HT) = 1$, $X(TH) = 1$, $X(TT) = 0$

Then the probability distribution of X is given by

X	0	1	2
$p(x) = P(X = x)$	$\frac{1}{4}$	$\frac{2}{4}$	$\frac{1}{4}$

Here, $E(X) = \sum_{x=0}^{2} xp(x)$

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

$$= \frac{2}{4}+\frac{9}{4}=1$$

Also, $E\left(X^2\right)=\sum_{x=0}^{2} x^2 p(x)$

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

$$= \frac{2}{4}+\frac{4}{4}$$

$$= \frac{3}{2}$$

Finally, $V(X)=E\left(X^2\right)-\left(E(X)\right)^2= \frac{3}{2}-1 = \frac{1}{2}$

The positive square root of the variance is called standard deviation (s.d.) which is given by

$$s.d. = +\sqrt{V(X)}=\frac{1}{\sqrt{2}}=\sigma, \text{ (say)}$$

Example 1.2.2 Let X be a continuous randon variable with pdf.

$$f(x)=\begin{cases}\theta e^{-\theta x}, & x>0, \theta>0\\ 0 & , x\le 0\end{cases}$$

Find $F(x)$, $E(X)$ and $V(X)$.

Solution : By defintion,

$$F(x) = P(X \le x),\ (x > 0)$$

$$= \int_{-\infty}^{0} f(x)dx+\int_{0}^{x} f(x)dx$$

$$= 0+\int_0^x \theta e^{-\theta x/dx}$$

$$= \left(1-e^{-\theta x}\right)$$

$$E(X) = \int_0^\infty xf(x)dx = \theta\int_0^\infty xe^{-\theta x}dx$$

$$= \theta\frac{\Gamma(2)}{\theta^2},$$

(using the gamma integral $\int_0^\infty x^{p-1}e^{-\theta x}dx = \frac{\Gamma(p)}{\theta^p}$, (if θ > 0, p > 0))

(The above integral can also be calculated by integration by pats)

As $\Gamma(2)=1$, we have $E(X)=\frac{1}{\theta}$.

Again, $E(X^2) = \int_0^\infty x^2 f(x)dx$

$$= \theta\int_0^\infty x^2e^{-\theta x}dx$$

$$= \theta\frac{\Gamma(3)}{\theta^3}=\frac{2}{\theta^2}$$

Finally, $V(X) = E(X^2) - (E(X))^2$

$$= \frac{2}{\theta^2}-\left(\frac{1}{\theta}\right)^2 = \frac{1}{\theta^2},$$

and s.d. = $\sqrt[+]{V(X)}=\frac{1}{\theta}$.

1.3.4 Conditional expectation and co-variance

We explain the idea of conditional expectation (or mean) by considering the joint distribution F(x,y) of two continuous random variable X and Y given by

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

for all pairs $(x,\ y) \in R^2$, two-dimensional real-line.

In this case, we have the result that $\frac{\partial^2 F(x,y)}{\partial x \partial y}$ exist and equal to f(x, y), the joint pdf of X and Y.

Also, let

$$g(x) = \int_{-\infty}^{\infty} f(x,y)dy\text{, marginal pdf of X,}$$

and $h(y) = \int_{-\infty}^{\infty} f(x,y)dx$, marginal pdf of Y.

We also define the conditional distributions of X and Y as follow :

The conditional pdf of X, given Y = y is $\frac{f(x,y)}{h(y)} = f_X\left(\frac{x}{y}\right)$, say and conditional cdf of X, given Y = y is

$$F_X\left(\frac{x}{y}\right) = \int_{-\infty}^{x} \frac{f(x,y)}{h(y)}dx$$

Similarly, for the conditional distribution of Y, given $X = x$. Thus,

The conditional *pdf* of Y, given $X = x$ is

$$f_Y\left(\frac{y}{x}\right)=\frac{f(x,y)}{g(x)}$$

and the conditional cdf of Y, given $X = x$, is

$$F_Y\left(\frac{y}{x}\right)=\int_{-\infty}^{y}\frac{f(x,y)}{g(x)}dy.$$

We now prove two useful properties of expectation :

(a) $E(X + Y) = E(X) + E(Y)$.

Let $f(x, y)$ be the joint p.d.f of X and Y.

Then $\mathrm{E}(X + Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x+y)f(x,y)dxdy$

$$=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}xf(x,y)dxdy+\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}yf(x,y)dxdy$$

$$=\int_{-\infty}^{\infty}x\left(\int_{-\infty}^{\infty}f(x,y)dy\right)dx+\int_{-\infty}^{\infty}y\left(\int_{-\infty}^{\infty}f(x,y)dx\right)dy$$

$$=\int_{-\infty}^{\infty}xg(x)dx+\int_{-\infty}^{\infty}yh(y)dy$$

$$= E(X) + E(Y).$$

Note that g(x) and g(y) are the marginal pdf's of X and Y, respectively.

(b) $E(XY)=E(X)E(Y)$, if X and Y are independent.

As in the previous case, let f(x,y) be the joint pdf of X and Y. Then

$$\mathrm{E}(XY) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf(x,y)dxdy$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy\,g(x)h(y)dxdy$$

(since f(x,y) = g(x) h(y) for independence of X and Y)

$$= \int_{-\infty}^{\infty} (xg(x)\left(\int_{-\infty}^{x} yh(y)dy)dx \right)$$

$$= \int_{-\infty}^{\infty} xg(x)E(Y)dx$$

$$= E(Y)\int_{-\infty}^{\infty} xg(x)dx$$

$$= E(Y)E(X).$$

The conditional mean of Y, given x, if it exists is given by

$$E\left(Y/x\right) = \int_{-\infty}^{\infty} yd\mathrm{F}_Y\left(y/x\right)$$

$$= \int_{-\infty}^{\infty} y\frac{f(x,y)}{g(x)}dy$$

The conditional variance of Y, given x, is given by

$$\int_{-\infty}^{\infty}(y-E(Y/x))^2\,dF_Y(y/x)$$

$$=\int_{-\infty}^{\infty}(y-E(Y/x))^2\frac{f(x,y)}{g(x)}dy.$$

Theorem 1.2.1 If $E(Y/x)$ exists for almost all x, then $E(Y)=EE(Y/X)$

Proof : We have $E(Y/x)=\int_{-\infty}^{\infty}y\frac{f(x,y)}{g(x)}dy.$

$$\therefore EE(Y/x)=\int_{-\infty}^{\infty}\left(y\frac{f(x,y)}{g(x)}dy\right)g(x)dx$$

$$=\int_{-\infty}^{\infty}y\left[\int_{-\infty}^{\infty}f(x,y)dx\right]dy$$

$$=\int_{-\infty}^{\infty}yh(y)dy$$

$= E(Y)$, which proves the result.

Thus, $E(Y) = EE(Y/X)$

Similarly it can be proved that

$V(Y)=EV(Y/X)+VE(Y/X)$, assuming the existence of relevant quantities.

Co-variance between *X* and *Y* : This is an extension of the concept of variance to the case of two random variables. We define co-variance between X and Y, written Cov(X,Y) as

$$\text{Cov }(X, Y) = E(X-E(X)\ (Y-E(Y)$$
$$= E\{XY-XE(Y) - E(X)Y + E(X)E(Y)\}$$
$$= E(XY) - E(Y)E(X) - E(X)E(Y) + E(X)E(Y)$$
$$(\because E\ (XE(Y))=E(Y)E(X),$$

as $E(Y)$ is a number and $E(x.b) = b$ E(x)))

$$= E(XY) - E(X)E(Y).$$

The following are inportant properties of Cov (X,Y) :

(1) $|Cov(X,Y)|^2 \leq V(X).V(Y)$

(2) Cov $(a + \text{b}X, c + dY) = bd$ Cov(X,Y)

(3) $V(X \pm Y) = V(X) + V(Y) \pm 2C\text{ov}(X,Y)$

(4) For four variables X,Y, U and V, Cov$(X+Y, U+V)$ = Cov(X,U) +Cov (X, V) + Cov(Y,U) + Cov (Y,V)

(5) Cov$(X,X) = V(X)$

Example 1.2.3 Let X and Y be two random variables. Then find Cov $(X+Y, X-Y)$

Solution. By the property of the co-variance, we have

Cov $(X+Y, X-Y)$ = Cov(X,X) + Cov$(X, -Y)$ + Cov(Y,X)+Cov$(Y,-Y)$

= Cov(X, X) – Cov (X, Y) +Cov(Y,X) + Cov $(Y, -Y)$

= $V(X) - V(Y)$. ($\because$Cov (X,Y)=Cov(Y,X)

1.4 Moments and moment generating function (mgf).

1.4.1 Concepts

The study of the probability distributions of a random variable is essentially the study of some numerical characteristics associated with them. These so-called parameters of the distribution play a key role in mathematical statistics. One such example of parameters is moment and its functions.

Discrete case :

Let X be a discrete random variable with pmf $p_k = \{X = k\}, k = 0,1,2...$

$$\text{with} \quad p_k \geq 0 \text{ and } \sum_{k=0}^{\infty} p_k = 1$$

Then, the rth raw moment of X is given by

$$\mu'_r = E(X^r) = \sum_{x=0}^{\infty} x^r p(x), \quad r = 0,1,2......$$

The rth central moment of X is similarly defined

as $\mu_r = \mathrm{E}(\mathrm{X} - \mathrm{E}(X))^r$

$$= E(X - \mu'_1)^r, (\because E(X) = \mu'_1 = \sum_{x=0}^{\infty} xp(x))$$

$$= \sum_{x=0}^{\infty} (x - \mu'_1)^r p(x)$$

Note that $\mu_2 = E(X - \mu'_1)^2$ = variance of X

On simplification,

$$\mu_2 = \mu'_2 - \mu'^2_1$$

Thus, the first order raw moment is the mean and second order central moment is the variance of a distribution.

Also, $\mu_0 = 1$ and $\mu_1 = 0$

Continuous case:

The above has a natural extension to the case of continuous random variable. Thus, let f(x) be the pdf of a continuous random variable, with

$$\int_{-\infty}^{\infty} f(x)dx = 1 \quad \text{and} \quad f(x) \geq 0.$$

The rth raw and central moment of X are then given by

$$\mu'_r = \int_{-\infty}^{\infty} x^r f(x)dx,$$

and $$\mu_r = \int_{-\infty}^{\infty} (x - \mu'_1)^r f(x)dx,$$

where r = 0,1,2, ...

Note that $\mu'_1 = 0$, then the raw and central moments are identical.

Moment generating function (mgf) :

The mgf of a random variable X (about zero) is defined as

M(t) = E(e^{tx}), (t is a parameter)

while the mgf about mean μ'_1 is defined as

$$M_{\mu'_1}(t) = E\left(e^{t(x-\mu'_1)}\right)$$

Note that $M_{\mu'_1}(t) = E\left(e^{tx}, e^{-t\mu'_1}\right)$

$$= e^{-t\mu'_1} E(e^{tx})$$

$$= e^{-t\mu'_1} \mathrm{M}(t)$$

It should be noted that there are distributions for which mgf does not exist.

1.4.2. Properties of mgf

Below we show that M(t) generates raw moments while $M_{\mu_1'}(t)$ generates central moments :

We have $M(t) = E(e^tX)$

$$= E\left(\sum_{r=0}^{\infty}\frac{(tX)^r}{\lfloor r}\right)$$

$$= \sum_{r=0}^{\infty}\frac{t^r}{\lfloor r}E(X^r), \text{ (since expectation is additive)}$$

$$= \sum_{r=0}^{\infty}\frac{t^r}{\lfloor r}\mu_r',$$

so that μ_r' can be obtained as the coefficient of $\frac{t^r}{\lfloor r}$ in M (t).

Again, $M_{\mu_1'}(t) = E\left(e^{t(X-\mu_1')}\right)$

$$= E\left(\sum_{r=0}^{\infty}\frac{\left(t(X-\mu_1')\right)^r}{\lfloor r}\right)$$

$$= \sum_{r=0}^{\infty}\frac{t^r}{\lfloor r}E(X-\mu_1')^r$$

$$= \sum_{r=0}^{\infty}\frac{t^r}{\lfloor r}\mu_r,$$

so that μ_r can be obtained as the coefficient of $\frac{t^r}{\underline{|r}}$ in $M_{\mu_1'}(t)$.

It can be also be shown that $\left.\frac{d^r M(t)}{dt^r}\right|_{t=0} = \mu_r'$

and $\left.\frac{d^r M_{\mu_1}(t)}{dt^r}\right|_{t=0} = \mu_r$

Two other important properties of mgf are given below :

1. The mgf of a linear function $Y = a + bX$, 'a' and 'b' being constant.

   Let $M^*(t)$ be the mgf of $Y= a + bX$. Then,

   $$M^*(t) = E(e^{t(a+bX)}) = E(e^{at}.\ e^{btX})$$

   $$= e^{at} E\ (e^{btX})$$

   $$= e^{at}M(bt), \text{ when}$$

   $$M(t) = E(e^{tX}), \text{ mgf of } X.$$

2. Let X_1 and X_2 be two vandom variables which are assumed to be independent. Then the mgf of

   $Y= X_1 + X_2$ is

   $$M^Y(t) \quad = E\ (e^{tY})$$

   $$= E\left(e^{t(X_1+X_2)}\right)$$

   $$= E\left(e^{tX_1}.e^{tX_2}\right)$$

   $$= E\left(e^{tX_1}\right).E\left(e^{tX_2}\right), \ (\because\ X_1 \text{ and } X_2 \text{ are independent})$$

   $$= M^{X_1}(t).M^{X_2}(t)$$

This shows that the mgf of the sum of two independent random variables in the product of their mgfs. In general, the mgf of the sum of a fixed number of independent random variables for which the mgf exists is the product of the mgf's of the summands.

1.4.3 mgf of some distributions

Example (a) : Let X have the pdf given by $f(x)=\begin{cases}\frac{1}{2}e^{-x/2}, x>0\\ 0, \text{otherwise}\end{cases}$

Then, $M(t)=E\left(e^{tX}\right)$

$$= \frac{1}{2}\int_0^\infty e^{tx}.e^{-x/2}dx$$

$$= \frac{1}{2}\int_0^\infty e^{-\frac{x}{2}(1-2t)}dx$$

$$= \frac{1}{2}\frac{\Gamma(1)}{\frac{1}{2}(1-2t)}, \text{ (Provided 1– 2t > 0)}$$

$$= (1-2t)^{-1},\ t<\frac{1}{2}.$$

Example (b) : Let X be discrete with pmf given by

$P(x)=\theta$, x = 1, $0<\theta<1$

= $1-\theta$, x = 0

Note that $p(1) + p(0) = \theta+(1-\theta)=1$

and $p(1) > 0, p(\text{o}) > 0$

The mgf of X about zero is given by

$$E\left(e^{tX}\right) = e^{t.1}p(i) + e^{t.o}p(o)$$

$$= \theta e^{t} + (1-\theta)$$

$$= 1+\theta(e^{t}-1)$$

Also, note that

mean is $\mu_1' = \left.\frac{dE(e^{tX})}{dt}\right|_{t=0}$

$$= \left[o+\theta(e^{t}-0)\right]_{t=0}$$

$$= \theta$$

Similarly, variance of X can be obtained by usig

$$\mu_2 = \mu_2' - \mu_1'^2$$

Here , $\mu_2' = \left.\frac{d^2E(e^{tX})}{dt^2}\right]_{t=0}$

$$= \left[\theta e^{t}\right]_{t=0}$$

$$= \theta$$

$$\therefore \mu_2 = \theta - \theta^2$$

$= \theta(1-\theta)$, giving the vaniance

The distribution of X given in this example is called Bernoulli distribution.

Example (c) : Let X be a continuous random variable with the pdf

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

This is called the Cauchy distribution which has the property that the mgf does not exist for this distribution. In other words,

$$E(e^{tX})=\frac{1}{\pi}\int_{-\infty}^{\infty}\frac{e^{tx}dx]}{1+x^2}\text{ does not exist}$$

Consequently, the mean of the distribution does not exist, while the median is at x = 0.

One very useful property of mgf is that the mgf uniquely determines a distribution function (d.f) and, conversely, if the mgf exists, it is unique.

1.5 Markov Chain, Chebyshev's Inequality

1.5.1 Markov chain—Concept

Markov chain generalizes the notion of statistical independence. Let the random variables X_n (n = 0,1,2,...) be defined as follows :

X_n = j, if event E_j (j = 1,2,....) is the outcome of the n^{th} trial.

The trials are said to be independent statistically if

$$P\left(X_n = j|X_0=i_o,...,X_{n-1}=i_{n-1}\right)=P(X_n=j),\text{ for all n } (\bigstar)$$

In Markov chain, it is assumed that the outcome of each new trial depends on the outcome of the directly preceding trial but is independent of the outcomes of all former trials.

Thus, the sequence $\{X_n\}$ constitutes a Markov chain if for all values of n and also for all values of X_0, X_1, X_2 ..... the condition

$$P\left(X_n=j|X_0=i_0,X_1=i_1,...,X_{n-1}=i_{n-1}\right)$$

$$=\quad P\left(X_n=j|X_{n-1}=i_{n-1}\right)\quad\rightarrow(**)$$

In this case, $\{X_n\}$ is called a Markov chain of order one.

Thus, the notion of Markov chain is arrived at by assuming that the outcome of each new trial depends on the outcome of the directly preceeding trial but is independent of the outcomes of all former trials.

Clearly, (**) is an extension of (*) to the case of one-step dependence.

As an illustration of a Markov chain, imagine a particle moving back and forth along the real line in discrete steps. Imagine also that the particle begins its motion at the point O and that it can move only one unit of distance to the left or to the right at each step. In particular, after its first stip it will be located at either x =1 or $x = -1$. Suppose that the probability that it moves to the right at each step is ‘p’ and the probability that it moves to the left is ‘1–p’ where 0< p<1. We might ask ‘what is the probability that the paricle will remain in some interval centered about the point ‘O’ after n steps?’

Example (a) : Consider a sequence of trial with a fair coin

i.e., P(Head) = $\frac{1}{2}$ = P(Tail)

Let X_n = number of heads in the first n trials.

If $X_{n-1} = k$, then X_n=K or $k + 1$

Thus, $P\left\{X_n = x_{in} \middle| X_{n-1} = k, X_{n-2} = x_{i_{n-2}}, ... X_1 = x_{in}\right\}$

$$= \begin{cases} \frac{1}{2}, if\ x_{in} = k\ or\ k+1 \\ 0, \text{otherwise} \end{cases}$$

It follows that $\{X_n\}$ is a Markov chain. Also, conditional probability is independent of n. Such a sequence is said to have stationary transition probabilities.

Example (b) :

Let $\{X_n\}$ be a sequence of independent r.v's with pmf

$P(\mathrm{X}_n = 1) = p, P(\mathrm{X}_n = -1) = 1 - p$, $0 < \mathrm{p} < 1$

Write $S_n = \sum_{i=1}^{n} X_i$. Show that $\{S_n\}$ is Markov sequence.

Solution : Note that $S_n = X_n + \sum_{i=1}^{n-1} X_i \quad = \mathrm{X_n + S_{n-1}}$

To show that $P\left(S_n = s_n \middle| s_{n-1} = s_{n-1}, S_{n-2} = S_{n-2}, \ldots, S_1 = s_1\right)$

$$= \quad P\left(S_n = p_n \middle| S_{n-1} = s_{n-1}\right)$$

Where s_1, s_2 ... s_{n-2}, s_{n-1} are fixed numbers.

Clearly, $P\left(S_n = s_n \middle| S_{n-1} = s_{n-1}\right)$

$$= P\left(X_n + S_{n-1} = s_n \middle| S_{n-1} = s_{n-1}\right)$$

$$= P\left(X_n = s_n - s_{n-1}\right)$$

$$= \begin{cases} p, \, if \;\; s_n = s_{n-1} + 1 \\ 1 - p, \, if \;\; s_n = s_{n-1}{}^{-1} \end{cases}$$

Thus, $P\left(S_n = s_n \middle| S_{n-1} = s_{n-1}\right)$ depends on s_{n-1} only, and not on s_1, $s_{2,}$ ...,s_{n-2}

So, $P\left(S_n = s_n \middle| S_{n-1} = s_{n-1}, S_{n-2} = s_{n-2,}\ldots, S_1 = s_1\right)$

$$= P\left(S_n = s_n \middle| S_{n-1} = s_{n-1}\right)$$

which shows that $\{S_n\}$ is a Markov sequence.

It can be seen that, by multiplicative rule of probability,

$$P\left(X_o = i_o, X_1 = i_{1,} \ldots, X_{n+1} = i_{n+1}\right)$$

$$= P\left(X_0 = i_o\right).P\left(X_1 = i_1 \middle| X_o = i_o\right).P\left(X_2 = i_2 \middle| X_0 = i_0, X_1 = i_1\right)$$

$$\ldots P\left(X_{n+1} = i_{n+1} \middle| X_0 = i_0, X_1 = i_1, \ldots, X_n = i_n\right)$$

$$= P\left(X_0 = i_0\right)P\left(X_1 = i_1 \middle| X_0 = i_0\right)P\left(X_2 = i_2 \middle| X_1 = i_1\right)..P\left(X_{n+1} = i_{n+1} \middle| X_n = i_n\right),$$

for all n, if $\{X_n\}$ is a Markov sequence.

Here, $P(X_0 = i_0)$ is called the initial probability

and $P\left(X_j = i_j \middle| X_{j-1} = i_{j-1}\right), (j = 1, 2, 3, \ldots, n+1)$

are called the one-step transition probabilities of the Markav chain (M.C)

These probabilities determine the joint distribution of any finite number of r.v.'s of the M.C. sequence. Hence, we conclude that the probability deistribution of the M.C. is completly determined if we are given the initial probability distribution and one-step transition probabilities.

Transition probability matrix (t.p.m) :

If $P\left(X_{n+1} = i_{n+1} \middle| X_n = i_n\right)$ depends only on $\left(i_{n+1,} i_n\right)$.

and not on n, the transition probabilities are said to be stationary. In this case.

$$P\left(X_n = j \middle| X_{n-1} = i\right) = P\left(X_1 = j \middle| X_0 = i\right)$$

$$= p_{ij}, \text{ say, for } i,j \in I,$$

where I is the set of integers including zero.

If $P = (p_{ij})$ is called the one-step transition probability matrix. On account of probabiliy properties, we have.

$$0 \le p_{ij} \le 1 \ , \ \sum_j p_{ij} = 1.$$

Hence, the elements of P are non-negative and the row-sum equals unity. Any matrix, with such a property is called a stochastic matrix. The matrix will be called doubly stochastic if the column sums are also equal to unity.

n-step transition probability :

Let us denote

$P(X_n = j | X_0 = i)$ by $p_{ij}^{(n)}$ which equals

$P(X_{n+m} = j | X_m = i)$, $\forall m = 1, 2, 3, \ldots,$

for stationary transition probabilities. This is called an n-step transition probability, and the matrix whose $(i_{,}j)$ the element is $p_{ij}^{(n)}$ is denoted by $P^{(n)}$ and is called the transition probability matrix. It is also a stochastic n-step matrix n-step. For statinary press $p^{(n)} = p \times p \times \ldots .p = p^n$ i.e, nth power of one step TPM

For stationary process $p^{(n)}=p\times px\ldots p=p^n$ie. nth powers of one-step TPM

Example (c) : (Random walk and gambler's ruin problem) :

Consider a gambler with initial capital x laying against an adversary with initial capital $a - x$. In every play, the winner will get one unit of money from the loser if he has money. If p is the probability that the gambler will win and $q(=1-p)$, that he will lose, fortunes (X_n) of the gambler after n plays, form a Markov chain with one-step transition probabilities given by $P(X_{n+1} = j | X_n = i) = p_{ij} = \begin{cases} p, \text{if } j = i+1, \\ q, \text{if } j = i-1, \\ o, \text{otherwise} \end{cases}$

$(n = 0,1,2,\ldots,)$ $(i,j = 1,2$ --- $a-1)$, $p_{oo} = 1 = p_{aa.}$

Thus, the transition matrix is

$$P = (p_{ij}) = \begin{pmatrix} 1 & 0 & 0 & 0 & \ldots & 0 \\ q & 0 & p & 0 & \ldots & 0 \\ 0 & q & 0 & p & \ldots & 0 \\ \vdots & & & & & \\ 0 & 0 & 0 & 0 & \ldots & p \\ 0 & 0 & 0 & 0 & \ldots & 1 \end{pmatrix}$$

The space I = {0,1,2,...,a} is called the state space. The states 'o' and 'a' are called absorbing, since the M.C cannot move out once it reaches one of these states. Absorption at 'o' denotes the ruin of the gambler and at 'a' denotes the ruin of the adversary, hence the win of the gambler.

In general, any state *i*, for which $P(X_{n+1} = i | X_n = i) = p_{ii} = 1$ is called absorbing. All other entries in the ith row are equal to zero. The above problem is also the same as the random walk problem, gambler's fortune performing the "walk" of one-step to the right or to the left every unit of time, with "o" or "a" as the absorbing barriers and x as the initial position.

Example (d) :

Let $P = (p_{ij})_{n\times n}$ and Q = $(q_{ij})_{n\times n}$ be two transition probability matrices (tpm) corresponding to two M.C.'s. Then, the product PQ is also a transition probability matrix.

Solution : Here, $P = (p_{ij})_{n\times n.}$ and $Q = (q_{ij})_{n\times n.}$

Then, the product $PQ = (r_{ij})_{n\times n}$ say, where r_{ij} is given by

$$r_{ij} = \sum_{k=1}^{n} p_{ik} q_{kj}, \text{ i,j } = 1,2,\ldots\ldots,\text{n}$$

Since $p_{ij} \geq o \,\&\, q_{ij} \geq 0$, hence $r_{ij} \geq 0$

Also, $\sum_{j=1}^{n} r_{ij} = i^{th}$ row sum of PQ, $(i = 1,2,....,n)$

$$= \sum_{j=1}^{n}\sum_{k=1}^{n} p_{ik}q_{kj},$$

$$= \sum_{k=1}^{n}\sum_{j=1}^{n} p_{ik}q_{kj}$$

$$= \sum_{k=1}^{n} p_{ik}\left(\sum_{j=1}^{n} q_{kj}\right)$$

But $\sum_{j=1}^{n} q_{kj}$ is the sum of the k$^{\text{th}}$ row of Q, so that $\sum_{j=i}^{n} q_{kj} = 1$, as Q is a stochastic matrix.

Thus, finally, $\sum_{j=i}^{n} r_{ij} = \sum_{k=1}^{n} p_{ik} = 1$, as P is also a stochastic martix.

It can be easily checked that if P and Q are doubly stochastic (i.e., their column sums are also unity), then the product PQ is also doubly stochastic.

1.5.2 Chebyshev's inequality

Theorem 1.5.1

Let h(X) be a non-negative function of a random variable X such that $E(h(X))$ exists. Then, for every $\in > 0$, $P(h(X) \geq \in) \leq \frac{E(h(X))}{\in}$

To prove the result, we assume X to be discrete. The continuous case is treated similarly.

Let $P\{X = x_k\} = p_k$, k = 1,2,.....

Then $E(h(X))=\sum_{k=1}^{\infty} h\left(x_k\right) p_k$

$$=\left(\sum_A+\sum_{A^c}\right) h\left(X_k\right) p_{k,}$$

where $A=\{k: h(x_k) \geq \in\}, A^c=\{k: h\left(x_k\right)<\in\}$

Then $\quad E(h(X)) \geq \sum_A R(x_k) p_k$

$$\geq \in \sum_A P_k=\in P(h(X) \geq \in)$$

$$P(h(X) \geq E) \leq \frac{E h(X)}{\in}$$

Corollary : Choose $h(X)=(X-\mu)^2, \in=k^2 \sigma^2$, when

$\sigma^2=E(X-\mu)^2$, $\mu=E(X)$, and k > 0.

Hence, $P\left((X-\mu)^2 \geq \in\right) \leq \frac{\sigma^2}{k^2 \sigma^2}$

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

This is called Chebyshev's inequality.

Chebyshev's inequality given an upper bound to the tail probabilities of a distribution.

Example (a) : Let $P(X=0)=1-\frac{1}{k^2}$, $\quad P(X= \pm 1)=\frac{1}{2 k^2}$, when k > I, a constant.

Here $E(X) = 0.\left(1-\frac{1}{k^2}\right)+1.\left(\frac{1}{2k^2}\right)-1.\left(\frac{1}{2k^2}\right) = 0$

$$E\left(X^2\right)=1.\frac{1}{2k^2}+1.\frac{1}{2k^2}=\frac{1}{k^2}$$

$$V(X)=\frac{1}{k^2}-0^2=\frac{1}{k^2}, \quad \sigma=\sqrt{\frac{1}{k^2}}=\frac{1}{k}$$

Hence $P\{|X-\mu|\geq k\sigma\}$

$$= p\{|X|\geq 1\}=\frac{1}{2k^2}+\frac{1}{2k^2}=\frac{1}{k^2},$$

so that Chebyshev's inequality $P\{|X-\mu|\geq k\sigma\}\leq\frac{1}{k^2}$ is satisfied.

Example (b) : Let X be distributed with pdt f(x) = 1 if 0 < x < 1, and = 0 otherwise. Then

$$E(X)=\int_0^1 xf(x)dx=\frac{x^2}{2}\Bigg]_0^1=\frac{1}{2}$$

$$E(X^2)=\int_0^1 x^2 f(x)dx=\frac{1}{3}x^3\Bigg]_0^1=\frac{1}{3}$$

$$V(X)=\frac{1}{3}-\frac{1}{4}=\frac{1}{12}, \quad \sigma=\frac{1}{\sqrt{12}}$$

By Chebyshev's inequality,

$$P\left\{\left|X-\frac{1}{2}\right|\leq 2\sqrt{\frac{1}{12}}\right\}\geq 1-\frac{1}{2^2}=0.75.$$

But the actual probability is

$$P\left\{\left|X-\frac{1}{2}\right|\le 2\sqrt{\frac{1}{12}}\right\}=P\left\{\frac{1}{2}-\frac{1}{\sqrt{3}}\le X\le\frac{1}{2}+\frac{1}{\sqrt{3}}\right\}$$

= 1, as X lies in 0 < x < 1.

The means that Chebyshev's inequality gives good approximation to the actual value.

1.5.3 Characteristic function

The mgf does not always exist. A generating function which always exists is the characteristic function (c.f), written as $\phi(t)$, where

$\phi(t)=E\left(e^{itx}\right)$, where $i=\sqrt{-1}$, and t is the parameter.

The relation between c.f. and m.g.f is that

$$\phi(t)=E\left(e^{it^{X}}\right)$$

= M (it),

where M(t) is the mgf of X, i.e., M(t) = E(e^{tX})

Example (a) : Consider the normal distribution with the pdf

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

$$-\infty<\mu<\infty, \sigma^2>0\,.$$

The characteristic function of this distribution is a complex integral, which on simplification, comes out to be $\phi(t)=e^{it\mu-t^2\sigma^2/2}$.

The characteristic function about mean μ will be given by

$$\phi_\mu(t) = E(e^{it(X-\mu)}$$

$$= e^{-it\mu}E\left(e^{itX}\right)$$

Thus, for the normal distribution above, are have

$$\phi_\mu(t) = e^{-it\mu}.e^{it\mu-t^2\sigma^2/2}$$

$$= e^{-t^2\sigma^2/2}$$

Example (b) : Let X be a discrete random variable with pmf

$$p(x) = \overset{n}{\underset{x}{C}} y p^x q^{n-x}, \text{ x = 0,1,....,n, 0 < p < 1, p + q = 1}$$

The characteristic function of X is then given by

$$\phi(t) = E\left(e^{itX}\right)$$

$$= \sum_{x=0}^{n} e^{itX}\, {}^{n}C_x\, p^n q^{n-x}$$

$$= \sum_{x=0}^{n} \overset{n}{\underset{x}{C}} \left(pe^{it}\right)^x q^{n-x}$$

$$= \left(q + pe^{it}\right)^n$$

One important property of the characteristic function is that the characteristic function of the sum of a fixed number of independent random variables is the product of the characteristic functions of the summands.

The characteristic function uniquely determines the underdying distribution. Given a characteristic function, it is possible to get back he distribution by using what is called the 'inversion theorem.'

1.5.4 Probability generating function (pgf)

Another important and useful generating function is the probability generating function (pgf) applicable to the case of discrete variable only. More specifically, here we assume that the variables under consideration is the integer-valued random variable.

Let X be a random variable and let

$$P(X = k) = p_K,\ k = 0,1,2,....$$

with $\sum_{k=0}^{\infty} p_k = 1$

Then the pgf of X is given by

$$G(t) = \sum_{k=0}^{\infty} t^k P(X = k)$$

$$= E(t^x),\ (\text{t is the parameter})$$

Example (a) :

Consider the Poisson distribution with pmf

$$P(X = k) = e^{-\theta} \frac{\theta^k}{\underline{|k}},\ k = 0,1,2,...$$

Then pgf G(t) is given by

$$G(t) = E(t^x)$$

$$= \sum_{k=0}^{\infty} t^k \frac{e^{-\theta}\theta^k}{\underline{|k}}$$

$$= e^{-\theta} \sum_{k=0}^{\infty} \frac{(\theta t)^k}{\underline{|k}}$$

$$= e^{-\theta}.e^{\theta t}$$

$$= e^{-\theta(1-t)}\ ,\ |t|\leq|1.$$

Note that if $G'(t)$ and $G''(t)$ be the derivatives of G(t) with respect to t once and twice, respectively, then $G'(1)=E(X)$

and $G''(1)+G'(1)-(G'(1))^2=V(X)$

Also $\left.\frac{d^k G(t)}{dt^k}\right|_{t=0} = P(X=k)$, k =0,1,2,....

Another property of the pgf is that the pgf of the sum of two integer-valued random variables is the product of their pgf's, if they are independent.

Thus, $G_{X_1+X_2}(t)=E(t^{X_1+X_2})$

$$=E\left(t^{X_1}.t^{X_2}\right)$$

$$= E\left(t^{X_1}\right)E\left(t^{X_2}\right) \quad (\because\ X_1\ \&\ X_2\ \text{independent})$$

$$= G_{X_1}(t),G_{X_2}(t)$$

1.6 Summary

In this unit,we have studied the concept of probability along with its different meanings. The classical definition of probability, the frequentist approach as well as axiomatic approach are all considered. The limitations of the classical definition of probability are also mentioned.

The important concept of a random variable along with its distribution is also discussed. The ideas of moment generating function (m.g.f), probability generating function (p.g.f) and characteristic function (ch.f) are also given. The functional relationships between them are discussed.

1.7 Exercises

1. Let A^C be the complement of the set A. Then $P(A^C) = 1-P(A)$
2. For A_1 and A_2 events, not necessarily mutually exclusive, $P(A_1 \cup A_2) \le P(A_1)+P(A_2)$. Discuss the situation when the equality holds.
3. For two events A and B, $P(A \cap B) \le P(B)$. Discuss the case when equality holds.
4. Express $P\left(A^c \cap B^c \cap C^c\right)$ in terms of the probabilities of A, B and C and their intersection.
5. Check that the function f(x) defined as follows is a valid p.d.f—

$$f(x) = \begin{cases} x, 0 \le x \le 1. \\ 2-x, 1 \le x \le 2, \\ 0, \text{otherwise} \end{cases}$$

6. Show that if X is a random variable such that

$P(a \le x \le b) = 1$, then $a \le E(x) \le b$, $V(X) \le \frac{1}{4}(b-a)^2$

7. Check whether $f(x) = 1-|x|, -1 \le x \le 1,$ is a proper p.d.f of a random variable X.

Unit 2 ❑ Theoretical Distribution

Structure

2.0 Objectives

The followings are discussed here:

- Definitions of probability mass function (pmf)
- Definition of probability density function (pdf)
- discrete distributions
- Continuous distributions
- Bivaviate distribution
- Weak Law of Large Numbers (WLLN)
- Central Limit Theorem (CLT)

2.1 Introduction

In many scientific investigations, we are confronted with huge numerical data that arise from what are technically called populations. In statistics, these populations are characterzed by probability. This unit considers the different properties of these populations and their uses in inferential problems. Different aspects, mainly, means and variances of these distributions are studied here.

2.2 Discrete Distribution

2.2.1. Uniform distribution

This distribution will occur if the different values of the random variable happen to be equally likely. Thus, if a fair die is thrown, the variable which is the number of points coming on the uppermost face takes the values 1, 2, 3, ..., 6, which are equally probable.

The *pmf* is given by

$f(x) = \frac{1}{k}$, $x = a, a + h, \ldots\ldots, a + (k - 1)h$, where '$a$' and '$h$' are fixed real numbers and 'k' is a fixed positive integer.

Clearly, $f(x) \geq 0$ for all x, and

$$\sum_{x=a}^{a+(k-1)h} f(x) = \sum_{x=a}^{a+(k-1)h} \left(\frac{1}{k}\right) = \frac{1}{k} \times k = 1$$

The mean is $a + (k-1)\,h$ $E(x) = \sum_{x=a}^{a+(k-1)h} xf(x)$

$$= \frac{1}{k}\left\{\frac{(2a+(k-1)h)k}{2}\right\} = a + \frac{h(k-1)}{2}$$

Similarly, $V(x) = E\,(X - \mu)^2$, where $\mu = E(X) = E\,(X^2) - (E\,(X))^2$.

Note that variance of the set $a, a + h, ..., a + (k - 1)h$ is same as that of $0, h, 2h, ...,(k - 1)h$, since variance is unaltered by change of origin. Again, variance of the set $0, h, 2h, ..., (k - 1)h$ is h^2 times variance of $0, 1, 2,.....,k - 1$ which is $h^2 \frac{k^2-1}{12}$.

With this in mind, define $Y = \frac{X-a}{h}$, so that Y takes the values $0, 1, ..., (k - 1)$. Now,

$P(X = x) = P\,(Y = \frac{x-a}{h}) = \frac{1}{k}$, $Y = 0, 1, 2,....., (k - 1)$,

so that $E(Y) = \sum_{y=0}^{k-1} y \left(\frac{1}{k}\right)$

$$= \left(\frac{1}{k}\right)\cdot\frac{(k-1).k}{2} = \frac{k-1}{2}$$

$$\therefore \quad E(X) = a + hE(Y)$$

$$= a + \frac{h(k-1)}{2}, \text{ as already obtained.}$$

We know that $V(Y) = \frac{k^2-1}{12}$

$$\therefore\ V(X) = h^2V(Y) = h^2\ \frac{k^2-1}{12}, \text{ as already established.}$$

Therefore, standard deviation

$$= \sqrt{V(X)} = |h| \sqrt{\frac{k^2 - 1}{12}}$$

2.2.2. Binomial distribution

Repeated independent trials with two possible outcomes, a 'success' with probability p and a 'failure' probability $1 - p$, are called Bernoullian trials. The most familiar examples of such trials are tosses of a coin, where occurrence of 'head' may be termed as success and that of 'tail' as failure. Our aim is to obtain the probability of x heads (x = 1, 1, 2, ..., n) in n trials (n is a fixed pre-assigned number).

The *pmt* of a Binomial Distribution is given by

$$f(x) = \binom{n}{x} p^x q^{n-x}, \; x = 0, 1, 2, ..., n \quad 0 < p < 1, \; q = 1 - p$$

Define parameter : The parameters of a distribution are constants that define the distribution, such as its mean, variance etc.

The mean of the distribution is given by

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

$$= \sum_{x=0}^{n} x \frac{\underline{|n}}{\underline{|x}\underline{|n-x}} p^x q^{n-x}$$

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

$$= np \sum_{x'=0}^{n-1} \frac{\underline{|n-1} p^{x'} q^{n-1-x'}}{\underline{|x'} \; \underline{|n-1-x'}} \begin{pmatrix} put \\ x' = x - 1 \end{pmatrix}$$

$$= np(q + p)^{n-1} = np, \; (\because q + p = 1)$$

Probability generating function of X is

$$E(t^x) = \sum_{x=0}^{n} t^x \cdot \binom{n}{x} p^x q^{n-x} = (q+pt)^n = G(t), \text{ say}$$

Then $G'(t) = n(q+pt)^{n-1} \cdot p$

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

$\therefore G'(1) = np(q+p)^{n-1} = np = E(X)$, as already shown

and $G''(t) = np(n-1)\ (q+pt)^{n-2} \cdot p$

$\therefore\ G''(1) = n(n-1)p^2$

Finally, $V(X) = G''(1) + G'(1) - (G'(1))^2$

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

$= np - np^2$

$= np\ (1-p)$

$= npq$ (putting $q = 1-p$)

and standard deviation $= \sqrt{npq}$.

Theorem : Let X follow a binomial distribution with parameters n and p, written as $X\sim b\ (n, p)$, then as $n \to \infty$ and $p \to 0$ keeping $np = \lambda$ (finite), we have $\underset{n\to\infty, p\to 0}{\text{lim it}}$

$$b(n, p) = \frac{e^{-\lambda}\lambda^x}{\underline{|x}},\ x = 0, 1, 2, \ldots\ldots.np = \lambda$$

Proof : Denoting *pmf* of X by $b(n, p)$, we have

$b(n, p) = \binom{n}{x} p^x q^{n-x},\ x = 1, 1, 2, \ldots\ldots.,\ np = \lambda$

which can be rewritten as

$$b(n, p) = \frac{\underline{|n}}{\underline{|x}\,\underline{|n-x}}\ p^x q^{n-x}$$

$$= \frac{n(n-1)(n-2).....(n-x+1)}{\underline{|x}} p^x q^{n-x}$$

$$= \frac{n^x}{\underline{|x}}\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\ldots\left(1-\frac{x-1}{n}\right)\cdot\frac{\lambda^x}{n^x}\left(1-\frac{\lambda}{n}\right)^{n-x} \text{(putting } p = \lambda/_n)$$

$$\frac{\lambda^x}{\lfloor x}\left(1-\frac{\lambda}{n}\right)^n \cdot \left(1-\frac{\lambda}{n}\right)^{-x}, \text{ as } n \to \infty$$

$$\frac{\lambda^x}{\lfloor x} e^{-\lambda}, \text{ since}$$

$$\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^n = e^{-\lambda} \text{ and } \lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^{-x} = 1, \text{ as } x \text{ is fixed}$$

Thus, $b(n, p) \simeq e^{-\lambda}\frac{\lambda^x}{\lfloor x}, \; x = 0, 1, 2, \ldots.., \infty$

$= f(x)$, say.

Some examples satisfying $n \to \infty$, $p \to 0$, but $np = \lambda$(finite) :

(1) Number of misprints per page of a book.

(2) Number of defects in a given area of photographic plate,

(3) Number of telephone calls received by the telephone operator in a given duration during peak hours of the day.

Mode of a binomial distribution :

Consider a binomial distribution with *pmt*.

$f(x) = \binom{n}{x} p^x q^{n-x}, \; x = 0, 1, 2, \ldots\ldots, n$

Let M_0 be the mode of the distribution. Then

$f(M_0) \geq f(M_0 + 1)$

and $f(M_0) \geq f(M_0 - 1)$, M_0 being an integer.

Note that mode is the value of the variable at which the probability is the maximum.

Here, $\dfrac{f(x)}{f(x-1)} = \dfrac{\binom{n}{x} p^x q^{n-x}}{\binom{n}{x-1} p^{x-1} q^{n-x+1}}$

$$= \frac{\cancel{\underline{|n}}}{\underline{|x}\,\underline{|n-x}} \cdot \frac{\underline{|x-1}\,\underline{|n-x+1}}{\cancel{\underline{|n}}} \cdot \frac{p}{q} = \frac{(n-x+1)p}{xq}$$

$\therefore f(M_0) \geq f(M_0 + 1)$ implies

$$\frac{f(M_0+1)}{f(M_0)} \leq 1$$

or, $\dfrac{(n-(M_0+1)+1)p}{(M_0+1)q} \leq 1$

or, $(n - M_0)p \leq M_0 q + q$

or, $np - M_0 p \leq M_0 q + q$

or, $M_0(p + q) \geq np - q$

or, $M_0 \geq np - 1 + p = (n + 1)p - 1$

i.e., $M_0 \geq (n + 1)p - 1 \rightarrow (1)$

Also, $\dfrac{f(M_0)}{f(M_0-1)} = \dfrac{(n-M_0+1)p}{M_0 q}$

$\therefore\ f(M_0) \geq f(M_0 - 1)$

$\Rightarrow$ $(n - M_0 + 1)\, p \geq M_0 q$

or, $np - M_0 p + p \geq M_0 q$

or, $M_0(p + q) \leq (n + 1)p$

or, $M_0 \leq (n + 1)p \rightarrow (2)$

$(n + 1)p - 1$ $(n + 1)p$

Case 1. If $(n + 1)p$ is not an integer, there must be a unique integer between $(n + 1)p - 1$ and $(n + 1)p$, satisfying (1) and (2).

We choose this integer to be M_0 which will be the unique mode. Thus, the greatest integer contained in $(n + 1)p$, i.e., $M_0 = [(n + 1)p]$ is the unique mode.

Case 2. If $(n + 1)p$ is an integer, then the two modes are M_0 and $M_0 - 1$, where $M_0 = (n + 1)p$.

Example : Consider a binomial distribution with $n = 6$ and $p = 0{\cdot}5$.

Here, $(n + 1)p = (6 + 1)\frac{1}{2} = \frac{7}{2} = 3{\cdot}5$.

So, the unique mode is a number M_0 (integer) such that

$M_0 \geq 3{\cdot}5 - 1 = 2{\cdot}52{\cdot}5 \qquad 3 \qquad 3{\cdot}5$

and $M_0 \leq 3{\cdot}5$

Thus, $M_0 = 3$

On the other hand, if $n = 7$, $p = 0{\cdot}5$, then

$(n + 1)p = (7 + 1)\frac{1}{2} = 4$

In this case, the two modes are 3 and 4.

Note : The mean and variance of a binomial *r.v.* X can also be obtained by first obtaining the *mgf* of X and then differentiating the *mgf*, as explained in the previous chapter.

2.2.3. Negative Bionomial Distribution

Consider the succession of trials of a random experiment which results either in a 'success' of a failure with probabilities p and $1 - p$ respectively. Let us compute the probability of observing excatly r successes, where $r \geq 1$ is a fixed integer.

Let X denote the number of failures that precede the r^{th} success. Then $X + r$ is the total number of trials needed to produce r success. This will happen if and only if the last trial results in a success and among the previous $r + x - 1$ trials there are exactly x failures. Since the trials are all independent, we have

$$P(X = x) = \binom{x+r-1}{x} p^r (1 - p)^x,\ x = 0,\ 1,\ 2,........$$

Equivalently,

$$P(X = x) = \binom{-r}{x} p^r (-q)^x,\ x = 0,1,\ 2,........,\ q = 1 - p.$$

Note that $\sum\limits_{x=0}^{\infty} p(X = x) = p^r \sum\limits_{x=0}^{\infty} \binom{-r}{x} (-q)^x = p^r(1 - q)^{-r} = p^r p^{-r} = 1.$

The *mgf* of the distribution is clearly

$M(t) \ = E(e^{tx})$

$$= \sum_{x=0}^{\infty} e^{tx} \cdot \binom{-r}{x} p^r (-q)^x$$

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

$$= p^r (1 - qe^t)^{-r}, \text{ for } qe^t < 1.$$

The mean and variance can be obtained in a number of ways, for example, by using *mgf.*

Here, $M(t) = p^r (1 - qe^t)^{-r}$

$\therefore M'(t) = p^r (1 - qe^t)^{-r-1}(-r) (0 - qe^t)$

so that $M'(0) = p^{-1} (-r)(-q) = \dfrac{rq}{p} = E(X)$, the mean.

Also, $M''(t) = (+rp^r)\, q\, [e^t(1 - qe^t)^{-r-1} + e^t (\in r - 1)(1 - qe^t)^{-r-2}(-qe^t))$

$= qrp^r [e^t (1 - qe^t)^{-r-1} + e^t qe^t(r + 1) (1 - qe^t)^{-r-2})$

so that $M''(0) = qrp^r [p^{-r-1} + q(r + 1) (p)^{-r-2}]$

$$= \frac{rqp^r}{p^{r+1}} [1 + q\,(r + 1)p^{-1}]$$

$$= \frac{rq}{p}\left[1 + \frac{q(r+1)}{p}\right] = \mu_2'$$

Thus variance $V(X) = \dfrac{rq}{p} + \dfrac{(rq)^2}{p} + r\dfrac{q^2}{p^2} - \dfrac{r^2q^2}{p^2}$

$$= \frac{rqp + rq^2}{p^2}$$

$$= \frac{rq(p+q)}{p^2} = \frac{rq}{p^2}$$

so that standard deviation is $\sqrt{\frac{rq}{p^2}} = \frac{\sqrt{rq}}{p}$.

The probability generating function (*pgf*) $G(t) = E(t^X)$ can be obtained by replacing e^t by t in *mgf*. Thus,

$G(t) = p^r (1 - qt)^{-r}, |t| \leq 1.$

2.2.4. Geometric distribution

Let X be the number of failures preceding the first success, say, occurrence of a head, in a series of independent tosses of a coin with P (Head) $= p$ and P (Tail) $= 1 - p = q$, say. The resulting distribution is called a geometric distribution or waiting time distribution.

The *pmf* of the distribution is given by

$f(x) = P(X = x) = pq^x, \ x = 0, 1, 2,.........$

The distribution is obtained as a special case of Negative Binomial Distribution with $r = 1$.

Consequently, the *mgf* and mean and variances are

$$M(t) = \frac{p}{1-qe^t}$$

$$E(X) = \frac{q}{p}$$

$$V(X) = \frac{q}{p^2}.$$

2.2.5. Poisson Distribution

This distribution can be obtained as a limiting case of binomial distribution. Consider a binomial distribution with *p.m.f.*

$f(x) = \binom{n}{x} p^x q^{n-x}, \ x = 0, 1, 2.........$

In many practical situations, n is very large while p is exceedingly small so that $np = \lambda$, a finite constant. Then the above *pmf* can be approximated as

$$f(x) = e^{-\lambda}\frac{\lambda^x}{\underline{|x}}, \; x = 0,\ 1,\ 2,\ldots\ldots$$

This distribution is known as the Poisson distribution.

Define parameter

The mean and variance of X can be obtained either by *mgf* or *pgf*. The *pgf* of X is given by

$$G(t) = E(t^X)$$

$$= \sum_{x=0}^{\infty} t^x \quad \frac{e^{-\lambda}\lambda^x}{\underline{|x}}$$

$$= e^{-\lambda}\sum_{x=0}^{\infty}\frac{(\lambda t)^x}{\underline{|x}}$$

$$= e^{-\lambda}\cdot e^{\lambda t} = e^{\lambda(t-1)}$$

$$\therefore G'(t) = e^{\lambda(t-1)}\cdot\lambda(1-0)$$

$$= e^{\lambda(t-1)}\cdot\lambda,$$

and $G''(t) = \lambda e^{\lambda(t-1)}\cdot\lambda(1-0)$

$$= \lambda^2 e^{\lambda(t-1)},$$

so that mean $= G'(1) = \lambda,$

and variance $= G''(1) + G'(1) - (G'(1))^2$

$$= \lambda^2 + \lambda - \lambda^2 = \lambda$$

Thus, for the Poisson distribution, both mean and variance are equal to λ.

Additive Properties of Binomial and Poisson random variables :

(1) Let X_1 and X_2 be independent Binomial random variables with parameters (n, p) and $(n_2\ p)$, respectively. Then $X_1 + X_2$ follows a binomial distribution with parameters $(n_1 + n, p)$.

(2) Let X_1 and X_2 be independent Poisson random variables with Parameters λ_1 and λ_2, respectively. Then the sum $X_1 + X_2$ follows a Poisson distribution with parameter $\lambda_1 + \lambda_2$.

2.3 Continuous Distributions

2.3.1 Uniform Distribution

The uniform distribution (also called rectangular distribution) has the same probability-density at all values throughout the range of the variable *X*. The *pdf* of *X* is defined, for some constants α & β, as

$$f(x) = \frac{1}{\beta-\alpha}, \text{ if } \alpha < x < \beta$$

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

The *pdf* can be shown graphically as follows :

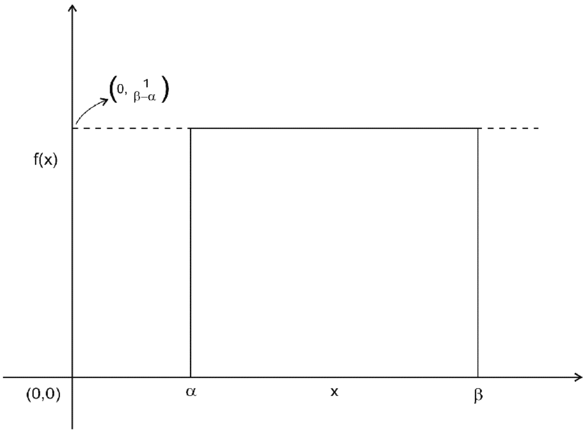


Fig. 2.1 A uniform distribution with range (α, β)

The mean is

$$E(X) = \int_{\alpha}^{\beta} xf(x)dx = \frac{1}{\beta-\alpha}\left[\frac{x^2}{2}\right]_{\alpha}^{\beta}$$

$$= \frac{1}{2(\beta-\alpha)}[\beta^2 - \alpha^2] = \frac{\alpha+\beta}{2},$$

$$E(X)^2 \quad = \int_{\alpha}^{\beta} x^2 f(x)dx = \frac{1}{\beta-\alpha}\left[\frac{x^3}{3}\right]_{\alpha}^{\beta}$$

$$= \frac{1}{3(\beta-\alpha)}\,[\beta^3 - \alpha^3] = \frac{\beta^2+\alpha\beta+\alpha^2}{3}$$

$$\therefore\ V(X) \quad = E(X^2) - (E(X))^2,$$

$$= \frac{\beta^2+\beta\alpha+\alpha^2}{3} - \frac{(\alpha+\beta)^2}{4}$$

$$= \frac{4\beta^2+4\alpha\beta+4\alpha^2-3\alpha^2-3\beta^2-6\alpha\beta}{12}$$

$$= \frac{(\beta-\alpha)^2}{12}$$

2.3.2. Exponential Distribution

The *pdf* of an exponentially distributed random variable X is

$$f(x) = \begin{cases} \beta^{-1}e^{-e/\beta}, x>0, \beta>0 \\ 0, \text{ otherwise} \end{cases}$$

If we take $\theta = \beta^{-1}$, then the above *pdf* can also be written as

$$f(x) = \begin{cases} \theta e^{-\theta x}, x>0, \theta>0 \\ 0, \text{ otherwise} \end{cases}$$

Define parameter

The *mgf* of X is

$$M(t) = E(e^{tx})$$

$$= \theta \int_{0}^{\infty} e^{tx}\, e^{-\theta x}\, dx$$

$$= \theta \int_0^{\infty} e^{-(\theta-t)x}\, dx$$

$$= \theta \frac{\Gamma(1)}{\theta - t},$$

$$= \left(1-\frac{t}{\theta}\right)^{-1}, \ |t| < \theta.$$

Then $\quad M'(t) = (-1)\left(1-\frac{t}{\theta}\right)^{-2}\left(-\frac{1}{\theta}\right)$

and $M''(t) = \left(\frac{1}{\theta}\right)(-2)\left(1-\frac{t}{\theta}\right)^{-3}\left(-\frac{1}{\theta}\right)$

$$= \frac{2}{\theta^2}\left(1-\frac{t}{\theta}\right)^{-3}$$

Thus, $E(X) = M'(0) = \frac{1}{\theta} = \beta,$

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

$$= M''(0) - \frac{1}{\theta^2}$$

$$= \frac{2}{\theta^2} - \frac{1}{\theta^2} = \frac{1}{\theta^2} = \beta^2.$$

The exponential distribution can also be obtained as a special case of a gamma distribution.

2.3.3. Erlangian Distribution

The standard form of a gamma distribution with shape parameter α (≥ 0) is given by

$$f(x) = \frac{1}{\Gamma(\alpha)} x^{\alpha-1} e^{-x}, \ x \geq 0.$$

If α is a positive integer, then the above distribution is called an Erlangian distribution. Note that for $\alpha = 1$, we have an exponential distribution.

Since α is a positive integer here, the *pdf* can be rewritten as

$$f(x) = \frac{1}{\lfloor\alpha - 1} x^{\alpha - 1} e^{-x}, \; x \geq 0, \; \alpha \geq 2.$$

Define parameter

Moment generating function (*mgf*) :

$$M(t) = E(e^{tX})$$

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

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

$$= \frac{1}{\Gamma(\alpha)} \frac{\Gamma(\alpha)}{(1-t)^{\alpha}}, \; t < 1.$$

$$= (1 - t)^{-\alpha}, \; t < 1.$$

$$= 1 + \alpha t + \frac{\alpha(\alpha+1)}{\lfloor 2} t^2 + \frac{\alpha(\alpha+1)(\alpha+2)}{\lfloor 3} t^3 + \ldots\ldots.$$

$$\therefore \quad \mu_1' = \text{mean} = \text{coefficient of } \frac{t'}{\lfloor 1} = \alpha,$$

$$\mu_2' = \text{coefficient of } \frac{t^2}{\lfloor 2} = \alpha(\alpha + 1).$$

$$\therefore \quad \mu_2 = \text{Variance} = \mu_2' - \mu_1'^2 = \alpha^2 + \alpha - \alpha^2 = \alpha$$

So, the mean and variance are equal.

Example (a) : For the Erlang distribution having *pmf*

$$f(x) = \frac{\left(\frac{x}{b}\right)^{c-1} e^{-\frac{x}{b}}}{b\left[(c-1)!\right]},\ x \geq 0,\ b > 0,\ c \text{ (integer)} > 0,$$

show that *mgf* is $(1 - bt)^{-c}$, $t > \frac{1}{b}$, mean = bc, variance = b^2c.

Solution : By definition, *mgf* is given by

$$M(t) = E(e^{xt}) = \frac{1}{b\left[(c-1)!\right]}\int_0^{\infty} e^{xt}\cdot\left(\frac{x}{b}\right)^{c-1} e^{-x/b}dx.$$

$$= \frac{1}{b\left[(c-1)!\right]}\cdot\frac{1}{b^{c-1}}\int_0^{\infty} x^{c-1} e^{-x\left(\frac{1}{b}-t\right)} dx$$

$$= \frac{1}{b\left[c-)!\right]}\cdot\frac{1}{b^{c-1}}\frac{\Gamma(c)}{\left(\frac{1}{b}-t\right)^{c}},\ t < \frac{1}{b}.$$

$$= \frac{1}{(1-bt)^{c}},\ t < \frac{1}{b}$$

$$= (1 - bt)^{-c},\ t < \frac{1}{b}.$$

Now, $M'(t) = (-c)\,(1-bt)^{-c-1}(-b) = bc(1 - bt)^{-(c+1)},$

and $M''(t) = -bc(c+1)(1-bt)^{-(c+1)-1}(-b)$,

$\therefore M'(0) = bc$, $M''(0) = b^2c\,(c + 1)$.

Hence, $\mu_1' = bc$, $\mu_2 = M''(0) - (M'(0))^2 = b^2c = b^2c\,(c + 1) - b^2c^2$.

2.3.4 Gamma Distribution

A random variable X with the following *pdf* is said to follow a gamma distribution:

$$f(x) = \frac{\theta^{\alpha}}{\Gamma(\alpha)} x^{\alpha-1} e^{-\theta x}, \; x > 0 \; ; \; \alpha, \theta > 0, \text{ Define parameter.}$$

and, = 0, otherwise.

$$\text{Mean } E(X) = \int_0^{\infty} x f(x) \, dx$$

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

$$= \frac{\theta^{\alpha}}{\Gamma(\alpha)} \cdot \frac{\Gamma(\alpha+1)}{\theta^{\alpha+1}} = \frac{\alpha}{\theta},$$

$$\text{and } V(X) = E(X^2) - (E(X))^2,$$

$$\text{where } E(X^2) = \frac{\theta^{\alpha}}{\Gamma(\alpha)} \int_0^{\infty} x^{\alpha+1} e^{-\theta x} dx$$

$$= \frac{\theta^{\alpha}}{\Gamma(\alpha)} \frac{\Gamma(\alpha+2)}{\theta^{\alpha+2}}$$

$$= \frac{(\alpha+1)(\alpha)}{\theta^2}$$

$$= \frac{\alpha^2 + \alpha}{\theta^2}$$

$$\therefore \quad V(X) = \frac{\alpha^2 + \alpha}{\theta^2} - \frac{\alpha^2}{\theta^2} = \frac{\alpha}{\theta^2}.$$

The *mgf* about *zero* is

$M(t) = E\,(e^{tX})$

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

$$= \frac{\theta^{\alpha}}{\Gamma(\alpha)} \frac{\Gamma(\alpha)}{(\theta - t)^{\alpha}}$$

On further simplification, we have

$$M(t) = \left(1 - \frac{t}{\theta}\right)^{-\alpha}, \ |t| < \theta,$$

The mean and variance can also be obtained by differentiating and putting $t = 0$, as discussed earlier.

2.3.5. Beta Distribution

A continuous random variable is said to follow a beta distribution with parameters m and n if its *pdf* is given by

$$f(x) = \frac{1}{B(n,m)} x^{m-1} (1-x)^{n-1}, \ 0 < x < 1, \ m, \ n > 0.$$

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

Here, $$E(X) = \int_0^1 x f(x)\, dx$$

$$= \frac{1}{B(n,m)} \int_0^1 x^m (1-x)^{n-1} dx$$

$$= \frac{B(m+1,n)}{B(n,m)}$$

$$= \frac{\Gamma(m+1)\Gamma(n).\Gamma(m+n)}{\Gamma(m+n+1)\Gamma(m)\Gamma(n)} = \frac{m}{m+n}$$

and $$E(X^2) = \frac{1}{B(m,n)} \int_0^1 x^{m+1} (1-x)^{n-1}\, dx$$

$$= \frac{B(m+2,n)}{B(m,n)}$$

$$= \frac{\Gamma(m+2)\Gamma(n)}{\Gamma(m+n+2)} \cdot \frac{\Gamma(m+n)}{\Gamma(m)\Gamma(n)} = \frac{(m+1)m}{(m+n+1)(m+n)}$$

Finally, on simplification,

$$V(X) = E(X^2) - (E(X))^2 = \frac{mn}{(m+n)^2(m+n+1)}$$

2.3.6. Normal Distribution

A continuous random variable X is said to follow a normal distribution if its *pdf* is given by

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/(2\sigma^2)}, \quad -\infty < x < \infty, \ -\infty < \mu < \infty, \ \sigma^2 > 0.$$

The constants μ and σ are the parameters of the distribution. The distribution is symmetric about μ.

Below we show that $E(X) = \mu$ and $V(X) = \sigma^2$

The *mgf* about zero is

$$\frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{tx - \frac{(x-\mu)^2}{2\sigma^2}} dx$$

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

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

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

$$= e^{\mu t+\frac{1}{2}t^2\sigma^2}, \text{ (using the fact that } \left. \int_{-\infty}^{\infty} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}} dx = \sigma\sqrt{2\pi} \right)$$

The moments of all order exist and may be computed from the *mgf.* Thus,

$$M'(t) = e^{\mu t+\frac{1}{2}t^2\sigma^2}(\mu+t\sigma^2) = M(t)\,(\mu + t\sigma^2).$$

$$\therefore M'(0) = \mu, \text{ the mean}$$

and $M''(t) = M'(t)\cdot(\mu + t\sigma^2) + M(t)\,(\sigma^2)$

Hence, $M''(0) = M'(0)\cdot\mu + M(0)\sigma^2$

$$= \mu^2 + \sigma^2 \quad (\because M(0) = 1)$$

Finally, $V(X) = (\mu^2 + \sigma^2) - \mu^2 = \sigma^2$

Characteristic function :

The characteristic function $\phi(t)$ of X can be obtained from the *mgf* using the relation $\phi(t) = M(it),\ i = \sqrt{-1}$.

Thus, $\phi(t) = M(it)$

$$= e^{\mu(it)+\frac{1}{2}(it)^2\sigma^2}$$

$$= e^{i\mu t-\frac{t^2\sigma^2}{2}}$$

Standard normal distribution : If $\mu = 0,\ \sigma = 1$, then X is said to follow a standard normal distribution having the *pdf*

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

The corresponding cumulative distribution function (*cdf*) is given by

$$F(x) = \int_{-\infty}^{x} f(x)dx$$

$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-x^2/2} dx$, which have been extensively tabulated for practical purposes.

If $X \sim N(\mu, \sigma^2)$, a normal distribution with mean μ and variance σ^2, then $Z = \frac{X-\mu}{\sigma}$ follows $N(0, 1)$. This follows from the following facts :

(1) $X = \mu + \sigma Z$,

(2) $E(X) = \mu + \sigma E(Z)$, so that $E(Z) = 0$

(3) $V(X) = \sigma^2 V(Z)$, so that $V(Z) = 1$

(4) $Z = \frac{1}{\sigma} \times - \frac{M}{\sigma}$, a linear function of the normals random variable X, hence Z also follows a normal distribution with mean zero and variance unity.

Example (a) :

By Chebyshev's inequality, if $E|X^2| < \infty$, $E(X) = \mu$ and $V(X) = \sigma^2$, then

$$P\{|X - \mu| > k\sigma\} \leq \frac{1}{k^2}, (k > 0).$$

For $k = 2$, we get

$$P\{|X - \mu| > 2\sigma\} \leq 0{\cdot}25,$$

and for $k = ,3$, we get

$$P\{|X - \mu| > 3\sigma\} \leq \frac{1}{9} = 0{\cdot}01235.$$

In this case, for $X \sim N(\mu, \sigma^2)$, we have

$$P\{|X - \mu| > k\sigma\} = P\{|Z| > k\}, Z \sim N(0, 1).$$

Using the normal table, it can be seen that

$P\{|Z| > 1\} = 0{\cdot}318$, $P\{|Z| > 2\} = 0{\cdot}046$

and $P\{ |Z| > 3\} = 0{\cdot}002$, while the estimated value is 0·01235 by Chebyshev's inequality.

Example (b) : Let $X \sim N(3, 4)$, then

$$P(2 \le X \le 5) = P\left(\frac{2-3}{2} \le \frac{X-3}{2} \le \frac{5-3}{2}\right)$$

$$= P\,(-0{\cdot}5 \le Z \le 1)$$

$$= P(Z \le 1) - P\,(Z \le -\,0{\cdot}5)$$

$$= P(Z \le 1) - P\,(Z \ge 0{\cdot}5)$$

(due to symnetry about zero)

$$= (1 - P\,(Z > 1) - 0{\cdot}3085$$

$$= (1 - 0{\cdot}1587) - 0{\cdot}3085$$

$$= 0{\cdot}5328.$$

Normal distribution is extensively used in many branches of physical, biological and social sciences. This distribution has occupied a unique position in statistics due to the so-called central limit theorem (CLT). It has been seen that under some mild assumptions, such as large sample size, many distributions can be approximated by a normal distribution.

Normal distribution possesses a reproductive property in the sense that if X_1 and X_2 are normal random variables, then $X_1 + X_2$ is also normal even if X_1 and X_2 are not independent.

2.3.7. Log-normal Distribution

A variable X is said to have a log–normal distribution if $\log_e X$ is normally distributed. As $\log_e X$ goes from $-\infty$ to $+\infty$, X goes from 0 to ∞.

Let $\log_e X$ follow $N(\xi, \delta)$, then the *pdf* of X can be shown to be

$$\frac{1}{\delta x\sqrt{2\pi}} exp\left[-\frac{1}{2\delta^2}(\log_e x - \xi)^2\right],\ x > 0.$$

Unlike the normal distribution which symmetric about the mean, log-normal distribution is positively skewed.

The mean and variance of the distribution are

$E(X) = exp\ [\xi + \delta^2/_2]$

and $V(X) = \omega^2(\omega^2 - 1)\ exp\ (2\xi)$,

where $\omega = exp\ (\delta^2/_2)$.

This distribution has wide application in reliability theory where the underlying situation is far from symmetry.

2.4 Sampling Distributions

2.4.1. Central Chi-square Distribution

Introduction :

Chi-square distribution is one of the three basic sampling distributions. The other two are the t-distribution and F-distribution, respectively.

Sampling distribution is the basis of statistical analysis. When we suggest a value for the unknown parameter based on the sample observations, the statistical properties of that suggested value will be needed. For this, we are to obtain the distribution of that suggested value, also called statistic. By the term 'sampling distribution', we shall mean the probability distribution of a statistic.

While discussing these three sampling distributions, we shall assume in all the cases that the random sample, namely, $X_1, X_2, ..., X_n$ has been drawn from a normal population $N(\mu, \sigma^2)$.

Distinguish between statistic and parameter:

A statistic is a function based on the sample, and hence it is a random variable, while a parameter is a constant that characterises a distribution.

Definition of a χ^2 (chi-square) random variable :

The sum of squares of n mutually independent normal random variables having zero mean and unit variance is called a central χ^2 with n degrees of freedom (*d.f.*). This is sometimes written as χ_n^2.

The *pdf* of a χ^2 with *n d.f.* is given by

$$f(\chi^2) = \frac{1}{2^{n/2}\Gamma\left(\frac{n}{2}\right)} exp\left[-\frac{1}{2}\chi^2\right](\chi^2)^{\frac{n}{2}-1} \; \chi^2 > 0,$$

where $\Gamma(')$ is the gamma function.

The χ^2 - distribution is related to gamma distribution and the properties of this distribution follow from those of the gamma distribution.

The important properties of the χ^2- distribution are :

(1) $E(\chi^2) = n,$

(2) $V(\chi^2) = 2n,$

(3) $\frac{\chi^2 - n}{\sqrt{2n}}$ tends to $N(0, 1)$, as $n \to \infty$

(4) χ^2 - distribution is positively skewed,

(5) If χ_1^2 and χ_2^2 are two independent χ^2 with n_1 & n_2 *d.f.* respectively, the sum $\chi_1^2 + \chi_2^2$ is itself a χ^2 with $(n_1 + n_2)$ degrees of freedom.

(6) When it is non-central? (Pl. mention)

Uses : (a) To test an assumed variance ——Let $X_1, X_2.....X_n$ be a random sample from $N(\mu, \sigma^2)$, whose both μ and σ^2 are unknown. Then, $\sum_{i=1}^{n}\left(\frac{Xi - \bar{X}}{\sigma_0}\right)^2$ follows a χ^2 with $(n - 1)$ degrees of freedom under $H_0 : \sigma = \sigma_0$.

(b) Chi-square distribution can also be used to obtain confidence interval for the unknown variance σ^2 for specified confidence co-efficient.

2.4.2. Central *t*-distribution

A *t*-statistic with *n* degrees of freedom is defined as $t = \dfrac{X}{\sqrt{\dfrac{Y^2}{n}}}$,

where $X \sim N(0,\ 1)$, Y^2 is a central $\aleph^2$ with *n* degrees of freedom and X & Y^2 are independent.

The *pdf* of a *t* statistic with *n* degrees of freedom is given by

$$f(t) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(\frac{n}{2}\right)}\left(1+\frac{t^2}{n}\right)^{-\frac{n+1}{2}}, \; -\infty < t < \infty .$$

Some properties of the distribution :

(1) The distribution is symmetrical about zero.

(2) Mean of the distribution is zero.

(3) If $n > 2$, the distribution has variance $n(n - 2)$.

(4) The statistic *t* follows asymptotically a normal distribution.

(5) We get a non-central *t* if E(x) is different from zero

Uses : (1) To test an assumed mean : If $X_1, X_2, \ldots\ldots, X_n$ *iid* $\sim N(\mu, \sigma^2)$, μ & σ^2 both unknown, then the statistic $\dfrac{\sqrt{n}(\bar{X}-\mu_0)}{S}$ (when S is the sample standard deviation with divison $n - 1$) follows a *t*-distribution with $n - 1$ *d.f.* under $H_0 : \mu = \mu_0$.

(2) To obtain the confidence interval for the mean when variance is unknown.

(3) To test equality of two means when the populations have same variance.

2.4.3. Central F-distribution

An *F*-statistic with $(n_1,\ n_2)$ degrees of freedom is given by

$$F = \frac{X_1^2/n_1}{X_2^2/n_2} = \frac{n_2}{n_1}.\frac{X_1^2}{X_2^2},$$

where X_1^2 and X_2^2 are independently distributed central χ^2 random variable with n_1 and n_2 *d.f.*, respectively.

The *pdf* of F is given by

$$f(F) = \frac{\Gamma\left(\frac{n_1+n_2}{2}\right)\left(\frac{n_1}{n_2}\right)^{n_1/2}}{\Gamma\left(\frac{n_1}{2}\right)\Gamma\left(\frac{n_2}{2}\right)} \frac{F^{\frac{n_1-2}{2}}}{\left(1+\frac{n_1}{n_2}F\right)^{\frac{n_1+n_2}{2}}}, \; 0 < F < \infty.$$

Properties:

(1) $E(F) = \frac{n_2}{n_2-2}$, (if $n_2 > 2$).

(2) $V(F) = \frac{2n_2^2(n_1+n_2-2)}{n_1(n_2-2)^2(n_2-4)}$, (if $n_2 > 4$)

(3) For $n_1 = 1$, $F = t^2$, t being a t-statistic with n_2 *d.f.*

(4) The statistic F follows a beta distribution

(5) When it is non-central : (Pl. mention)

Uses :

(1) To test equality of the variances of two normal populations.

(2) To obtain confidence interval for the variance ratio.

(3) To test equality of several means as in the case of analysis of variance.

2.5 Bivariate Distribution

2.5.1. Bivariate normal Distribution

In unit 2 of chapter 1, we have given the basic ideas of bivariate distributions in general along with the marginal and conditional distributions. Here, we consider the case of bivariate normal distribution.

The joint *pdf* of the pair (x, y) is said to be of the bivariate normal form if it is given by

$$f(x, y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-e^2}} exp\left[-\frac{1}{2(1-\rho^2)}\left\{\left(\frac{x-\mu_x}{\sigma_x}\right)^2\right.\right.$$

$$\left.\left.-2\rho\left(\frac{x-\mu_x}{\sigma_x}\right)\left(\frac{y-\mu_y}{\sigma_y}\right)+\left(\frac{y-\mu_y}{\sigma_y}\right)^2\right\}\right]$$

Where $\mu_x = E(X)$, $\mu_y = E(Y)$, $\sigma_x{}^2 = V(X)$, $\sigma_y{}^2 = V(Y)$ and $\rho =$ correlation co-efficient between X and Y, $|\rho| \leq 1$.

In brief, we write $(X, Y) \sim N_2\ (\mu_x, \mu_y; \sigma_x{}^2, \sigma_y{}^2; \rho)$

The role of the correlation co-efficient :

If $\rho = 0$, then $f(x, y)$ can be written as

$$f(x, y) = \frac{1}{2\pi\sigma_x\sigma_y} exp\left[-\frac{1}{2}\left\{\left(\frac{x-\mu_x}{\sigma_x}\right)^2+\left(\frac{y-\mu_y}{\sigma_y}\right)^2\right.\right],$$

$$= \frac{1}{\sigma_x\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu_x}{\sigma_x}\right)^2} \cdot \frac{1}{\sigma_y\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{y-\mu_y}{\sigma_y}\right)^2}$$

$= g(x), h(y)$, (say),

where $g(x)$ and $h(y)$ are the marginal distribution of X and Y, respectively. This shows that in this case X & Y are independently distributed.

The conditional distribution of Y, given $X=x$, is univariate normal with mean and variance given by

$$E(Y/x) = \text{regression of } Y \text{ on } x$$

$$= \mu_y + \rho\frac{\sigma_y}{\sigma_x}(x - \mu_x),$$

and $V(Y \mid x) = \sigma_y{}^2(1-\rho^2)$.

In other words, $Y \mid x \sim N\left(\mu_y + \rho\frac{\sigma_y}{\sigma_x}(x - \mu_x),\ \sigma_y{}^2(1-\rho^2)\right)$.

Similarly, the conditional distribution of X, givey $Y=y$, is given by a univariate normal distribution with mean and variance given by

$$E(X/y) = \mu_x + \rho\frac{\sigma_x}{\sigma_y}(y - \mu_y),$$

$$V(X/y) = \sigma_x{}^2(1-\rho^2).$$

In other words, $X/Y \sim N(\mu_x + \rho\frac{\sigma_x}{\sigma_y}(y - \mu_y)),\ \sigma_x{}^2(1-\rho^2))$.

Note that $E(X/y)$ is the regression of X on y.

Thus, both the regression equations are linear.

The joint *mgf* of X and Y in case of bivariate normal distribution is given by

$$M(t_1, t_2) = exp\ [t_1\mu_x + t_2\mu_y + \tfrac{1}{2}(t_1{}^2\sigma_x{}^2 + t_2{}^2\sigma_y{}^2 + 2\rho\,\sigma_x\sigma_y\,t_1t_2)],$$

where t_1 and t_2 are parameters.

Independence of X and Y for $\rho = 0$ follows from $M(t_1, t_2)$ as in this case $M(t_1, t_2)$ reduces to the product of two *mgf's* of $N(\mu_x, \sigma_x{}^2)$ and $N(\mu_y, \sigma_y{}^2)$.

2.5.2. Weak law of large numbers (WLLN)

Let $\{X_n\}$ be a sequence of random variables and let $S_n = \sum_{k=1}^{n} X_k$, $n = 1, 2.....$ we say that $\{X_n\}$ obeys the weak law of large numbers (WLLN) with respect to the sequence of constants $\{B_n\}$, $B_n > 0$, $B_n \uparrow \infty$, if there exists a sequence of real constants A_n such that $B_n^{-1}(S_n - A_n) \xrightarrow{P} 0$ as $n \to \infty$. A_n are called centering constants, and B_n norming constants.

Example (a) : Let X_1, X_2.......be *iid* random variables with common distribution $b(1, p)$, binomial with parameter 1 and p.

Then $E(X_i) = p$, $V(X_i) = p(1 - p)$

and we have

$$E\left(\frac{S_n}{n}\right) = \frac{E(S_n)}{n} = \frac{np}{n} = p$$

$$V\left(\frac{S_n}{n}\right) = \frac{V(S_n)}{n^2} = \frac{npq}{n^2} = \frac{pq}{n} \to 0, \text{ as } n \to \infty.$$

(Note that $S_n = \sum_{k=1}^{n} X_k$ follows Binomial distribution with parameters n and p)

The above result implies that $\frac{S_n}{n} \xrightarrow{P} p$, as $n \to \infty$

So, here $B_n = n$, and centering constants

$A_n = E(S_n) = np$.

Thus, we have proved that $\frac{S_n - np}{n} \xrightarrow{P} 0$, so that WLLN obeys in this example.

Example 1 : Let x_i assume two values i and $-i$ with equal probabilities. Show that WLLN cannot be applied to the sequence x_1, x_2,.....,x_n,.....

Solution : We have been given,

for each i, $p(x_i = i) = \frac{1}{2} = p\ (x_i = -i)$

$(i = 1, 2, 3, ...)$

$\therefore E(x_i) = i \cdot \left(\frac{1}{2}\right) + (-i) \cdot \frac{1}{2} = 0$

Define $S_n = \sum_{i=1}^{n} x_i$

Also, $V(x_i) = E(x^2_i) - (E\ (x_i))^2, 2$

$= i^2 \cdot \frac{1}{2} + i^2 \cdot \frac{1}{2}$

$= i^2.$

$\therefore E(S_n) = 0$, and $V(S_n) = \sum_{i=1}^{n} i^2 = \frac{n(n+1)(2n+1)}{6}$

Note that $\frac{V(S_n)}{n^2} = \frac{(n+1)(2n+1)}{6n} \not\to 0$, as $n \to \infty$

Therefore, WLLN does not hold for the sequence $\{x_n\}$.

An alternative definition of WLLN :

The WLLN is said to held for the sequence $\{X_n\}$ of $\pi.v.$'s. if $\left\{\frac{S_n}{n}\right\}$, where $S_n = \sum_{i=1}^{n} X_i$, converges in probability to a constant. For this, we need to check the following conditions :

(1) $\frac{Var(S_n)}{n^2} \to 0$,

(2) $\frac{E(S_n)}{n} \to C$, a constant.

Example 2. If x_i can take two values i^{α} and $-i^{\alpha}$ with equal probabilities. Then show that WLLN can be applied to the sequence $\{x_n\}$, if $\alpha < \frac{1}{2}$.

Solution : Here, $E(x_i) = \frac{1}{2}i^{\alpha} + \frac{1}{2}(-i^{\alpha}) = 0$

and $V(x_i) = E(x_i^2) - (E(x_i))^2$

$$= \frac{1}{2}i^{2\alpha} + \frac{1}{2}i^{2\alpha}$$

$$= i^{2\alpha}$$

Define $S_n = \sum_{i=1}^{n} x_i$

then $V(S_n) = \sum_{i=1}^{n} V(x_i)$, (assuming $x_1, x_2, \ldots\ldots, x_n$ to be independent)

$$= \sum_{i=1}^{n} i^{2\alpha}$$

$$= 1^{2\alpha} + 2^{2\alpha} + 3^{2\alpha} + \ldots\ldots\ldots + n^{2\alpha}$$

$$\simeq \int_0^n x^{2\alpha}dx \text{, (by Euler's summation formula)}$$

$$= \left.\frac{x^{2\alpha+1}}{2\alpha+1}\right]_0^n$$

$$= \frac{n^{2\alpha+1}}{2\alpha+1}$$

$$\therefore \frac{V(S_n)}{n^2} = \frac{n^{2\alpha+1}}{(2\alpha+1)n^2}$$

$$= \frac{n^{2\alpha-1}}{2\alpha+1} \to 0, \text{ as } n \to \infty$$

Provided $2\alpha - 1 < 0$,

or, $\alpha < \frac{1}{2}$.

Thus, for $\alpha < \frac{1}{2}$, the WLLN holds for the sequence $\{x_n\}$.

Khintchine's WLLN :

Let $\{X_n\}$ be a sequence of independently and identically distributed random variables with mean

$E(X_i) = \mu, \forall i$, Then

$$\bar{X}_n \xrightarrow{p} \mu, \text{ i.e., } \bar{X}_n \text{ tends to } \mu \text{ in probability,}$$

$$\text{where } \bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

This is called Khintchine's WLLN.

Symbolically, we write $\bar{X}_n \overset{P}{\rightarrow} \mu$, to mean the following :

A sequence of $\pi.v.$'s $\{X_n\}$ is said to converge in probability to X_0, if, for given $\in$ and δ, both > 0, there exists a N such that

$$P\ \{|X_n - X_0| > \in\} < \delta, \text{ for all } n \geq N,$$

$$\text{i.e., } \lim_{n\to\infty} P\{|\ X_n - X_0\ | > \in\} = 0, \text{ for every } \in > 0.$$

Example:

The variable X_i (i = 1, 2,......) assumes the value $2^{i-2\log i}$ with probability 2^{-i}. Examine if the WLLN holds in this case.

Solution : $E(X_i) = \sum_{i=1}^{\infty} 2^{-i} \cdot 2^{i - 2\log i}$

$$= \sum_{i=1}^{\infty} 2^{-2\log i}$$

$$= \sum_{i=1}^{\infty} \frac{1}{2^{\log i^2}} = \sum_{i=1}^{\infty} \frac{1}{i^2}.$$

But the series $\sum_{i=1}^{\infty} \frac{1}{i^2}$ is a convergent series.

So, $E(X_i)$ is finite. Hence, by Khintchine's theorem, the sequence X_1, X_2....., $X_{n,}$.......satisfies the WLLN. This is called Khintchine's WLLN.

Example : Consider Poisson's scheme of sampling : n independent Bernoulli trials are performed, the i^{th} trial with probability of success p_i, and failure q_i.

Let X_i = 1, or 0 if the i^{th} trial results in 'success' or in 'failure'.

Then $E(\bar{X}_n) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} E(X_i)$

$$= \frac{1}{n}\sum_{i=1}^{n} p_i = \bar{p},$$

and $V(\bar{X}_n) = \frac{1}{n^2}\sum_{i=1}^{n} V(X_i)$, since X_1, X_2,......, X_n are independent.

$$= \frac{1}{n^2}\sum_{i=1}^{n}\left[E\left(X_i^2\right) - \left(E(X_i^2)\right)\right]$$

$$= \frac{1}{n^2}\sum_{i=1}^{n}\left[p_i - p_i^2\right]$$

$$= \frac{1}{n^2}\sum_{i=1}^{n} p_i q_i \text{ , (where } q_i = i - p_i)$$

Note that $p_i q_i \le \frac{1}{4}$, $\forall i$,

$$\therefore V(\bar{X}_n) = \frac{1}{n^2}\sum_{i=1}^{n} p_i q_i$$

$$\le \frac{1}{n^2}\sum_{i=1}^{n}\left(\frac{1}{4}\right) = \frac{1}{4n},$$

which tends to zero, as $n \rightarrow \infty$.

Thus, $\bar{X}_n \rightarrow \bar{P}$, in probability.

This is known as Poisson's WLLN.

2.5.3. Central limit theorem (CLT)

Let the distribution of a random variable Y depend on a parameter n, and if there exits two quantities μ and σ (which may or may not depend on n) such that

$$\lim_{h\rightarrow\infty} P\left[\frac{Y-\mu}{\sigma} \le t\right] = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{t} \exp\left(e^{-x^2/2}\right)dx, \text{ for all } t,$$

then we say that y is asymptotically normally distributed with mean μ and variance σ^2. We also say that $(Y-\mu)/\sigma$ follows the central limit law.

The following theorem states the classical problem where the variables are independent.

The Lindeberg-Levy Theorem—

Let $\{X_k\}$ be a sequence of independent and identically distributed random variables with

$E(X_k) = \mu$ and $V(X_k) = \sigma^2 < \infty$. Then

$$\frac{\sum_{K=1}^{n}(X_K - \mu)}{\sigma\sqrt{n}}$$

is asymptotically normal with mean 0 and variance 1.

Note that

$$\frac{\sum_{K=1}^{n}(X_K - \mu)}{\sigma\sqrt{n}} = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}},$$

so that the sample mean $\bar{X}$ is asymptotically normal with mean μ and variance σ^2/μ.

Theorem (Liapounov)

Let $\{X_n\}$ be a sequence of mutually independent random variables such that for some $\delta(>0)$,

$E\{|X_k - \mu_k|\}^{2+\delta}$ exists for every k = 1, 2, 3, ..., where

$E(X_k) = \mu_k$ and $V(X_k) = \sigma_k^2$.

Then, if the condition

$$\lim_{n\to\infty} \frac{1}{\left(\sum_{k=1}^{n}\sigma_k^2\right)^{1+\frac{d}{2}}} \sum_{k=1}^{n} E\left\{|X_k - \mu_k|^{2+\delta}\right\} = 0$$

is satisfied, the $\frac{\left(S_n - \mu_{(n)}\right)}{\sigma_{(n)}}$ is asymptotically normal with mean 0 and variance 1, where

$$\mu_{(n)} = \sum_{k=1}^{n}\mu_k, \quad \sigma_{(n)}{}^2 = \sum_{k=1}^{n}\sigma_k^2 .$$

The Liapounov theorem, as stated about, gives only a sufficient condition for $\{S_r\}$, where $S_n = \sum_{k=1}^{n} X_k$, to follow central limit theorem (CLT). The proof is omitted.

Theorem (Lindeberg - Feller)

The following theorem, known as Lindeberg-Feller CLT, gives a necessary and sufficient condition for the sum $S_n = \sum_{k=1}^{n} X_k$ to have an asymptotically normal distribution.

This is the central limit theorem (CLT) in the most general form. We state the theorem without proof.

Statement of Lindeberg-Feller CLT :

Let $\{X_n\}$ be a sequence of independent r.v.s with the cumulative distribution function (c.d.f.) of X_k as $F_x(x)$, $E(X_k) = \mu_k$ and $V(X_k) = \sigma_k^2 < \infty$.

Then,

(i) S_n is asymptotically normal $N(\mu_{(n)}, \sigma_{(n)})$,

(ii) $\lim_{n \to \infty} \max_{1 \le k \le n} \frac{\sigma_k}{\sigma_{(n)}} = 0$,

if and only if

$$A_n(\in) = \frac{1}{\sigma_{(n)}^2} \sum_{k=1}^{n} \int_{|x_k - \mu_k| \ge \in \sigma_{(n)}} (x - \mu_k)^2 \, dF_k(x) \to 0,$$

as $n \to \infty$, holds for every $\in > 0$.

We omit the proof of the theorem.

Below we consider an example where CLT holds, but WLLN does not hold.

Example :

Consider the sequence of mutually independent r.v.'s $\{X_n\}$ such that

$$P\left(X_k = k\right) = P\left(X_k = -k\right) = \frac{1}{2\sqrt{k}}, \quad P\left(X_k = 0\right) = 1 - \frac{1}{\sqrt{k}}, k > 0$$

Choose δ = 1. Then CLT holds if

$$\lim_{n\to\infty} \frac{e}{\sigma} = 0, \text{ where } \rho^3 = \sum_{k=1}^{n} E\left|X_k - \mu\right|^3$$

$$\text{and } \sigma = \sqrt{\sum_{k=1}^{n} \sigma_k^2}$$

Here $\mu = E\left(X_k\right) = k \cdot \frac{1}{2\sqrt{k}} - k \cdot \frac{1}{2\sqrt{k}} = 0$

and $E\left|X_k\right|^3 = k^3 \cdot \frac{1}{2\sqrt{k}} + k^3 \cdot \frac{1}{2\sqrt{k}} = k^{\frac{5}{2}}$

and $V\left(X_k\right) = E\left(X_k^2\right) = k^2 \cdot \frac{1}{2\sqrt{k}} + k^2 \cdot \frac{1}{2\sqrt{k}} = k^{\frac{3}{2}}$

Therefore,

$$\lim_{n\to\infty} \frac{e}{\sigma} = \lim_{n\to\infty} \frac{\left(\sum_{k=1}^{n} k^{\frac{5}{2}}\right)^{\frac{1}{3}}}{\left(\sum_{k=1}^{n} k^{\frac{3}{2}}\right)^{\frac{1}{2}}} \simeq \lim_{n\to\infty} \frac{\left(\int_0^n x^{\frac{5}{2}} dx\right)^{\frac{1}{3}}}{\left(\int_0^n x^{\frac{3}{2}} dx\right)^{\frac{1}{2}}}$$

$$= \lim_{n\to\infty} \frac{A}{g\left(n\right)}, \text{ say}$$

where $$\frac{\left(\int_0^n x^{\frac{5}{2}}dx\right)^{\frac{1}{3}}}{\left(\int_0^n x^{\frac{3}{2}}dx\right)^{\frac{1}{2}}} = \frac{\left(n^{\frac{7}{2}}.\frac{2}{7}\right)^{\frac{1}{3}}}{\left(n^{\frac{5}{2}}\cdot\frac{2}{5}\right)^{\frac{1}{2}}} = \left(\frac{n^{\frac{7}{6}}}{n^{\frac{5}{4}}}\right)\cdot\frac{2^{\frac{1}{3}-\frac{1}{2}}\sqrt{5}}{7^{\frac{1}{3}}}$$

$$= \frac{1}{n^{\frac{1}{12}}}\cdot\frac{\sqrt{5}}{2^{\frac{1}{6}}7^{\frac{1}{3}}} = \frac{A}{g(n)}$$

where $A = \frac{\sqrt{5}}{\sqrt[6]{2}\sqrt[3]{7}}$, and $g(n) = \sqrt{n^6}$

Then, $\lim\limits_{n\to\infty}\frac{A}{g(n)} = 0$.

Hence CLT holds.

But $$\lim_{n\to\infty}\frac{V(S_n)}{n^2} = \lim_{n\to\infty}\frac{n^{\frac{5}{2}}\cdot\frac{2}{5}}{n^2} = \lim_{n\to\infty}\frac{\sqrt{n}\cdot 2}{5} = \left(\frac{2}{5}\right)\lim_{n\to\infty}\sqrt{n} \neq 0$$

Hence WLLN does not hold.

2.6 Summary

The chapter elaborates the distinction between a discrete random variable, and a continuous random variables. Several standard distributions of discrete random variables are considered here, and m.g.f.'s as well as means and variances of these distributions are derived.

Similarly, some distributions of continuous random variables, including the well-known normal distribution, are considered here. The m.g.f.'s of these distributions have been calculated and the means and variances are derived from these m.g.f.'s.

The bivariate normal distribution is also considered and the marginal and conditional distributions of this bivariate distribution are also derived. The ideas of conditional mean and variance are also given.

2.7 Exercises

1. Write the m.g.f. of a normal random variable X having p.d.f. N(0,1). Also, state its mean and variance.
2. For a binomial distribution, the mean is always greater than the variance. Discuss this phenomenon.
3. Show that exponential distribution can be obtained as a special case of gamma distribution. Hence, obtain its mean and variance.
4. Obtain the m.g.f. of a log-normal distribution starting from a normal distribution.
5. Obtain the m.g.f. of a Poison distribution. Hence obtain the p.g.f. from this m.g.f. Find also $E(X)$ and $V(X)$.
6. Consider the trinomial distribution with p.m.f.

$$P(X=x,Y=y)=\frac{\underline{|n}}{\underline{|x}\,\underline{|y}\,\underline{|n-x-y}}p_1^x p_2^y p_3^{n-x-y}$$

where x, y are non-negative integers such that $x + y \le n$, $p_1, p_2, p_3 > 0$ and $p_1 + p_2 + p_3 = 1$.

Then $\quad E(Y/x)=(n-x)\dfrac{p_2}{1-p_1},$

and $\quad E(X/y)=(n-y)\dfrac{p_1}{1-p_2}.$

7. Suppose X has the probability function given by

$$P(X = a + bk) = \text{C},\ \text{k} = 1, 2, \ldots, \text{N}.$$

where $C(>0)$ is a constant. Find the constant C and the mean and variance of X.

8. Show that the normal distribution N (μ, σ^2) is symmetric about its mean μ.

9. Show that X has a symmetric distribution if and only if X and $-X$ are identically distributed.

10. Let $f_1, f_2, ..., f_k$ be density functions on the interval (a, b), b>a. Show that

    (i) $\sum_{i=1}^{k} f_i$ cannot be a density function on the interval (a, b).

    (ii) $\sum_{i=1}^{k} \alpha_i f_i$, $0 \le \alpha_i \le 1$, $\sum_{i=1}^{k} \alpha_i = 1$ is a density function on (a, b).

11. A fair coin is tassed once. Let X be the number of heads and Y be the number of tails. Are X and Y identically distributed? Is $P(X = Y) = 1$?

Examines related to WLLN and CLT.

Unit 3 ❑ Survey Methodology

Structure

3.0 Objectives

The followings are discussed here:

- Sample Survey
- Complete enumeration
- Errors in Survey
- Ramdom Sampling

- Simple random sampling with replacement (SRSWR)
- Simple Random Sampling with Replacement (SRSWOR)

3.1 Introduction

In Scientific investigations, we draw samples from populations. We need these samples to be representative of the populations. if they are not, then the inferences drawn from than will be biased. The theory of probability helps in drawing representative samples. This reduces the errors in survey as well. For operational simplicity, SRSWR or SRSWOR is usually employed in real life problems.

3.2 Sample survey and complete enumeration

The practice of drawing a 'sample' from a 'population' and then drawing inferences about this population is quite old.Technically speaking, a population is the set or collection of all conceivable and identifiable units (i.e., individuals) under study, a sample means only a part (i.e., a subset) of the population. Data are collected only on the individuals that form the sample. It is desirable that the sample should be a good representative of the underlying population.

By sample survey, we mean studying the population on the basis of the data collected from a sample. Complete enumeration, on the other hand, means studying all the units of the population. This is called census.

There are some advantages of sample survey over complete enumeration, or census. These are :

(a) greater speed in execution,

(b) less cost compared to total inspection of all units,

(c) greater accuracy of the results,

(d) possibility of estimation of error in drawing inferences.

Regarding the point (a) above, it is clear that as only a part of the population is considered in a sample survey, less time will be required for study, and hence greater speed can be achieved.

Regarding the cost aspect namely, (b), it can be said that the cost per unit involved in a sample survey would be larger compared to that in a complete enumeration owing to the employment of skilled and trained workers, sophisticated equipments for data collection, and so on. But the total cost in a sample survey is likely to be much less than that in a complete enumeration because of the smaller number of units involved in the study.

3.2.1 Sampling and Non-sampling Error

To understand the point (c) above, it is necessary to distinguish between 'sampling error' and 'non-sampling error'. The discrepancy between the estimate supplied by the sample, and the true value of the population characteristic under study is termed as sampling error. Thus, this error arises solely due to the fact that we are studying a part of the population, and not the whole of it. This type of error, therefore, is absent in a complete enumeration. On the contrary, errors due to omission, wrong tabulation, fatigue, miscalculation, and so on, constitute what are called non-sampling errors. These errors are likely to be present both in sample survey as well as complete enumeration.

It is to be understood that both these errors, namely, sampling error and non-sampling error, are present in a sample survey while a complete enumeration is affected by the non-sampling error only whose magnitude will be very high if the population is large enough. We can, however, control both the errors present in a sample survey, firstly, by using appropriate statistical inference procedure, and, secondly, by employing skilled and adequately trained personnel for the survey.

Finally, regarding the point (d), it is possible to give statistical measure of the error involved in the inference regarding the population.

3.3 Different types of sampling

Broadly speaking, these are two types of sampling : (a) Judgement sampling, and

(b) Probability sampling. Judgement samples are those which are selected rather subjectively by the investigator and, as a result, no scientific and reliable conclusions can be drawn from these types of samples. For example, in estimating the quality of apples in a basket, the sampler may choose apples only from the upper portion which are usually of the good type. Hence, the sample does not represent the population, here apples of the whole basket, adequately.

Probability samples, also called random samples, are selected in such a way that, at each draw, each unit in the population has got some preassigned and known probability of being included in the sample. Thus, let there be a population of N units, N being finite. Aslo, the p_i be the probability of selecting the i^{th} unit (i = 1, 2,..,N) in the first draw. Then, $p_i > 0$ and $\sum_{i=1}^{n} p_i = 1$. If $p_i = \frac{1}{N}$ i.e., the probability of selection is the same for all units, then the sample drawn is called a simple random sample (SRS), and the selection procedure is termed as the simple random sampling.

3.3.1 Simple Random Sampling With Replacement (SRSWR) and Simple Random Sampling Without Replacement (SRSWOR)

There are two ways of drawing a simple random sample : (a) with replacement, (b) without replacement. In the first place, ach unit being drawn is returned to the population before the next draw. In the second case, the unit drawn in a draw is not returned before the next draw and the process is continued till a specified number of units are taken.

Clearly, in SRSWR, the population size remains the same, but in SRSWOR, the population size decreases in each step. Also, in case of SRSWR, the same population member may appear more than once in the sample, while in SRSWOR, all units of the sample are distinct.

Let N be the number of units in the population, N being finite. If SRSWOR is employed, then at the first draw, a unit will have probability $\frac{1}{N}$ of being selected, at the second draw, a unit will be selected, with probability $\frac{1}{N-1}$, etc. If the sample size is n, then the n^{th} sample unit will be selected with probability $\frac{1}{N-n+1}$.

In case of SRSWR, the probability of selecting the first unit is $\frac{1}{N}$, the second unit $\frac{1}{N}$,..., the n^{th} unit also with $\frac{1}{N}$.

3.3.2 Sampling Errors in SRSWR and SRSWOR

Suppose there are N units in a population and Y_1, Y_2,....,Y_N are the corresponding values of the variable y under study. In many cases, the population size N will be known, but the population mean or total unknown. Our aim is to estimate these population parameters on the basis of random samples.

Define standard error of an estimator.

Theorem 1. Let y_1, y_2,....,y_n be the values of y for an SRSWOR of size n obtained from a population of size N with members having the values Y_1, Y_2,...,Y_N. Then,

$E(\bar{y}) = \bar{Y}$, ($\bar{y}$ and $\bar{Y}$ being the sample and population mean).

Also, $v\ (\bar{y}) = \frac{\sigma^2}{n} \cdot \frac{N-n}{N-1}$, where σ^2 is the population variance and is given by

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}\left(Y_i - \bar{Y}\right)^2.$$

The positive square-root of the variance $V(\bar{y})$ is called the sampling error (same as standard error) of $\bar{y}$ which estimators $\bar{Y}$.

Thus, *i.e.*, $(\bar{y}) = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$

Theorem 2. Let y_1, y_2,...,y_n be the values of y for an SRSWR of size n obtained from a population of size N. Then,

$E(\bar{y}) = \bar{Y}$, the population mean

$V(\bar{y}) = {}^{\sigma^2}\!/_{n}$, and

i.e., $(\bar{y}) = {}^{\sigma}\!/_{\sqrt{n}}$.

Note that SRSWOR is better than SRSWR since $\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}} \le \frac{\sigma}{\sqrt{n}}$.

The quantity $\frac{N-n}{N-1}$ is called 'finite population correction' (f.p.c.)

3.4 The method of drawing random samples, random numbers

3.4.1 Method of drawing random samples

Suppose we want to draw a random sample of size 3 from a population of 10 marbles. Here, $n = 3$, $N = 10$. This can be done by putting 10 marbles in an urn, and then taking 3 from it after properly mixing all the marbles. The situation becomes complicated if N is large enough or the population units are human beings, villages, etc. The problem is then solved by using what are called random numbers. As the first step, we prepare a list of all the units in the population, which is termed as a 'sampling frame'.

Definition of a random number series : A random sampling number series is an arrangement, which may be worked upon either as linear or rectangular, in which each place has been filled in with one of the digits 1,2,....,9,0. The digits occupying any place is selectd at random from these ten digits and independently of the digits occurring in other places. Thus, a series of numbers (read either as one digit, two digited, three digited, etc.) are said to be random if the following conditions are satisfied :

(a) the numbers occur with equal frequency (i.e., probability) in the long run,

(b) the numbers occur independently of one another.

There are published tables of random numbers which are used in practice.

Pseudo-random number : Please mention about computer generated random numbers using linear congratied sources.

Mean and standard error of sample proportion :

Sometimes we are interested in estimating the proportion p of individuals in the population having a certain attribute A. Clearly, population proportion of individuals possessing the attribute 'not - A' is q, $q = 1 - p$, $0 < p < 1$. This type of situation arises when we estimate, say, the proportion of voters supporting an issue of national interest. The following theorem may be of interest in this regard :

Theorem 3. Let f be the number of units having a certain attribute in an SRSWOR of size n. Then $f/_n = \hat{p}$, (say), will be the sample proportion. Let N be the population size and p the population, then

$$E(\hat{p}) = p$$

$$V(\hat{p}) = \frac{pq}{n}\frac{N-n}{N-1},\ q = 1 - p,$$

and *i.e.*, $(\hat{p}) = \sqrt{\frac{pq}{n}}\sqrt{\frac{N-n}{N-1}}$,

where $\frac{N-n}{N-1}$ is the finite population correction (*f.p.c.*).

Note that if n is small compared to N, then for large N, $\frac{N-n}{N-1} = \frac{1-n/_N}{1-1/_N} \simeq 1$, as

$\frac{n}{N}$ and $\frac{1}{N}$ both tend to zero for large N.

As in the earlier case, for SRSWR, we have

$$E(\hat{p}) = p,$$

$$V(\hat{p}) = \frac{pq}{n}$$

and *i.e.*, $(\hat{p}) = \sqrt{\frac{pq}{n}}$.

Clearly, if the population is large, then SRSWR and SRSWOR are identical.

3.4.2. Random numbers–their uses and properties

The following example will explain the use of random numbers in selecting random sample from a population :

Example 1. Suppose we have to select a random sample of 35 voters from a list of 345 voters.

(Reference : Statistics for the Social Sciences by Gun, A.M. and Aich, A.B., World Press, P. 104.)

Solution : First of all, the voters are serially numbered from 1 to 345. Here N = 345, which is a three-digited number. As a source of random numbers, one may use either (a) *A Million Random Digits,* published by Rand Corporation (1955), or (b) *Random Sampling Numbers* (Tracts for Computers. XV) by L.H.C. Tippett (1927), or (c) the random number series given in the ISI publication *Formulas and Tables for Statistical Work* by Rao, Mitra, Mathai and Ramamurthi.

Since we need three-digited numbers, we can select arbitrarily columns 1 to 3, or 4 to 6, or 7 to 9, etc., and read the numbers vertically downwards, rejecting all numbers greater than 345 as well as the number 000. The selected voters will be those whose serial numbers correspond to the chosen random numbers. We continue the process till the required sample size is reached. The above procedure, however, often involves the rejection of a large number of random numbers. To avoid this huge rejection, we divide a random number by 345 and choose the voter with a serial number between 1 and 344 that corresponds to the remainder if it is different from zero, and the serial number 345 when the remainder is zero. It is however, necessary to reject the random numbers from 691 to 999 (as also the number 000) in the above procedure, because these will correspond to Voter Nos. 1 to 309 and so each will get a larger chance of being selected, viz. 3/999, while Voter Nos. 310 to 345 will have the probability 2/999 each of being selected.

If we draw an SRSWOR, then we have to ignore any repetition of numbers. In the case of an SRSWR, we have to retain these numbers.

The next example explain the estimation of the standard error, or sampling error of the mean.

To do this we note that the population variance σ^2 is estimated by

$$\hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2$$

$$= \frac{1}{n-1}\left\{\sum_{i=1}^{n} y_i^2 - n\bar{y}^2\right\} = s^2, \text{ say.}$$

Also, note that $E(\hat{\sigma}^2) = \sigma^2$.

In the following example, observations on a sample of size 20 are given for a large population :

***Example* : 2** To estimate the average monthly rent per flat in a big locality, a random sample of 20 flats was selected and the following data on monthly rent (in Rs.) were obtained :

385 715 615 475 800 735 525 390 720 500

565 900 525 485 650 575 435 825 415 480

Estimate the average monthly rent. Also obtain an estimate of the standard error of the estimate.

We shall assume that the total number of rented flats in the locality is large compared to 20 and hence shall igmore the f.p.c.

Since $\sum_{i=1}^{n} y_i$ = 11715,

the estimate of the average monthly rent is

$\bar{y} = \frac{11715}{20}$ = 585.75 rupees.

The estimate of the standard error of the estimate of average monthly rent is (since N may be supposed to be very large)

$\hat{s.e.}(\bar{y}) = \frac{s}{\sqrt{n}}$, where s^2 = $\frac{1}{n-1}\left[\sum_{i=1}^{n} y_i^2 - n\bar{y}^2\right]$

Henre, n = 20, $\bar{y}$ = 585.75, $\sum_{i=1}^{20} y_i^2$ = 7305325.

On simplification, we obtain s^2 = 23329.674 and so

$\hat{s.e.}(\bar{y}) = \frac{152.7405}{\sqrt{20}}$ = 34.15 rupees.

Properties of random numbers : From the definition of random numbers given in 3.3.1, it follows that if the random numbers are read with a decimal before them i.e., the number 3459, say, is read as 0·3459, then the resulting numbers will constitute observations from a uniform distribution $U(0, 1)$. This result is used in drawing random samples from standard distributions, as explained in Unit 4.

3.4.3 Different tests for random numbers

The following tests are commonly used as test for random numbers. These tests can be used to any part of the series, read as single digited, two-digited, three-digited numbers, and so on. These tests are explained below :

(a) Frequency test : Here, the observed frequencies of the ten digits from 0 to 9 are noted. Then the observed and expected frequencies are compared by χ^2 - test under the assumption that the numbers are actually random in which case each digit has the probability $\frac{1}{10}$ to occur.

(b) Serial test : Here the series of numbers are read as two-digited numbers, i.e., 00, 01, ...,99 and their observed frequencies are noted. Then, the χ^2 - test is applied to test the hypothesis that the numbers are random in which case the probability for each is $\frac{1}{100}$.

(c) Gap test : We first pick out the successive occurrences of zeros, say, and find the gaps between them. The frequencies of such gaps are obtained and the hypothesis of randomness, according to which the gap is x with probability $\frac{9^x}{10^{x+1}}$ $x = 0, 1, 2,,,,,,$ is tested by using an appropriate χ^2.

(d) Poker test : Here, the series of numbers are read as four (or five)-digited numbers. There are five possibilities – *aaaa*, *abcd*, *aabb*, *aaab*, *aabc*. Under the hypothesis of randomness, their probabilities are obtained. Finally, observed and expected frequencies are compared by an appropriate χ^2 - test.

3.5 Random number generation using inverse transformation technique

Example 1. Exponential distribution

Consider the exponential distribution with *pdf*

$$f(x) = \beta^{-1} e^{-x/\beta}, \; x > 0,$$

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

The distrtibution function $F(x)$ of this distribution is given by

$$F(x) = \int_0^x f(x)dx$$

$$= \beta^{-1} \int_0^x e^{\frac{-x}{B}} dx$$

$$= \beta^{-1} \left[\frac{e^{-x/\beta}}{\left(-\frac{1}{\beta}\right)} \right]_0^x$$

$$= 1 - e^{-x/\beta}$$

$$\therefore \; e^{-x/\beta} = 1 - F(x),$$

or, $\left(-x/\beta\right) = \log_e(1 - F(x))$.

or, $x = -\beta \log_e (1 - F(x))$.

As explained earlier, $F(x)$ is obtained from the random number table by considering the numbers with decimals before each of them, e.g., 0.6752 is taken to be $F(x)$ if 6752 number appears (reading the table as four-digited numbers). Here, we use the result that $F(x)$ follows a uniform distribution $U(0,1)$.

Thus, for each given $F(x)$, we obtain a value of x, as shown above.

This is called inverse transformation technique for random number generation.

In order to generate samples of large size are may as pseudo-random numbers as the values of $F(x)$.

Example 2. Consider the Cauchy distribution with location parameter θ and scale parameter λ as

$$f(x) = (\pi\lambda)^{-1}\left[1+\left\{\frac{x-\theta}{\lambda}\right\}^2\right],\ \lambda > 0,\ -\infty < x < \infty.$$

The cumulative distribution function $F(x)$ is the

$$F(x) = \int_{-\infty}^{x} f(x)dx$$

$$= \frac{1}{2}+\frac{1}{\pi}\tan^{-1}\left(\frac{x-\theta}{\lambda}\right).$$

Therefore, $\tan^{-1}\left(\frac{x-\theta}{\lambda}\right) = \pi(F(x) - \frac{1}{2})$,

or, $\frac{x-\theta}{\lambda} = \tan(-\frac{\pi}{2}+\pi F(x))$,

or,$x = \theta+\lambda\tan(-\frac{\pi}{2}+\pi F(x))$.

Then, for each given $F(x)$ (which is obtained from random number tables), we get a value of x by using the above relation. This x will be an observation from the Cauchy distribution stated above. The procedure will be repeated till we reach the required number of observations.

Example 3. The *pdf* of a Pareto distribution, which fits income for all higher income group, is given by

$$f(x) = ak^a x^{-(a+1)},\ k > 0,\ a > 0,\ x \geq k.$$

The cumulative distribution function $F(x) = P(X \leq x)$ is

$$F(x) = 1 - \left(\frac{k}{x}\right)^a, \ k > 0, \ a > 0, \ x \geq k.$$

$$\therefore \left(\frac{k}{x}\right)^a = 1 - F(x),$$

or, $$\left(\frac{k}{x}\right) = (1 - F(x))^{\frac{1}{a}}$$

or, $$\left(\frac{x}{k}\right) = (1 - F(x))^{-\frac{1}{x}} = (1 - F(x))^{\frac{1}{a}}$$

or, $$x = k\ (1 - F(x))^{-\frac{1}{a}}$$

Hence, for each $F(x)$, we get an x by using the above relation.

Example 4. The standard form of a gamma distribution is given by

$$f(x) = \frac{x^{\alpha-1}e^{-x}}{\Gamma(\alpha)}, \ x \geq 0.$$

Note that the cumulative distribution function $F(x)$ is

$$F(x) = P(X \leq x)$$

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

$$= \frac{1}{\Gamma(\alpha)} \underset{\alpha}{\Gamma}(x), \text{ where}$$

$\underset{\alpha}{\Gamma}(x)$ is the incomplete gamma function given by

$$\underset{\alpha}{\Gamma}(x) = \int_0^x x^{\alpha-1} e^{-x} dx.$$

Thus, $\underset{\alpha}{\Gamma}(x) = \Gamma(\alpha) F(x)$, so that for given $F(x)$ (obtained from the random number tables), the right hand side is known completely (since α is a known constant).

Also, $\Gamma_\alpha(x)$ is tabulated for different x by Pearson.

Hence, x can be obtained by inverse interpolation from given values of $\Gamma_\alpha(x)$. This is a random number from the gamma distribution with shape parameter α.

We may generate gamma from exponential samples (this method is easier) as gamma is sum of indop exponential.

3.6 Summary

In this Unit, we have studied the basis of sample survey which is a branch of applied statistics. We have defined a random sample, distinguished between random sampling with replacement and without replacement. The sampling errors in these cases are also indicated.

The uses of random number series are also discussed. Their properties are stated with examples.

The technique, called inverse transformation technique, for generating random numbers is discussed taking some standard statistical distribution. The technique can be extended to other distributions as well.

3.7 Exercises

1. Define random sample. Distinguish between SRSWR and SRSWOR. State the corresponding sampling errors.
2. Define finite population correction (*f.p.c.*). What happens to *f.p.c.* if sample size is (a) 25%, (b) 50% and (c) 75% of the population size?
3. Define random number series. State their uses.
4. State some common tests for random numbers.
5. Generate a random sample of size five from a normal distribution with mean 50, and variance 25. [as hint plese mention the polar transformation]

Unit 4 ❑ Estimation Theory

Structure

4.0 Objectives

The followings are discussed here:

- Meaning of Parameters
- Estimation of Parameters
- Difference between parameters and Statistics

- Properties of Good Estimators
- Methods of Estimation

4.1 Introduction

We have distinguished between a sample and a population. In many practical situations, nothing or very little is known about the population. The problems of drawing inferences about the population then arise which constitute the topics of statistical inference.

The logic behind statistical inference is clearly 'inductive' in the sense that it involves generalization from particular cases. Because, here random samples are first drawn from the populations, certain features of these samples, say, mean and variance, etc., are then calculated and using these values the corresponding features of the populations are then guessed. The purpose of the present chapter as well as the following chapter is to make this process of guessing as objective as possible.

The type of statistical inference usually takes one of the two forms : (a) estimation of some unknown features of the population, (b) testing about a tentative assumption about the unknown parameter of the population. The first one will be studied in the present chapter, while the second will be taken up in the next chapter.

As examples, consider data on family-size obtained from a random sample of families of Kolkata, which may be used to guess the average family-size in the city. This is a case of estimation. On the other hand, in studying the percentage of voters showing allegiance to a particular political party, our tentative assumption might be that the proportion of voters sympathetic to that party is, say, 0·48. This is the example of statistical hypothesis which would test this assumption or belief.

4.2 Statistic and Parameter

To begin with, it is important to distinguish between a 'statistic' and 'Parameter'. A statistic is a function of the sample observations while a parameter is a characteristic of the population. Thus, a statistic is a random variable, and a parameter is a constant. However, there is a branch of statistics, called Bayesian School, where a parameter itself is also considered to be a random variable. We would not discuss this approach here.

It is also important to distinguish between an 'estimate' and an 'estimator'. If an unknown parameter θ, say, population mean, is estimated by the sample mean $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$, then the particular value obtained from a given sample X_1, X_2, ...,X_n is called an estimate, while in the general case the random variable $\bar{X}$ is termed as estimator.

In the following discussions, we shall assume that we are given a random sample X_1, X_2,......,X_n from some distribution having the *pdf* $f(x)$ in the continuous case, (or *pmf* $f(x)$ in the discrete case). Some or all the parameters of the distribution may be unknown. Then any function $T(X_1, X_2, ...,X_n)$ of $X_1, X_2,....,X_n$ will be called a statistic. An estimation $\hat{\theta}$ of the parameter θ is itself a statistic, and we may write $\hat{\theta} = T\ (X_1, X_2,.....,X_n)$.

The criteria of selecting the right type of estimator $\hat{\theta}$ are the topical of discussion of this chapter.

We first mention certain desirable properties to be satisfied by a good estimator.

(a) Unbiasedness :

An estimator T of θ will be called unbiased if $E(T) = \theta$, for all θ. This would mean that the mean of the sampling distribution of T is θ. In non-technical terms, unbiasedness means that if we estimate θ by T a large number of times, *i.e.*, by drawing repeated random samples of a fixed size n, then, on the whole, we shall neither overestimate nor underestimate the true value of θ. An estimator which is not unbiased is called a biased estimator. For a given parameter, there may exist a large number of unbiased estimator.

(b) Minimum Variance :

The property of unbiasedness is, in itself, not a very important criterion. This is because expectation is only one of the many measures of location. Moreover, an estimator possessing this property is not unique. A natural course of action would be that of selecting the unbiased estimator having the smallest variable. If such an estimator exists, it is called the minimum variance unbiased estimator (MVUE). An MVUE, if exists, is unique.

(c) Consistency :

Consistency is essentially a large-sample property of an estimator. A statistic T, which is a function of sample observations, depends on the size of the sample, say, n. By varying n, we get a sequence $\{T_n\}$. If how this sequence converges to a parameter θ in some sense, then T will be called a consisted estimator of θ.

For a consistent estimator T_n of θ (the suffix n in T_n indicates its dependence on n), the difference $|T_n - \theta|$ should decrease as n increases.

Note that $|T_n - \theta|$ is also a random variable being the function of T_n, and hence the statement that '$|T_n - \theta| \to 0$ as $n \to \infty$' is not a deterministic statement. Thus, there may exist situations where the difference $|T_n - \theta|$ is not small even through n is large.

The consistency property of an estimator, however, ensures that the proportion of such cases tends to zero as n increases. That is,

$$P\ \{|T_n - \theta| > \in\} \to 0, \text{ as } n \to \infty.$$

Or, equivalently,

$$P\ \{|T_n - \theta| \leq \in\} \to 1, \text{ as } n \to \infty.$$

Either of the above conditions is taken as the difinition of consistency of T_n for θ. In practice, the verification of the above conditions is not at all easy. However, the following result helps us to decide whether a given statistic (or estimator) is consistent for a parameter θ :

Let $\{T_n\}$ be a sequence of estimators for the parameter θ. If, now,

$$E(T_n) \to \theta, \text{ as } n \to \infty$$

and $\quad V(T_n) \to 0, \text{ as } n \to \infty,$

Then T_n is consistent for θ.

The above two conditions provide a set of sufficient conditions for consistency which have been found very useful in practice.

Example : Let X be distributed in the Poisson from with parameter θ. Show that the only unbiased estimator if *exp* $[-(k+1)\theta]$, $k > 0$, is $T(x) = (-k)^x$

Solution : Let us write $\gamma(\theta) = exp\ (-(K+1)\theta)$

$$\text{Then } E(T(x) = \sum_{x=0}^{\infty}(-k)^x \frac{e^{-\theta}\theta^x}{\lfloor x}$$

$$= e^{-\theta}\sum_{x=0}^{\infty}\frac{(-k\theta)^x}{\lfloor x}$$

$$= e^{-\theta}e^{-k\theta} = e^{-\theta(k+1)} = \gamma(\theta)$$

So, $T(x)$ is an unbiased estimator of $\gamma(\theta) = exp\ (-(k+1)\theta$.

We note that $\gamma(\theta) > 0$. But its estimate $T(x) > 0$, if x is even,

and $T(x) < 0$, if x is odd.

Example : Let $X_1, X_2......,X_n$ be *iid* $N(\mu,\sigma^2)$. If S^2 is the sample variance defined by $S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$, then S^2 is consistent, though biased.

Solution : Let us write $\frac{ns^2}{\sigma^2} = \sum_{i=1}^{n}\left(\frac{X_i - \bar{X}}{\sigma}\right)^2$ which follows χ^2 with $(n-1)$ degree of freedom. From the property of a χ^2– distribution, we have

$$E\left(\frac{ns^2}{\sigma^2}\right) = (n-1)$$

$$\text{or, } E(S^2) = \left(\frac{n-1}{n}\right)\sigma^2$$

$$= \left(1 - \frac{1}{n}\right)\sigma^2,$$

which is less than σ^2.

Thus, S^2 is a biased estimate of σ^2 and it also underestimates σ^2.

Also $V\left(\frac{ns^2}{\sigma^2}\right) = 2(n - 1)$.

$$\text{or, } V(S^2) = \sigma^4 \frac{2(n-1)}{n^2}$$

$$= 2\sigma^4\left(\frac{1}{n} - \frac{1}{n^2}\right)$$

Hence, $E(S^2) \to \sigma^2$, as $n \to \infty$

and $V(S^2) \to 0$, as $n \to \infty$.

Thus, by the sufficient conditions for consistency, S^2 is seen to be consistent for σ^2.

Example : Consider the truncated Poisson distribution, truncated at $x = 0$, with *pmf*.

$$f(x) = (1 - e^{-\theta})^{-1}\ \frac{e^{-\theta}\theta^x}{\lfloor x}, \ x = 1, 2, \ldots\ldots.$$

Based on a single observation x, find an unbiased estimator of $1 - e^{-\theta}$.

Solution : Based on a single observation x, let us define an estimator $T(x)$ as

$T(x) = 0$, when x is odd

$= 2$, when x is even.

$$\text{Then } E(T(x)) = \sum_{x=1}^{\infty} T(x)\ \frac{1}{1-e^{-\theta}}\frac{e^{-\theta}\theta^x}{\lfloor x}$$

$$= \frac{e^{-\theta}}{(1-e^{-\theta})}\sum_{x=1}^{\infty} T(x)\ \frac{\theta^x}{\lfloor x}$$

$$= \frac{e^{-\theta}}{1-e^{-\theta}}\left[2\cdot\frac{\theta^2}{\lfloor 2} + 2\cdot\frac{\theta^4}{\lfloor 4} + 2\cdot\frac{\theta^6}{\lfloor 6} + \ldots\right]$$

$$= \frac{2e^{-\theta}}{(1-e^{\theta})}\left[\frac{\theta^2}{\lfloor\underline{2}}+\frac{\theta^4}{\lfloor\underline{4}}+\frac{\theta^6}{\lfloor\underline{6}}+....\right]$$

$$= \frac{2e^{-\theta}}{1-e^{-\theta}}\left[\frac{e^{\theta}+e^{-\theta}}{2}-1\right]$$

$$= \frac{(1-e^{-\theta})^2}{(1-e^{-\theta})}$$

$$= (1 - e^{-\theta})$$

So, $T(x)$ is unbiased for $1 - e^{-\theta}$.

Sufficiency :

Let $\underline{X} = (X_1, X_2, ..., X_n)$ be a sample from a population with *pmf* (or *pdf*) $\int_{\theta}(x)$, where θ is a unknown parameter.

A statistic $T(\underline{X})$ is said to be sufficient for θ if and only if the conditional distribution of $\underline{X}$, given $T = t$, does not depend on θ.

Example : Let $X_1, X_2, ..., X_n$ be *i.i.d.* Bernoulli random variables with parameter p. Consider the statistic T such that $T = X_1 + X_2 + ... + X_n$ which follows $b(n, p)$, i.e., a binomial distribution with parameters n and p.

The conditional distribution of $X_1, X_2, ..., X_n$, given $T = t$ (fixed), is

$P(X_1 = x_1, X_2 = x_2, ..., X_n = x_n \mid T = t)$

$$= \frac{p[(X_1 = x_1, X_2 = x_2, ..., X_n = x_n) \cap (T = t)]}{p(T = t)}$$

$$= \frac{p(X_1 = x_1, X_2 = x_2, ..., X_n = x_n)}{\binom{n}{t} p^t q^{n-t}}, \text{ (putting } \sum_{i=1}^{n} x_i = t \text{)}$$

$$= \frac{\overset{n}{\underset{i=1}{\pi}}\left\{p^{x_i}(1-p)^{(1-x_i)}\right\}}{\binom{n}{t}p^t q^{n-t}}, \text{ (note } q = 1 - p)$$

$$= \frac{p^t(1-p)^{n-t}}{\binom{n}{t}p^t q^{n-t}}$$

$$= \frac{1}{\binom{n}{t}}, \text{ which is independent of } p.$$

This shows that the statistic $T = X_1 + X_2 + ... + X_n$ is sufficient for the parameter p.

Not every statistic is sufficient. Let X_1, X_2 be *i.i.d by* $p(\lambda)$. Consider the statistic $T = X_1 + 2X_2$ for estimating unknown λ.

We have $P(X_1 = 0, X_2 = 1|X_1 + 2X_2 = 2)$

$$= \frac{P[(X_1 = 0, X_2 = 1) \cap (X_1 + 2X_2 = 2)]}{P(X_1 + 2X_2 = 2)}$$

$$= \frac{P(X_1 = 0, X_2 = 1)}{P(X_1 = 0, X_2 = 1) + P(X_1 = 2, X_2 = 0)}$$

$$= \frac{e^{-\lambda} \cdot \frac{e^{-\lambda}\lambda'}{\underline{|1}}}{e^{-\lambda} \cdot e^{-\lambda}\frac{\lambda}{\underline{|1}} + \frac{e^{-\lambda}\lambda^2}{\underline{|2}} \cdot \frac{e^{-\lambda}\lambda^0}{\underline{|0}}}$$

$$= \frac{\lambda}{\lambda + \frac{\lambda^2}{2}}$$

$$= \frac{1}{1 + (\frac{\lambda}{2})}, \text{ which depends on } \lambda$$

Thus, the statistic $T(X_1, X_2) = X_1 + 2X_2$ cannot be considered to be sufficient for the parameter λ.

4.2.1 Fisher-Neyman Factorization Theorem

The following theorem, called Fisher-Neyman factorization theorem, gives a constructive method of determining sufficient statistics :

Let $X_1, X_2,...,X_n$ be discrete (or continuous) with joint *pmf* (or *pdf*) $f_\theta\,(x_1, x_2,...,x_n)$. Then the statistics $T_1, T_2,....,T_k$ will be sufficient for θ (which may be a scalar or vector) if and only if the joint *pmf* $f_\theta\,(x_1, x_2,.....,x_n)$ is expressible in the form

$f_\theta\,(x_1, x_2,...,x_n) = g_\theta\,(t_1, t_2,....,t_k).h(x_1, x_2,...,x_n)$ where $g_\theta\,(\cdot)$ depends on θ and $h(\cdot)$ is independent of θ, and is a function of $x_1, x_2,...,x_n$ only.

Example : Let $X_1, X_2,...,X_n$ be *iid.* $P(\lambda)$ with the *p.m.f.* of X_i

as $f\,(x_i) = \dfrac{e^{-\lambda}\lambda^{x_i}}{\underline{|x_i}}$, $x = 0,\ 1,\ 2,....$

The joint *p.m.f.* of $X_1, X_2,...,X_n$ is

$$f(x_1, x_2,...,x_n) = \prod_{i=1}^{n}\left\{\frac{e^{-\lambda}\lambda^{x_i}}{\underline{|x_i}}\right\}$$

$$= e^{-n\lambda}\lambda^{\sum\limits_{i=1}^{n}x_i}\cdot\prod_{i=1}^{n}\frac{1}{\underline{|x_i}}$$

$$= g_\lambda\,(t).h(x_1, x_2,....,x_n)$$

when $g_\lambda\,(t) = (e^{-n\lambda}\lambda^{t}),\ t = \sum\limits_{i=1}^{n}x_i$

and $h(x_1, x_2,...,x_n) = \prod\limits_{i=1}^{n}\left(\dfrac{1}{\underline{|x_i}}\right)$

So, $t = \sum_{i=1}^{n} x_i$ is sufficient for λ.

Efficiency :

The concept of efficiency is related to variance of an estimator. We expect an estimator to be better if its variance is small. In other words, the reciprocal of the variance is large. This reciprocal of the variance of an estimator is termed as its efficiency. Thus, given two estimators T_1 and T_2, we will choose T_1 if $V(T_1) < V(T_2)$, or *eff* $(T_1) >$ *eff* (T_2), whose *eff* $(T_1) = \frac{1}{V(T_1)}$.

Complete family of distributions :

A statistic T for θ is said to be complete if for any function $\psi(T)$, we have

$E(\psi(T)) = 0$ for all θ.

$\Rightarrow \psi(T) = 0$, almost everywhere.

In this case, we say that the underlying distribution is also complete.

Thus, if T is complete, then there is no non-trivial unbiased estimator of zero based on T. The only such estimator is zero itself.

Example. Let $X_1, X_2, .., X_n$ be a random sample from some Poisson distribution $P(\theta)$ with common *p.m.f.*

$$f_\theta(x) = \frac{e^{-\theta}\theta^x}{\underline{|x}}, \; x = 0, 1, 2, ...$$

Here, $T = \sum_{i=1}^{n} X_i$ is a sufficient statistic for θ having the distribution $P(n\theta)$ so that the *p.m.f.* of T is

$$f_\theta(t) = \frac{e^{-n\theta}(n\theta)^t}{\lfloor t}, \; t = 0, 1, 2, \ldots\ldots$$

Consider a function $\psi(T)$ of T and write

$$E(\psi(T)) = \sum_{t=0}^{\infty} \Psi(t) \frac{e^{-n\theta}(n\theta)^t}{\lfloor t}$$

$$= e^{-n\theta} \sum_{t=0}^{\infty} a(t)\theta^t,$$

where $a(t) = \dfrac{n^t \Psi(t)}{\lfloor t}$

Thus, $E(\psi(T)) = 0$, for all θ,

$$\Rightarrow \sum_{t=0}^{\infty} a(t)\theta^t = 0, \text{ for all } \theta.$$

But, we know from algebra that a convergent power series which is identically zero, must have all the co-efficiants as zero.

Thus, $a(t) = 0$, for $t = 0, 1, 2, \ldots\ldots.$

or, $\psi(t) = 0$, for $t = 0, 1, 2, \ldots\ldots$

Hence, T is complete and, also, the Poisson family of distributions is complete.

Define Completeness.

4.2.2 Basu's Theorem

We first define an ancillary statistic which will be needed in stating the theorem.

A statistic $V(\,X)$ is said to be ancillary if its distribution does not depend on θ.

An ancillary statistic by itself contains no information about θ.

Statement of the theorem : Let θ be the parameter of a distribution $P(x)$, which may be a *p.m.f.* or *p.d.f.* Also, let T be a statistic which is complete as well as sufficient. Then any ancillary statistic $V(X)$ is independent of T.

Proof : If V is ancillary, then by definition the probability $P(V\in \text{A}) = p_A$, say, is independent of θ for all values of A. Define the conditional probability

$$P(V\in \text{A})/T = t) = \eta_A(t), \text{ say.}$$

But $\quad E(\eta_A(T)) = p(V\in A)$

$$= p_A, \text{ (using the result that } E(P(A/B) = P(A))$$

or, $E(\eta_A(T) - p_A) = 0,$

By completeness, $\eta_A(T) - p_A = 0,$

$$\text{i.e., } \eta_A(T) = p_A,$$

$$\text{or,} \qquad p(V\in A/T) = p(V\in A) \ \underline{X}$$

So, V and T are independent.

4.3 Minimum Variance Unbiased estimator, (MVUE) Cramer-Rao Inequality

As already mentioned, an unbiased estimator is not unique. Thus, there may be a number of unbiased estimator corresponding to a given parameter. It is, therefore, natural to select the one having the minimum variance. In this section, we discuss an inequality, called Cramer-Rao inequality (or C-R inequality), also called Cramer-Rao bound, which gives a lower bound to the variance of an unbiased estimator.

Given an unbiased estimator, we can use this inequality to check if the lower bound of the variance is attained. This is helpful to identify the estimator under consideration as a MVUE or not.

Theorem : Let $T(X_1, X_2,...,X_n)$ be an unbiased estimator of θ based on a random sample $X_1, X_2,....,X_n$ from a *p.m.f.* or *p.d.f.* $f_\theta(x)$. Then, **under some regularity conditions,**

we have $V(T) \geq \dfrac{1}{E\left(\dfrac{\partial L}{\partial \theta}\right)^2}$, (1)

where $L = \log_e f_\theta(x_1, x_2,...,x_n)$, $f_\theta(x_1, x_2,...,x_n)$ being the joint density of $X_1, X_2,....X_n$.

An unbiased estimator T of θ attaining the equality in (1) will be called MVUE or minimum variance bound (MVB) estimator, but the converse is not true.

Proof of the theorem :

For simplicity, we write the multiple integral *w.r.t.* $x_1, x_2,...,x_n$ as

$\int\limits_A (....)d\underline{X}$, where A means the entire permissible range.

Note that, $\int\limits_A f_\theta(\underline{X})d\underline{X} = 1$. (Total probability).

Differentiating *w.r.t.* θ, we have

$$\frac{\partial}{\partial\theta}\int_A f_\theta(\underline{X})\, d\underline{X} = 0$$

$$\text{or, } \int_A \frac{\partial}{\partial\theta} f_\theta(\underline{X})\, d\underline{X} = 0$$

$$\text{or, } \int_A \int_A \left(\frac{\partial L}{\partial\theta}\right) f_\theta(\underline{X})\, d\underline{X} = 0$$

$$\text{or, } E\left(\frac{\partial L}{\partial\theta}\right) = 0 \rightarrow (2)$$

Also, $\theta = E(T)$, (since T is unbiased for θ)

$$= \int_A t\, f_\theta(\underline{X})\, d\underline{X}$$

Differentiating *w.r.t.* θ, we have

$$1 = \int_A t \cdot \frac{\partial}{\partial\theta} f_\theta(\underline{X})\, d\underline{X}$$

$$= \int_A t\left(\frac{\partial}{\partial\theta}\log_e f_\theta(\underline{X})\right) f_\theta(\underline{X})\, d\underline{X}$$

$$= E(T\frac{\partial}{\partial\theta}\log f_\theta(\underline{X}))$$

$$= E(T\frac{\partial L}{\partial\theta}).$$

But $\cos(T, \frac{\partial L}{\partial\theta})$

$$= E\left((T-\theta)\left(\frac{\partial L}{\partial\theta} - E\left(\frac{\partial L}{\partial\theta}\right)\right)\right.$$

$$= E\left((T-\theta)\frac{\partial L}{\partial\theta}\right), \left(\because E\left(\frac{\partial L}{\partial\theta}\right) = 0\right)$$

$$= E\left(T\frac{\partial L}{\partial\theta}\right).$$

By Cauchy - Schwarz Inequality, we have

$$[\cos(T, \frac{\partial L}{\partial\theta})]^2 \le V(T)\cdot V(\frac{\partial L}{\partial\theta})$$

or, $1 \le V(T)\cdot E\left(\frac{\partial L}{\partial\theta}\right)^2, \left(\because V\left(\frac{\partial L}{\partial\theta}\right) = E\left(\frac{\partial L}{\partial\theta}\right)^2\right)$

or, $V(T) \ge \dfrac{1}{E\left(\frac{\partial L}{\partial\theta}\right)^2}.$

It can be shown further that

$$\mathrm{E}\left(\frac{\partial L}{\partial \theta}\right)^2 = -E\left(\frac{\partial^2 L}{\partial \theta^2}\right),$$

so that the above inequality reduces to

$$V(T) \geq -\frac{1}{E\left(\frac{\partial^2 L}{\partial \theta^2}\right)}, \text{ (assuming } E\left(\frac{\partial^2 L}{\partial \theta^2}\right) \neq 0)$$

It is to be noted that the above inequality will hold under some regularity conditions such as differentiability within the integration etc.

Regularity conditions : The following conditions taken together are called regularity conditions and the situation where these conditions hold is called a regular estimation case. The Cramer-Rao Inequality (or bound) is derived under these conditions only.

It is assumed that θ is a single parameter varying over the parameter space (H) and that $X_1, X_2,...,X_n$ are all continuous with joint pdf $f\theta(x_1, x_2, \ldots x_x)$. For the sake of semplicity, we write the multiple integral $\int\int_A \ldots . \int(\ldots)\prod_{i=1}^{n} dx_i$ as $\int_A (\ldots)d\underline{X}$. The discrete case follow similarly – multiples integrals being replaced by multiple sums.

I. (H) is a non-degenerate open-interval.

II. $\frac{\partial}{\partial \theta}f_\theta(x_1, x_2,...,x_n)$ exists for all $\theta \in$ (H)

III. $\frac{\partial}{\partial \theta}\int_A f_\theta(x_1, x_2,...,x_n)\, d\underline{X} = \int_A \frac{\partial}{\partial \theta} f_\theta(x_1, x_2,...,x_n)\, d\underline{X}$

IV. $\frac{\partial}{\partial \theta} \int_A t\, f_\theta(x_1, x_2,....,x_n)\, d\underline{X} = \int_A t\, \frac{\partial}{\partial \theta}f_\theta(x_1, x_2,...,x_n)\, d\underline{X}.$

V. $E\left\{\dfrac{\partial \log_e f_\theta(x_1,x_2,..,x_n}{\partial\theta}\right\}^2$ exists and positive for all $\theta \in$ (H).

Here, A is the domain of x_1, x_2,..,x_n for which f_θ (x_1, x_2,...,x_n) is positive.

An example where the above regular conditions are not satisfied is the reactangular distribution $R(0, \theta)$.

Example : Consider X_1, X_2,..,X_n as a random sample from $N(\mu,\sigma^2)$. Then the sample mean $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ is distributed as $N\left(\mu, \frac{\sigma^2}{n}\right)$. Then question that we may ask : Is $\bar{X}$ a MVUE of μ?

It is well-known that $\bar{X}$ is unbiased for μ. To check if $\bar{X}$ has the minimum variance, we calculate $E\left(\frac{\partial^2 L}{\partial\mu^2}\right)$ as follows :

We have $L = \log_e \left\{\prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{X_i-\mu}{\sigma}\right)^2}\right\}$

$$= \text{const} - \frac{1}{2}\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2$$

$$= \text{const} - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(X_i - \mu)^2$$

Therefore, $$\frac{\partial L}{\partial\mu} = -\frac{1}{2\sigma^2}\cdot 2\sum_{i=1}^{n}(X_i - \mu)(-1)$$

$$= \frac{1}{\sigma^2}\{n\bar{X} - n\mu\}$$

$$= \frac{n}{\sigma^2}(\bar{X} - \mu),$$

$$\text{or, } \frac{\partial^2 L}{\partial \mu^2} = -\frac{n}{\sigma^2}$$

Hence, finally, $-E\left(\frac{\partial^2 L}{\partial \mu^2}\right) = \frac{n}{\sigma^2}$

so that the lower bound of the variance of an unbiased estimator is $\sigma^2/_n$ which is also the variance of $\bar{X}$ in this case. Hence, $\bar{X}$ is a MVUE of μ.

Example : Let $X_1, X_2, \ldots, X_n$ be a random sample from a Bernoulli distribution with parameter θ, which is assumed to be unknown.

The *pmf* is given by

$$f(x) = \theta^x (1 - \theta)^{1-x},\ x = 0, 1$$

so that $\log_e f(x) = x \log_e \theta + (1 - x) \log_e)(1 - \theta)$

$$\therefore \frac{\partial \log e f(x)}{\partial \theta} = \frac{x}{\theta} - \frac{1-x}{1-\theta} = \frac{x-\theta}{\theta(1-\theta)}$$

Consequently, $E\left(\frac{\partial \log e f(x)}{\partial \theta}\right)^2 = \frac{1}{\theta^2(1-\theta)^2} E\,(x - \theta)^2,$

$$= \frac{\theta(1-\theta)}{\theta^2(1-\theta)^2},\ (\because E(x - \theta)^2 = V(x) = \theta(1 - \theta))$$

$$= \frac{1}{\theta(1-\theta)}.$$

The Cramer-Rao lower bound to the variance of an unbiased estimator for θ is

$$\frac{1}{nE\left(\frac{\partial \log_e f(x)}{\partial\theta}\right)^2} = \frac{1}{n.\frac{1}{\theta(1-\theta)}} = \frac{\theta(1-\theta)}{n}$$

Consider the sample mean $\bar{X}$ which is unbiased for θ. Also, we know that

$$V(\bar{X}) = \frac{V(X)}{n} = \frac{\theta(1-\theta)}{n},$$

which coincides with the Cramer-Rao lower bound. Thus, $\bar{X}$ is the MVUE of θ.

Note : If $X_1, X_2,...,X_n$ are *i.i.d.* with common *pmf* (or *pdf*) $f(x)$, then

$$E\left(\frac{\partial \log_e f(X_1,X_2,..,X_n)}{\partial\theta}\right)^2$$

$$= nE\left(\frac{\partial \log_e f(x)}{\partial\theta}\right)^2,$$

so that the Cramer-Rao lower bound is

$$V(T) \geq \frac{1}{nE\left(\frac{\partial \log_e f(x)}{\partial\theta}\right)^2}$$

In fact, it can be further shown that

$$V(T) \geq -\frac{1}{nE\left(\frac{\partial_2 \log_e f(x)}{\partial\theta^2}\right)}.$$

4.3.1 Condition for equality of Cramer-Rao Inequality

The equality in the Cramer-Rao Inequality is achieved if

$$\frac{\partial L}{\partial\theta} = C(\theta)\,(T - \theta)$$

or, $L = \int C(\theta)\,(t - \theta)\,d\theta$ + constant, or, $\log_e f_\theta(x_1, x_2,...,x_n) = \int C(\theta)(t-\theta)\,d\theta$ + constant,

On further simplification, we get

$$f_\theta(x_1, x_2,...,x_n) = e^{\int C(\theta)(t-\theta)d\theta} \cdot m,$$

$$= g_\theta(t).m, \text{ where}$$

m is free of θ, but may depend on $x_1, x_2,...,x_n$

Thus, in order that θ may have T as an MUVE, T must be a sufficient statistic for θ.

Example : Suppose that $X_1, X_2,...,X_n$ are a random sample from $N(M, \theta)$, where μ is the mean and θ is the variance of the distribution. Suppose also that μ is known. Without any loss of generality, let us assume that $\mu = 0$.

Then the *pdf* of the common X is given by

$$f(x) = \frac{1}{\sqrt{2\pi\theta}} e^{-\frac{1}{2}\frac{x^2}{\theta}}$$

Then, the logarithm of $f(x)$ is

$$\log_e f(x) = \text{const} - \frac{1}{2}\log_e \theta - \frac{1}{2}\frac{x^2}{\theta}$$

Hence $\dfrac{\partial \log_e f(x)}{\partial\theta} = -\dfrac{1}{2\theta} + \dfrac{1}{2}\dfrac{x^2}{\theta^2}$

and $\dfrac{\partial^2 \log_e f(x)}{\partial\theta^2} = \dfrac{1}{2\theta^2} - \dfrac{x^2}{\theta^3}$

$$\therefore E\left(\frac{\partial^2 \log_e f(x)}{\partial\theta^2}\right) = \frac{1}{2\theta^2} - \frac{V(x)}{\theta^3}$$

$$= \frac{1}{2\theta^2} - \frac{1}{\theta^2},$$

$$= \frac{1-2}{2\theta^2},$$

$$= -\frac{1}{2\theta^2}.$$

Then Cramer-Rao lower bound to the variance of an unbiased estimator of θ is

$$\frac{-1}{nE\left(\frac{\partial^2 \log_e f(x)}{\partial\theta^2}\right)}$$

$$= \frac{2\theta^2}{n}.$$

Define $S_0{}^2$ as $S_0{}^2 = \frac{1}{n}\sum_{i=1}^{n} x_i{}^2$

But $\sum_{i=1}^{n} x_i^2/_\theta$ follows $a\chi^2$ - distribution with n df

$$\therefore E\left(\sum_{i=1}^{n} x_i^2/_\theta\right) = n$$

or, $E\left(\frac{1}{n}\sum_{i=1}^{n} x_i^2\right) = \theta,$

so that $\frac{1}{n}\sum_{i=1}^{n} x_i^2$ is an unbiased estimator of θ. Its variance is

$$V\left(\sum_{i=1}^{n} x_i^2/_\theta\right) = 2n$$

or, $$V\left(\frac{1}{n}\sum_{i=1}^{n}x_i^2\right)=\frac{2n\theta^2}{n^2}$$

$$=\frac{2\theta^2}{n},$$

which coincides with the Cramer-Rao lower bound, as derived above. Thus, $\frac{1}{n}\sum_{i=1}^{n}x_i^2$ is the minimum variance unbiased estimator (MVUE) of the variance.

Note : It is to be remembered that the Cramer-Rao lower bound is valid for the distributions that satisfy certain regularity conditions, including interchangeability of differentiation and integration. This requires, among other things, that the range of x is independent of θ. So, Cramer-Rao lower bound will not be valid for the density

$$f(x)=\frac{1}{\theta},\ 0\le x\le\theta.$$

4.4 Method of generating minimum variance unbiased estimator (MVUE), Rao-Blackwellization

We have seen in the preceeding section that given an unbiased estimates T of a parameter θ of some distribution, it is possible to check whether this estimates is the minimum variance unbiased estimator by comparing its variance with the lower bound of the Cramer-Rao Inequality. In case, this proposed estimator is not an MVUE, there is no way to improve the estimator in this approach.

Rao-Blackwellization is a technique of generating a minimum variance unbiased estimator (MVUE). If we are given an unbiased estimator U, then we can improve upon U by forming a new estimator $\phi(T)$ based on U and a sufficient statistic T.

The following theorem, called Rao-Blackwell Theorem, enables us to obtain an MVU estimator from any unbiased estimator by using a sufficient statistic.

4.4.1 Rao-Blackwell Theorem

Let the statistic $U = U(X_1, X_2,....,X_n)$ be an unbiased estimator of $\gamma(\theta)$, a function of θ. Also, $T(X_1, X_2,...,X_n)$ is a sufficient statistic for θ.

Define $\phi(T) = E(U/T)$, which is the conditional mean of U, given T.

Then (i) $E(\phi(T)) = \gamma(\theta)$

and (ii) $V(\phi(T) \leq V(U)$,

where $V(\cdot)$ stands for variance.

Proof : We have $\underset{\theta}{E}(U) = \underset{\theta}{E}E(U/T)$

$$= \underset{\theta}{E}(\phi(T), \text{ for all } \theta.$$

But $\underset{\theta}{E}(U) = \gamma(\theta)$, so that $\underset{\theta}{E}(\phi(T)) = \gamma(\theta)$.

This shows $\phi(T)$ is also unbiased for $\gamma(\theta)$.

Again, $\underset{\theta}{V}(U) = \underset{\theta}{E}V(U/T) + \underset{\theta}{V}E(U/T)$, for all θ.

$$= \underset{\theta}{E}V(U/T) + \underset{\theta}{V}(\phi(T))$$

But $\underset{\theta}{E}V(U/T) \geq 0$,

$$\therefore \underset{\theta}{V}(U) \geq \underset{\theta}{V}(\phi(T))$$

or, $\underset{\theta}{V}(\phi(T)) \leq \underset{\theta}{V}(U)$, for all θ.

Thus, $\phi(T)$ is an MVUE of $\gamma(\theta)$.

Uniqueness : Let $\phi(T)$ and $\phi_1(T)$ be both MVUE for $\gamma(\theta)$.

Then, $\underset{\theta}{E}(\phi(T)) = \gamma(\theta)$, for all θ,

and $\underset{\theta}{E}(\phi_1(T) = \gamma(\theta)$, for all θ.

$\therefore \underset{\theta}{E}(\phi(T) - \phi_1(T)) = 0$, for all θ. → (1)

If the distribution of T is complete, then the last equation (1) would suggest that

$\phi(T) = \phi_1(T)$, almost everywhere. i.e., $\phi(T)$ is unique.

Thus, the estimator obtained by using Rao-Blackwellization is not only an MVUE, but is also unique if the underlying distribution is complete.

4.4.2 Application of Rao-Blackwell and Lehmann-Scheffe theorems

Example 3.1 Let $X_1, X_2,...,X_n$ be a random sample from a Poisson distribution with *p.m.f.*

$$f_\theta(x) = \frac{e^{-\theta}\theta^x}{\lfloor x}, \ x = 0, 1, 2,...; \ \theta > 0.$$

Find an MVUE for $P(X = k)$, assuming that θ is unknown.

Solution : First note that $P(X = k) = \dfrac{e^{-\theta}\theta^k}{\lfloor k} = \gamma(\theta)$, say.

Let $X_1, X_2,...,X_n$ be the given random sample from $f_\theta(x)$.

Definea random variable Y such that

$Y = 1$, if $X_1 = k$

$= 0$, otherwise.

Then, $E_\theta(Y) = 1 \cdot P(X_1 = k) + o \cdot P(X_1 \neq k)$

$$= \frac{e^{-\theta}\theta^k}{\lfloor k}$$

$$= P(X = k) = \gamma(\theta).$$

Thus, Y is an unbiased estimator of $\gamma(\theta)$.

Again, it is known that $T = \sum_{i=1}^{n} X_i$ is a sufficient statistic for θ.

Consider now the statistic $\phi(t)$ defined by

$$\phi(t) = \underset{\theta}{E}(Y/t)$$

$$= P(X_1 = k/T = t)$$

$$= \frac{P\{(X_1 = k) \cap (T = t)\}}{P(T = t)}$$

$$= \frac{P(X_1 = k, \sum_{i=1}^{n} X_i = t - k)}{P(T = t)}$$

$$= \frac{\frac{e^{-\theta}\theta^k}{\underline{|K}} \cdot \frac{e^{-(n-1)\theta}(n-1)\theta)^{t-k}}{\underline{|t-k}}}{\frac{e^{-n\theta}(n\theta)^t}{\underline{|t}}}, \; (\because \sum_{i=1}^{n} X_i \text{ follows a Poisson}$$

distribution with parameter $(n - 1)\theta$)

$$= \frac{\underline{|t}}{\underline{|k}\,\underline{|t-k}} \cdot \frac{(n-1)^{t-k}}{n^t}$$

$$= \binom{t}{k} \frac{(n-1)^{t-k}}{n^t}.$$

Thus, an MVUE of $P(X = k)$ is $\binom{T}{k}\frac{(n-1)^{T-k}}{n^T}$, which is also unique as the Poisson distribution is complete.

Example 3.2 : Let $X_1, X_2, \ldots, X_n$ be *i.i.d.* $b(1, p)$. Then $\sum_{i=1}^{n} X_i = T$ is sufficient for p. Again, this distribution is complete. Let us find the MVUE of $d(p) = p(1 - p)$.

Solution.

We have $E(nT) = nE(T)$

$= n.np$ ($\because T$ follows $b(n, p)$)

$= n^2p$.

and $E(T^2) = V(T) + (E(T))^2$

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

$\therefore E(nT - T^2) = n(n-1)\, p\,(1-p)$, on simplification.

Finally, $E\left(\dfrac{nT-T^2}{n(n-1)}\right) = p\,(1-p)$, so that

$\dfrac{nT-T^2}{n(n-1)}$ is unbiased for $d(p) = p\,(1-p)$.

Also, $E\left(\dfrac{nT-T^2}{n(n-1)}/T\right) = \dfrac{nT-T^2}{n(n-1)}$

which by Rao-Blackwell theorem, is the MVUE of $p(1-p)$.

Example 3.3 : Let $X_1, X_2, \ldots\ldots, X_n$ be $N(\theta, 1)$. Find the MVUE of $d(\theta) = \theta^2$.

Solution. We first obtain an unbiased estimator of $d(\theta) = \theta^2$.

We have $V(\bar{X}) = E(\bar{X}^2) - (E(\bar{X}^2))$,

or, $\quad \dfrac{1}{n} = E(\bar{X}^2) - \theta^2$,

or, $\quad \theta^2 = E(\bar{X}^2 - \dfrac{1}{n})$, so that

$\bar{X}^2 - \dfrac{1}{n}$ is an unbiased estimator of θ^2.

Again, since variance is known, the sample mean $\bar{X}$ is sufficient for θ, so that $\bar{X}^2 - \dfrac{1}{n}$ is a function of the sufficient statistic. In other words,

$$E(\bar{X}^2 - \frac{1}{n} \mid \bar{X}) = \bar{X}^2 - \frac{1}{n}.$$

Thus, $\bar{X}^2 - \dfrac{1}{n}$ is an MVUE of $d(\theta) = \theta^2$.

Furthermore, $\bar{X}$ is a complete sufficient statistic, or equivalently, the normal distribution is complete.

Hence, $\bar{X}^2 - \frac{1}{n}$ is the unique MVUE of $d(\theta) = \theta^2$.

Example 3.4 : Let X follow $N(\mu, \sigma^2)$ i.e., a normal distribution with unknown mean μ and unknown variance σ^2. Consider a random sample X_1, X_2,...,X_n from the distribution. Define

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

being respectively the sample mean and sample variance. Then $\bar{X}$ and S^2 are statistically independent.

We first note that

(a) $\bar{X}$ and S^2 are jointly sufficient for (μ, σ^2),

and (b) ($\bar{X}$, S^2) has a complete family of distributions.

In other words, the pair ($\bar{X}$, S^2) is complete sufficient for (μ, σ^2).

It may be noted that $\frac{ns^2}{\sigma^2} = \sum_{i=1}^{n}\left(\frac{X_i - \bar{X}}{\sigma}\right)^2$ has a χ^2 - distribution with $(n-1)$ degrees of from. In other words, $\sum_{i=1}^{n}(X_i - \bar{X})^2$ has a distribution which is free of μ, so that it is an ancillary statistic for μ, as it contains no information for μ.

Also, $\bar{X}$ is sufficient for μ. By Basu's theorem we thus have the important result that $\bar{X}$ and S^2 are independently distributed. This is called a characterising property of a normal distribution.

4.5 Method of Maximum Likelihood Method of Moments

The most important of all the methods of estimation is the method of maximum likelihood. Its importance lies in the fact that it generally yields very good estimators as judged from various criteria.

The first define the likelihood function. Let $X_1, X_2, ..., X_n$ be a random sample from some pdf (or *pmf*) $f_\theta(x)$, and let $f_\theta(x_1, x_2, ..., x_n)$ be their joint distribution.

For given $x_1, x_2, ..., x_n$ and unknown θ, the function $f_\theta(x_1, x_2, ..., x_n)$ may be looked upon as a function of θ only, and we may write

$$L(\theta) = f_\theta (x_1, x_2, ..., x_n).$$

In case of a random sample $x_1, x_2, ..., x_n$ we have

$$L(\theta) = \prod_{i=1}^{n} f_\theta(x_i)$$

which is known as livelihood function.

4.5.1 Maximum Likelihood Estimate or (MLE) :

The MLE estimator of θ, say, $\hat{\theta}$ is such that

$$L(\hat{\theta}) \geq L(\theta), \text{ for all } \theta \in (\text{H}).$$

or, $$L(\hat{\theta}) = \sup_{\theta \in (H)} L(\theta)$$

In practice, it is easier to work with loge $L(\theta)$

than L (θ), so that we have

$$\log L(\hat{\theta}) \geq \log L(\theta), \text{ for all } \theta.$$

or, $$\log L(\hat{\theta}) = \sup_{\theta \in (H)} \log L(\theta)$$

When the supremum is attained at an interior point of (H) and $L(\theta)$ is a differentiable function of θ throughout the interval (H), the partial derivative of loge $L(\theta)$ will vanish at the point $\hat{\theta}$, where $\hat{\theta}$ is the solution of

$$\frac{\partial \log_e L(\theta)}{\partial \theta} = 0.$$

This equation is called the likelihood (or log-likehood) equation and $\hat{\theta}$ is the maximum likelihood estimate.

Example 4.1: Consider the rectangular distribution $R(0,\theta)$

whose *pdf* is

$$f_\theta(x) = \frac{1}{\theta},\ 0 \le x \le \theta.$$

Given the random sample $X_1, X_2, ..., X_n$, the likelihood function is

$L(\theta) = f_\theta\ (x_1, x_2, ..., x_n)$

$$= \frac{1}{\theta^n},\ \theta \ge x_i\,(i = 1,2,...,n)$$

Here, the likehihood equation $\frac{\partial \log_e L(\theta)}{\partial\theta} = 0$ has no solution.

Let $X_{(1)} \le X_{(2)} \le X_{(3)} \le X_{(n)}$ be the ordered observations of the sample, with $X_{(n)}$ being the largest observation.

So, $L(\theta) = \frac{1}{\theta^n}$, for $\theta \ge x_{(n)}$

But $L(\theta)$ will be the maximum if θ takes the minimum value, which is $X_{(n)}$ here.

So, $\hat{\theta}_{ML} = X_{(n)}$ is the maximum likelihood estimator (MLE) of θ for the rectangular distribution.

Example 4.2 : Consider a set of n Bernoullian trials with probability of success θ. With the i^{th} trial we associate a variable X_i with the probability mas function.

$$f(x_i) = \theta^{x_i}(1-\theta)^{1-x_i}, x_i = \theta, 1$$

The joint probability mass function of $x_1, x_2, ..., x_n$ is

$$f(x_1, x_2, ..., x_n) = \theta^{\sum_{i=1}^{n} x_i} (1-\theta)^{n-\sum_{i=1}^{n} x_i}$$

$$= \theta^{n\hat{\theta}} (1-\theta)^{n-n\hat{\theta}}, \text{ where } \hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

The function $f(x, x_2, ..., x_n)$, treated as a function of θ and written as $L(\theta)$, is called the likelihood function. Thus,

$$L(\theta) = \theta^{n\hat{\theta}} (1-\theta)^{n-n\hat{\theta}}$$

Taking logarithm,

$$\log_e L(\theta) = n\hat{\theta} \log_e \theta + (n - n\hat{\theta}) \log_e (1-\theta)$$

Solving $\frac{\partial \log_e L(\theta)}{\partial \theta} = 0$, we have

$$0 = \frac{1}{L(\theta)} L'(\theta) = \frac{n\hat{\theta}}{\theta} + \frac{n(1-\hat{\theta})}{1-\theta}(-1)$$

or, $\frac{n\hat{\theta}}{\theta} = \frac{n(1-\hat{\theta})}{1-\theta}$ or, $\theta(1-\hat{\theta}) = \hat{\theta}(1-\theta)$

or, $\theta(1-\hat{\theta}+\hat{\theta}) = \hat{\theta}$ or, $\theta = \hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} x_i$,

which is the maximum-likehihood estimator (MLE) of the parameters θ, and is seen to be the sample mean.

Example 4.3: Let $X_1, X_2, ..., X_n$ be a random sample from a Poisson distribution with parameters λ. Then the likelihood function is

$$L(\lambda)=\frac{e^{-n\lambda}\lambda^{\sum_{i=1}^{n}X_i}}{\prod_{i=1}^{n}(X_i!)}$$ and log-likelihood is

$$\log_e L(\lambda)=(-n\lambda)+\left(\sum_{i=1}^{n}X_i\right)\log_e\lambda-\sum_{i=1}^{n}\log_e(X_i!)$$

So, the equation $0=\frac{\partial\log_c L(\lambda)}{\partial\lambda}$ gives $-n+\frac{\sum_{i=1}^{n}X_i}{\lambda}=0$

or, $\hat{\lambda}=\frac{1}{n}\sum_{i=1}^{n}X_i$,

which is the maximum-likelihood estimator (MLE) of λ. Thus, in this example, the sample mean is the MLE of λ.

Example 4.4 : Consider the MLE of θ of the exponential distribution $f(x)=\frac{1}{\theta}e^{-\frac{x}{\theta}}$, $0 < x < \theta$, $\theta > 0$.

Based on a random sample X_1, X_2, ..., X_n, the likelihood equation is

$$L(\theta)=\left(\frac{1}{\theta}\right)^{m}e^{-\frac{1}{\theta}}\sum_{i=1}^{n}X_i$$

So, $0=\frac{\partial\log_e L(\theta)}{\partial\theta}$ gives

$$0=\frac{\partial}{\partial\theta}\left[-n\log_e\theta-\frac{1}{\theta}\sum_{i=1}^{n}Xi\right]=-\frac{n}{\theta}+\frac{1}{\theta^2}\sum_{i=1}^{n}X_i$$

Or, $\hat{\theta}=\frac{1}{n}\sum_{i=1}^{n}X_i$, which is the MLE of θ.

Example 4.5 : Suppose X_1, X_2, ..., X_n, are independent random observations from a normal distribution with mean μ and variance σ^2.

Case 1 : μ known and σ unknown.

In this case, the likelihood function is

$$L(\sigma) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \exp\left[-\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{2\sigma^2}\right]$$

The log-likelihood is given by

$\log_e L(\sigma)$ = constant independent of σ

$$+(-n)\log_e \sigma - \frac{\sum_{i=1}^{n}(X_i-\mu)^2}{2\sigma^2}$$

Hence, the equation $0 = \frac{\partial \log_e L(\sigma)}{\partial\sigma}$ gives

$$0 = -n/\sigma - \left(\frac{1}{2}\right)(-2)\sigma^{-3}\sum_{i=1}^{n}(X_i-\mu)^2$$

or, $\frac{1}{\sigma^3}\sum_{i=1}^{n}(X_i-\mu)^2 = \frac{n}{\sigma}$ or, $\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2$

Finally, $\hat{\sigma} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2}$, which is the MLE of σ.

Case 2 : μ unknown and σ^2 known.

In this case, we obtain the maximum likelihood estimator by first writing the likelihood function as

$L(\mu)$ = constant independent of μ + exp $\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(X_i-\mu)^2\right]$,

so that the log-likelihood is given by

$$\log_e L(\mu) = \text{constant } -\frac{1}{2\sigma^2}\sum_{i=1}^{n}(X_i-\mu)^2$$

Solving $0=\frac{\partial \log_e L(\mu)}{\partial\mu}$, we have

$$0=-\frac{1}{2\sigma^2}\sum_{i=1}^{n}2(X_i-\mu)(-1),$$

or, $0=\sum_{i=1}^{n}(X_i-\mu)$ or, $\hat{\mu}=\frac{1}{n}\sum_{i=1}^{n}X_i$,

which is the sample mean.

Thus, in this case, the sample mean is the MLE of μ.

Case 3 : Both μ and σ are unknown.

Here, the likelihood function is given by

$$L(\mu,\sigma)=\frac{1}{\left(\sigma\sqrt{2\pi}\right)^n}\exp\left[-\sum_{i=1}^{n}(X_i-\mu)^2/\left(2\sigma^2\right)\right]$$

The log-likelihood is given by

$$\log_e L\,(\mu,\sigma) = -n\log_e\left(\sigma\sqrt{2\pi}\right)-\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{2\sigma^2}$$

Since both μ and σ are unknown, the MLE's of μ and σ are obtained by solving the simultaneous equations :

$$0=\frac{\partial \log_e L(\mu,\sigma)}{\partial\mu},\quad 0=\frac{\partial \log_e L(\mu,\sigma)}{\partial\sigma}$$

On solving these equations, we obtain the MLE's of the unknown μ and σ as

$\hat{\mu} = \bar{X}$, the sample mean

and $\hat{\sigma} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2}$, the sample standard deviation.

The properties of maximum likelihood estimation :

Apart from their intuitive appeal, maximum likelihood estimates possess several nice properties :

1. Consistency : The maximum likelihood estimators (MLE) $\hat{\theta}$ of θ is, under general conditions, a consistent estimator.
2. Asymptotic normality: Under general conditions, $\hat{\theta}$ is esymptotically normally distributed with mean θ and variance $\frac{1}{nE\left(\frac{\partial \log_e f(x)}{\partial\theta}\right)^2}$, which the lower bound of the variance of Cramer-Rao Inequality.
3. Among all asymptotically normal consistent estimators of θ, $\hat{\theta}$ is generally efficient.
4. If there exists a sufficient statistics for θ, then $\hat{\theta}$ is also sufficient, or is a function of a sufficient statistic.
5. Unbiasedness : Generally, $\hat{\theta}$ is not unbiased.
6. Invariance : If $\hat{\theta}$ is the maximum likelihood estimator of θ, then $\psi\left(\hat{\theta}\right)$ is also the maximum likelihood estimator of $\psi\left(\hat{\theta}\right)$, ψ being a single-valued function of θ with a unique inverse.
7. May not always exist.

4.5.2 Method of Moments :

The particularly simple method of estimating parameters is the method of moments. Supposting these are k parameters to be estimated, the method consists of the following steps :

1. To express the moments of the theoretical distribution in terms of the parameters.
2. To equate the first k theoretical moments which are expressed in terms of the k parameters to the corresponding sample moments.
3. Finally, to solve the k resulting equations to determine the k parameters. The k parameters will, therefore, be obtained in terms of the k sample moments, *i.e.*, $m_1'(=\bar{x})$, m_2', ..., m_k'.

Mathematically, the process can be explained as follows :

Let $X_1, X_2, ..., X_n$ be a random sample from some distribution having k parameters.

Define $$m'_\sigma = \frac{1}{n}\sum_{i=1}^{n} X_i, \ (r = 1, 2, 3, ..., k),$$

where m'_r is the rth order sample raw moment, about zero and $m'_1 = \bar{X}$. Similarly, define

$$\mu'_r = E(X^r), (r = 1, 2, ..., k).$$

where μ'_r is the r^{th} order population raw moment about zero, and $\mu'_r = E(X)$.

We first equate μ'_r with m'_r to have the k simultaneous equations.

$$\mu'_r = m'r \ (r = 1, 2, ..., k).$$

But each μ'_r depends on k parameters. By solving the k equations for k unknown parameters, we get the moment estimates of the parameters.

Example 4.6 : Consider the binomial distribution $X \sim b\ (n, p)$, where p is unknown, we have

$$E(X) = np \quad \text{or,} \quad p = \frac{E(X)}{n}$$

By equating $E(X)$ with the sample mean $\bar{X}$, we have an estimate of p as

$$\hat{p} = \frac{\bar{X}}{n}.$$

This $\hat{p}$ will be the moment estimate of p .

4.6 Summary

In this chapter, the theory of estimation has been discussed. Terms like 'statistic', 'parameter', 'estimator', etc. have been defined. The purpose of estimation has been elaborated.

The properties of estimators, e.g., unbiasedness, consistency, sufficiency and efficiency have been explained with examples. Completeness of a distribution is defined. The role of Basu's theorem has been discussed with example.

Cramer-Rao Inequelity has been stated and proved. Numerous example sare given. The method of deriving the minimum variance unbiased estimator (MVUE) has been thoroughly explained. The importance of Rao-Blackwell theorem has been elaborated with numerous examples.

The method of maximum likelihood estimation and moment method are also mentioned.

4.7 Exercises

1. Distinguish between the term 'statistic' and 'parameter', with examples.
2. Why do we need to study estimation? What do we estimate? Does it give the exact value of the quantity that we assume to be unknown?
3. Mention some good properties of the estimator. which do you think to be the most important?
4. Give an example of an estimator in each of the following cases :

   (a) unbiased and consistent, (b) biased but consistent,

   (c) sufficient but biased, (d) sufficient and unbiased.
5. Describe the maximum likelihood method of estimation. What are the properties of a maximum likelihood estimator?
6. Let X follow a binomial distribution $b(n,\theta)$, where θ is unknown. Obtain an unbiased estimetor of θ. Is it also UMVUE?

7. Let X follow $N(M, \sigma^2)$ bistribution, why μ and σ^2 both unknown. Obtain sufficient statistics for μ and σ^2. Are they unbiased?

8. Find the maximum-likelihood estimator of $\frac{1}{P}$, based on a single observation x, for the distribution

$$f(x) = P(1-P)^{x-1}, \text{ for } x = 1, 2, 3, \ldots$$

9. Let $X \sim P(\lambda)$, a Poisson distribution with parameter λ. Show that X is the UMVUE of λ.

10. Find the lower bound for the variance of an unbiased estimator of θ based on a sample of size n for the distribution

$$f(x) = \frac{1}{\theta} e^{-x/\theta}, \ x > 0, \ \theta > 0.$$

Is this estimator UMVUE?

Unit 5 □ Testings Statistical Hypothesis

Structure

5.0 Objectives

The followings are discussed here:

- Statistical Hypothesis
- Null and Alternative Hypothesis
- Type I and Type II error
- Level of Significance
- Acceptance and Rejection Region
- Power of a test

5.1 Introduction

The purpose of this unit is to decade whether a statement about an unknown parameter is true. As the decision is based on samples, and not on the entire data, there is a posibility of ariving at wrong conclusions. Testing procedures are avoidable where the probability of such wrong decisions is minimized. These are the issues discussed in this unit.

Another aspect of testing is confidence interval. A point estimator assigns a single value to an unknown parameter, but an interval estimator assigns an interval to the parameter which is supposed to cover it with high probability. This interval is called confidence interval. We discuss all these with examples.

5.2 Population and Sample

A population is the set or collection of all conceivable and identifiable units under study. On the contrary, a sample is a part of the population.

In statistical inference, we draw conclusions about the unknown characteristics of the population, such as its mean, variance, skewness, etc., on the basis of data collected from the sample. A sample, being a part of the population, may not give exact value of the population characteristics. For example, the population mean of the students heights, may be 165 cm. while the sample estimates it as 160 cm, say. This difference is termed as sampling fluctuation, and it arises due to the fact that we are studying only a part of the population.

Since we draw conclusion on the basis of samples, it is desirable that the samples should be representative of the population. The more this is ensured, the better will be the inferences about the unknown characters of the population. Random sampling technique is the most widely used technique of drawing representative samples from a population.

5.2.1 Statistical Hypothesis

A statistical hypothesis is a tentative statement or assertion about the unknown population characteristic (s).

We distinguish between two types of hypothesis, namely, null hypothesis, say,

H_O and aternative hypothesis, say, H_1. In general, H_o is a statement about the unknown parameter, say θ, which we believe to be true. When a null hypothesis. H_O is rejected, we accept the alternative hypothesis H_1 with some degree of confidence associated with it. For example, let $X_1, X_2, ..., X_n$ be a random sample from a normal distribution N(θ,1), where θ is unknown. The only krowledge about θ is that it lies on the real line. Then, our H_O and H_1 might be $H_O : \theta = \theta_o$ and $H_1 : \theta = \theta_1, \theta_0 \neq \theta_1$, respectively. Here, we test H_0 against H_1. If on the basis of the data, H_0 is rejected, we accept H_1 as a plausible value of θ.

We consider a second example. In coin-tesing experiments, one frequently assumes that the coin is fair, i.e., the probability p of getting heads is $\frac{1}{2}$. How does one test whether the coin is fair? If one is guided by intuition, a reasonable procedure would be to toss the coin n times, say, and count the number of heads. If the proportion of heads out of n tosses does not deviate"too much" from $p = \frac{1}{2}$, one would tend to conclude that the coin is fair.

5.2.2 Type I and Type 2 error

Since we take our decision either to accept or to reject the hypothesis H_0 on the basis of a random sample, we are likely to commit two types of errors :

(a) We may reject H_0 when H_0 is true. This is the case of rejecting a true hypothesis. This error is termed as 'Type 1 error'.

(b) We may accept H_0 when H_0 is false. This is the case of accepting a false hypothesis. This error is termed as 'Type 2 error'.

This is indicated in the following table :

Table 5.1

Type 1 and Type 2 error

True situation/Decision made	H_0 true	H_0 false

H_0 rejected	Type 1 error	correct decision
H_0 accepted	correct decision	Type 2 error

Level of significance : The level at which the probability of type 1 error is set is called the level of significance and is denoted by α. This being the probability of rejecting a true hypothesis, should be very small, such as 0.05, 0.01, etc.

Power of a test : The probability of Type 2 error, i.e., the probability of accepting a false hypothesis, is denoted by β, which is also small. Then, $1-\beta$ is the power of the test which is the probability of rejecting a false hypothesis. The larger the power of a test, the better will be its performance.

Uniformly Most Powerful (UMP) test :

Consider the hypothesis $H_0 : \theta = \theta_0$ against $H_1 : \theta = \theta_1$. If $\theta_0$and θ_1 are both numbers, then the hypothesis are simple. If θ_0 is a number, but θ_1 is a set, say, subset of the real line, then H_1 is composite. The case of UMP test arises in the latter case.

Let us consider the case of testing of the simple hypothesis

$$H_0 : \theta = \theta_0$$

against the composite alternative hypothesis

$$H_1 : \theta \neq \theta_0$$

A test is said to be UMP test if it satisfies the following two conditions :

1. The test is of size α,

2. Its power is greater than or equal to the power of any other test of the same size.

Note that simple hypothesis completely specifies a distribution, while a composite hypothesis does not. Moreover, in case of a composite alternative hypothesis of the form $H_1 : \theta \neq \theta_0$ we will get a power curve$\beta(\theta)$, say, for varying θ instead of a single power. Our aim is to obtain a test that maximies power (or miximizes type 2 error) subject to given level of significance. The now celebrated Neyman-Pearson Lemma

(NP Lemma) gives a complete solution to this problem.

5.2.3 Acceptance and rejection region

A test of a statistical hypothesis practically means a demarcation of the sample space into two regions–acceptance region, and rejection region (also called critical region). For example, in the case of coin-tossing experiments for testing $p=\frac{1}{2}$, observed values of heads for away from 50 will constitute the rejection region if we toss the coin 100 times. The suitable demarcation of the sample space into acceptance and critical region is the main purpose of testing statistical hypothesis.

It is to be clearly understood that we cannot control or minimize Type 1 and Type 2 error simultaneously. This will be clear from the following consideration. If we always accept H_0, whatever be the data, then there will be no occasion of rejecting H_0 whether it is true or false, and we have $\alpha=0$, but in this case $\beta=1$, as in this case a false hypothesis will be accepted.

Conversely, if we always reject H_0 whatever be the data, then there will be occasion when a true hypothesis H_0 will be rejected and we have $\alpha=1$, but in this case $\beta=0$, as a false hypothesis H_0 will not be accepted. Thus, both α and β cannot be made zero simultaneously.

That is why, we keep α fixed at a lower level and minimize β, or maximize $1-\beta$, which is the power of the test.

5.3 Confidence co-efficient, Confidence interval etc.

5.3.1 Relation between tests and confidence intervals

A confidence interval is a statement about the unknown population parameter.While a point estimator guesses the value of the unknown parameter by a single number, an interval estimator identifies an interval,which is a random interval, which is likely to contain the unknown parameter. We attach a probability, called confidence co-efficient, to this interval so that the resulting interval now becomes confidence interval with confidence co-efficient $1-\alpha$, α being the level of significance of a test as discussed earlier. Clearly, larger is the value of $1-\alpha$, better is the chance that the confidence internal will contain the unknown parameter.

Thus, let $X_1, X_2, \ldots, X_n$ be a random sample from some population with $p.d.f. \underset{\theta}{f}(x), or\ p.m.f. p_\theta(x)$. Then, a point estimator $T(X_1, X_2, \ldots, X_n)$ estimates θ by a single number, a confidence interval $\left(l(X_1, X_2 \ldots, X_n), u(X_1, X_2 \ldots, X_n), (u>l)\right)$, estimates θ by an interval, which is random, and we write

$$P(l \leq \theta \leq u) = 1-\alpha,$$

Where $1-\alpha$ is the confidence co-efficient. Thus, if $\alpha = 0{\cdot}05, 0{\cdot}01$, etc. then the confidence co-efficients are 0·95, 0·99, respectively.

5.3.2 Tests and confidence internal for mean

(a) Let $X_1, X_2, \ldots, X_n$ be a random sample from $N(\mu, \sigma^2)$, σ^2 being known. We want to test $H_0: \mu = \mu_0$ against $H_1: \mu \neq \mu_0$. This is the case, when H_0 is simple but H_1 is composite.

Consider $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ as a point estimator of μ. Then, we known that

$\bar{X}$ follows $N\left(\mu, \frac{\sigma^2}{n}\right)$

Hence, $Z = \frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}}$ follows N(0,1).

This Z is the standard normal variable and is the test statistic in the problem. We then reject the null hypothesis $H_0: \mu = \mu_0$ if

$$\left|\frac{\bar{X}-\mu_0}{\frac{\sigma}{\sqrt{n}}}\right| \geq Z_{\frac{\alpha}{2}} \rightarrow (1)$$

when $Z_{\alpha/2}$ is the critical value to be obtained from a normal table corresponding to a given α. For α=0·05, the critical value $Z_{\alpha/2}=1{\cdot}96$. Similarly, for other values ofα.

The confidence interval for μ can be obtained from (1) by replacing μ_0 by μ, and rewriting it, in terms of probability, as (see Fig. 5.1)

$$P\left\{\left|\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\right| \geq Z_{\alpha/2}\right\} = \alpha,$$

or, $$P\left\{\left|\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}}\right| \leq Z_{\alpha/2}\right\} = 1-\alpha,$$

$$P\left\{\bar{X}-z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X}+z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right\} = 1-\alpha$$

so that $\left(\bar{X}-Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \bar{X}+Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$is the $100(1-\alpha)\%$

confidence interval for μ.

(b) Consider the case where X follows $N(\mu,\sigma^2)$, both μ and σ^2 unknown. To test $H_0:\mu=\mu_0$ against $H_1:\mu\neq\mu_0$.

In this case $\frac{\bar{X}-\mu_0}{s/\sqrt{n}}$, where $s=\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(X_1-\bar{X}\right)^2}$

under H_0 follows a t-distribution with (n–1) d.f., so that for α level of significance, the critical region is

$$\left|\frac{\bar{X}-\mu_0}{s/\sqrt{n}}\right|\geq t_{\alpha/2};(n-1)$$

when $t_{\alpha/2};(n-1)$ is to be obtained from a t-distribution.

Introducing probability and rewriting the above, we have on replacing μ_0 by μ,

$$P\left\{\left|\bar{X}-\mu\right|\geq t_{\alpha/2;(n-1)}\frac{s}{\sqrt{n}}\right\}=\alpha,$$

$$\text{or, } P\left\{\left|\bar{X}-\mu\right|\leq t_{\alpha/2;(n-1)}\frac{s}{\sqrt{n}}\right\}=1-\alpha\,,$$

$$\text{or, } P\left\{\bar{X}-t_{\alpha/2;n-1}\frac{s}{\sqrt{n}}\leq\mu\leq\bar{X}+t_{\alpha/2;n-1}\frac{s}{\sqrt{n}}\right\}=1-\alpha\,.$$

Thus, the interval $\left(\bar{X}-t_{\alpha/2;n-1}\frac{s}{\sqrt{n}},\bar{X}+t_{\alpha/2;n-1}\frac{s}{\sqrt{n}}\right)$

is the $100(1-\alpha)\%$ confidence interval for μ when σ^2 is unknown.

Thus, in the above two cases, we have seen that the case of testing and confidence interval are somehow related. If one knows the rejection region, or its complement the acceptance region, then the confidence interval with a specified confidence co-efficient can be determined in a straight forward manner.

5.3.3 Tests and confidence interval for variance

Let $X_1, X_2, ... X_n$ be iid random variables from $N(\mu, \sigma^2)$, when both μ and σ^2 are unknown. Our aim here is to test $H_0 : \sigma^2 = \sigma_0^2$ against $H_1 : \sigma^2 \neq \sigma_0^2$, and also obtain a confidence interval for σ^2.

First we note that

$$\chi^2_{n-1} = \frac{\sum_{i-1}^{n}(X_i - \bar{X})^2}{\sigma_0^2} = \frac{ns^2}{\sigma_0^2} \text{ follows } \chi^2_{n-1},$$

a chi-square distribution with (n–1) d.f., under H_0.

So, under the alternative $H_1 : \sigma^2 \neq \sigma_0^2$, we reject $H_0 : \sigma^2 = \sigma_0^2$ in favour of H_1 if

either observed $\chi^2_{n-1} < \chi^2_{1-\alpha/2;n-1}$

or, observed $\chi^2_{n-1} > \chi^2_{\alpha/2;n-1}$,

When the critical values $\chi^2_{1-\frac{\alpha}{2};n-1}$ and $\chi^2_{\frac{\alpha}{2};n-1}$

are obtained from the χ^2 – table for given α.

To obtain the confidence interval for σ^2, we consider the probability

$$P\left\{\chi^2_{1-\frac{\alpha}{2};n-1} \leq \chi^2_{n-1} \leq \chi^2_{\frac{\alpha}{2};n-1}\right\} = 1 - \alpha,$$

Or, $$P\left\{\chi^2_{1-\frac{\alpha}{2};n-1} \leq \frac{ns^2}{\sigma^2} \leq \chi^2_{\frac{\alpha}{2};n-1}\right\} = 1 - \alpha,$$

On writing the value of χ^2_{n-1} and replacing σ_0^2 by σ^2.

Further simplification gives

$$P\left\{\frac{1}{\chi^2} \le \frac{\sigma^2}{ns^2} \le \frac{1}{\chi^2_{1-\frac{\alpha}{2};n-1}}\right\} = 1-\alpha,$$

Or, $$P\left\{\frac{ns^2}{\chi^2_{\frac{\alpha}{2};n-1}} \le \sigma^2 \le \frac{ns^2}{\chi^2_{1-\frac{\alpha}{2};n-1}}\right\} = 1-\alpha$$

so that $\left(\frac{ns^2}{\chi^2_{\frac{\alpha}{2};n-1}}, \frac{ns^2}{\chi^2_{1-\frac{\alpha}{2};n-1}}\right)$ is the

$100(1-\alpha)\%$ confidence interval for the variance σ^2.

Test and confidence interval for proportion :

Let f denote the number of success in n Bernoullian trials. Then the sample proportion of successes is

$$\hat{P} = \frac{f}{n}$$

To test $H_0 : p = p_0$ against $H_1 : p \neq p_0$, where p is the unknown proportion of successes, we consider the test statistic

$$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_o(1-p_0)}{n}}}$$

Which, under H_0, is distributed as $N(0,1)$, for large sample. Thus, for a level of significance α, we reject the null hypothesis if

$$|Z| \ge Z_{\frac{\alpha}{2}}$$

5.4 Large Sample Theory of Testings

5.4.1 Large Sample Theory

So far we have considered testing of hypothesis which are exact and true irrespective of the sample size. But we have situation in practice where the exact testing procedures will be difficult to apply as the underlying distributions may be far from being normal. Howerve, if the sample size is large, the test procedures can be simplified by virtue of what is known as central limit the reom (CLT). It states that, under some general conditions on the distribuion of a variable X, a standardized variable of the form (X-mean)/standard derivation follows a standard normal distribution, provided the sample size is large. In other words, the assumption of normality of the population can be dropped if the sample size is large. One can then apply the standard results of normal theory for estimation and testing purposes.

The critical regions of tests depend on the nature of the alternative hypothesis, as shown in the following figure (Fig. 5.1). All the three tests have the same level of significance, namely, α.

(i) $H_1 : \theta > \theta_0$

(ii) $H_1 : \theta < \theta_0$

(iii) $H_1 : \theta \neq \theta_0$

Critical regions in the three cases are given, respectively, by (i) $z > z_\alpha$, (ii) $z < -z_a$, and (iii) $z < z_{\alpha/2}$ or $z > z_{\alpha/2}$.

We summarize below the different steps in calculating confidence interval of a parameter θ with confidence co-efficient $1-\alpha$:

Step 1. Estimate θ by $\hat{\theta}$, say

Step 2. Argue that $z = \dfrac{\hat{\theta}-\theta}{s.e.\left(\hat{\theta}\right)}$ follows N(0,1) for large n. Estimate s.e. $\left(\hat{\theta}\right)$ if necessary.

Step 3. Write $p\left\{-z_{\alpha/2} \leq \dfrac{\hat{\theta}-\theta}{s.e.\left(\hat{\theta}\right)} \leq z_{\alpha/2}\right\} = 1-\alpha$.

Finally,

Step 4. The $100(1-\alpha)\%$ confidence interval is

$$\left(\hat{\theta}-z_{\alpha/2}s.e.\left(\hat{\theta}\right),\hat{\theta}+z_{\alpha/2}s.e.\left(\hat{\theta}\right)\right).$$

$\aleph^2$ – test of Goodness of fitness :

Suppose we have a hypothetical population which is completely specified, there being no unknown parameter in its distribution. Even if some of the parameters are unknown, these are estimated from the given data. Let us visualize the population as being composed of *k* mutually exclusive classes, and let us suppose that, according to the hypothesis, the population proportion in the *i*th class is $P_i^0, i=1,2,....,k$. If the frequency in the *i*th class in a random sample of size *n* from this distribution is f_1, then under the hypothesis, the quantity

$$\sum_{i=1}^{k}\frac{\left(f_i-nP_i^0\right)^2}{n\chi_i^0}=\sum_{i=1}^{k}\frac{f_i^2}{nP_i^0}-n,$$

is approximately a χ^2 with (k–1) degrees of freedom (d.f.), provided nP_i^0 is large enough for each i.

If α is the chosen level of significance, then our test procedure consists in the rejection of the hypothesis of agreement between the observed distribution and the hypothetical distribution if

$\sum_{i=1}^{k}\frac{f_i^2}{nP_i^0}-n$ exceeds $\chi^2_{\alpha;}(k-1)$, and accepted otherwise.

This test is called a test for goodness of fit since it tests the closenesrs between the expected frequencies nP_i^0 and observed frequencies $f_i, i=1,2,....,k$. This is the celebrated Pearsonian χ^2 –test for goodness of fit.

5.5 Tests based on x², t and f-distributions

Tests based on χ^2, *t* and F–distributions have already been considered while discussing theoretical and sampling distributions in chapter 2. Here, the main consideration is that the parent distribution is normal.

χ^2– **test :**

Two uses of χ^2 – distribution are already mentioned in unit 2 (confidence interval for variance) and unit 3 (Chi-square test for goodness of fit).

Example 1. A manufacture claims that the lifetime of a certain brand of batteries by his factory has a variance 5000 (hours)^2. A sample of size 26 has a variance of 7200 (hours)^2 Assuming normality, let us test $H_0 : \sigma^2 = 5000$ against $H_1 : \sigma^2 \neq 5000$ at level $\alpha = 0{\cdot}02$.

$$\text{Here, } \frac{\sigma_0^2}{n-1}\chi^2_{1-\frac{\alpha}{2};n-1} = \frac{5000}{25} \times 11{\cdot}524$$

$$= 2304{\cdot}8$$

$$\text{and } \frac{\sigma_0^2}{n-1}\chi^2_{\frac{\alpha}{2};n-1} = \frac{5000}{25} \times 44{\cdot}314$$

$$= 8862{\cdot}8$$

Here, χ^2 – values for $\alpha = 0{\cdot}02$ are obtained from χ^2 – table. Also, s^2 is the sample variance with divisor (n–1). Since

s^2 is neither $\leq 2304{\cdot}8$, nor $\geq 8862{\cdot}8$,

we accept H_0 at $\alpha = 0{\cdot}02$ level of significance.

t-test for testing population mean, the population variance being unknown

Example 2.

In a pollution study of a river in India, the concentration of lead in the upper sedimentary layer of the bottom is measured from 15 sediment samples of thousand cubic centimetres each. The sample mean is 0·38 and the unknown population standard deviation is estimated from the sample as s = 0·12 (divisor n–1)

Then, under $H_0 : \mu_0 = 0{\cdot}48$ (against $H_1 : \mu_0 \neq 0{\cdot}48$), we have that $t = \frac{\sqrt{n}\left(\bar{X} - \mu_0\right)}{s}$ follows t-distribution

with (n–1) defrees of freedom.

Here, $t = \frac{\sqrt{15}\left(0{\cdot}38 - 0{\cdot}48\right)}{0{\cdot}12}$

$= -3{\cdot}2275$

For $\alpha = 0{\cdot}05$, we have $t_{0{\cdot}025;14} = 2{\cdot}145$.

Since $|t| = 3{\cdot}2275 > 2{\cdot}145$, we reject H_0 at 5% level of significance, and conclude that the mean concentration of lead is different from 0·48.

Other examples of t-test are already considered in unit 3 of chapter 2.

F-test for testing homogeneity of variances of two populations :

The F-distribution and its uses are discussed in unit 3 of chapter 2. Here we discuss one numerical example in this regard.

Example 3. Consider two samples from two populations of size $n_1 = 9$ and $n_2 = 10$ whose sample sums of squares of deviations from the means are 36 and 42, respectively. To test the hypothesis that the two populations have identical variance.

The unbiased estimates of the variances of the two populations are $\frac{36}{8}$ and $\frac{42}{9}$.

The observed F-statistic is

$F = \frac{42/9}{36/8}$ with 9 and 8 degrees of freedom,

$= 1{\cdot}04,$

while $F_{9,8;0.05} = 3{\cdot}39$.

Since F<3·39, it is not significant and we conclude that the samples may be drawn from the same population, *i.e.*, we accept the hypothesis $H_0 : \sigma_1^0 = \sigma_2^2$ against $H_1 : \sigma_1^0 \neq \sigma_2^2$.

The unbiased estimates of the variances of the two populations are $\frac{36}{8}$ and $\frac{42}{9}$.

The observed F-statistic is

$$F = \frac{42/9}{36/8} \text{ with 9 and 8 degrees of freedom,}$$

$$= 1{\cdot}04,$$

while $F_{9,8;0.05} = 3{\cdot}39$.

Since F<3·39, it is not significant and we conclude that the samples may be drawn from the same population, *i.e.*, we accept the hypothesis $H_0 : \sigma_1^0 = \sigma_2^2$ against $H_1 : \sigma_1^0 \neq \sigma_2^2$.

5.6 Summary

In this chapter, we have studied the theory of testing of statistical hypothesis. After studying this chapter, you would know null hypothesis, alternative hypothesis, Type 1 and Type 2 error, level of significance and power associated with a testing procedure. You would also know UMP test.

The ideas of confidence interval, confidence co-efficient, the difference between point and interval estimation have been explained. These are elaborated with numerical examples.

Finally, large sample testing procedure is explained.

5.7 Exercises

1. Define Type 1 and Type 2 error.
2. Distinguish between level of significance and power of a test.
3. What are acceptance and rejection region?
4. How does point and interval estimation differ?

5. Derive testing procedures for the following cases :

(a) $H_0 : \mu = \mu_0$, against $H_1 : \mu \neq \mu_0$ for the normal case $N\left(\mu, \sigma^2\right)$, σ^2 unknown.

(b) $H_0 : \mu = \mu_0$ against $H_1 : \mu \neq \mu_0$ for the normal case $N\left(\mu, \sigma^2\right), \neq \sigma^2$ known.

6. Define confidence interval for σ^2 in the normal case $N\left(\mu, \sigma^2\right)$, where μ is not known.

7. Write a note on Goodness of Fit test.

Unit 6 □ Correlation and Regression

Structure

6.0 Objectives

The followings are discussed here:

- Scatter diagram
- Correlation Co-efficient
- Least Square Method (LSM)
- Regression Equations and their uses
- Partial and Multiple correlation co-efficient

6.1 Introduction

Most of our discussions so far confined to a single variable. In statistical work we often have to deal with problems involving more than one variable. Let us consider, for the moment, two variables. Our interest lies in studying the relationship between the two variables.

For example, we may be interested in finding the relationship, if any, between

1. the performance of students in two subjects, say, mathematics and statistics or accountancy and economics,

2. the heights of father and eldest son,

3. the index of wholesale prices and that of agricultural production,

4. the ages of bride-groom and bride.

A rough idea about the association between two variables can be obtained by studying what is called a 'scatter diagram'. Here, two variables are plotted along two perpendicular axes, namely, x-axis and y-axis. After plotting the data in a graph, one obtains a clear picture of the variation in one variable with respect to the other. Scatter diagram is explained below in detail.

6.2 Scatter Diagram

Given a set of pairs of observations on two variables, x and y, a rough picture about their interrelationship can be obtained by drawing a *scatter diagram*. Such a diagram can be obtained by taking, say, x along the horizontal axis and y along the vertical and then plotting each pair of values (x_i,y_i) as a point with respect to these axes. Thus the diagram will be a conglomeration of points (x_i,y_i). The pattern of this conglomeration will vary depending on the nature of the relationship between the variables. For example, consider the following three scatter diagrams (fig.6.1, parts *a–b*).

In fig. 6.1(a) the general tendency of the points (x, y) is such that as one variable increases the other also increases. In fig. 6.1(b) the general tendency of the points (x, y) is such that as one variable increases the other decreases. But in fig. 6.1(c), x

and y are uncorrelated in the sense that with an increase or decrease in one variable, the other generally remains unchanged. The three situations are said to be, respectively, the case of *positive correlation*, that of *negative correlation* and that of *zero correlation*.

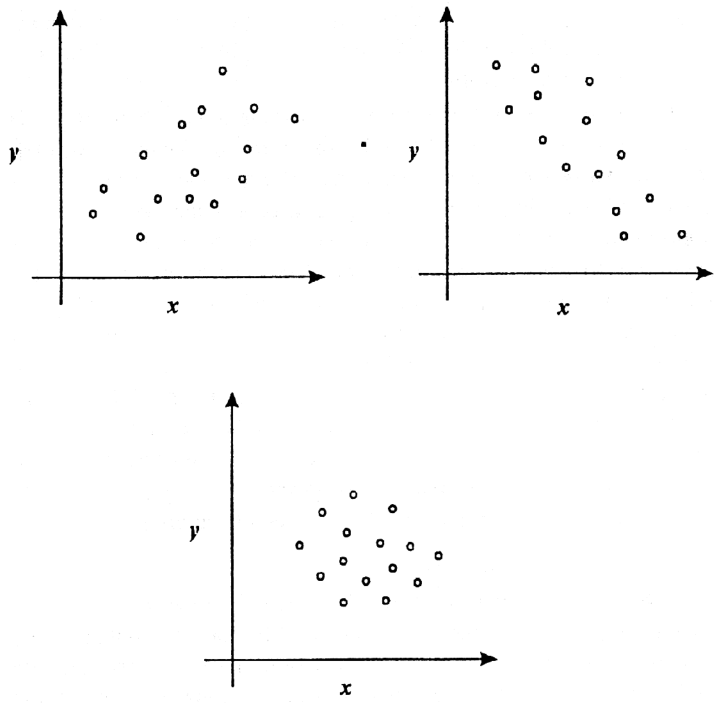


Fig. 6.1 : (Clock-wise from upper left-hand corner) Scatter diagrams for (a) positive, (b) negative and (c) near- zero correlation.

6.3 Correlation Co-efficient and its properties

Assuming that the relationship between the variables x and y is linear, or approximately so, as may be revealed by the scatter diagram, we can give a measure of the strength of this relationship. This numerical measure is called the correlation co-efficient.

Suppose $(x_i,y_1),(x_2,y_2),\ldots,(x_n,y_n)$ be n pairs of observations on (x,y). Then the correlation co-efficient between x and y, written as r_{xy}, is defined as

$$r_{xy} = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{s_x s_y}, \ldots\ldots(1)$$

where the numerator is the covariance between x and y, and the denominators are the standard deviations of x and y. Note that $r_{xy} = r_{yx}$, so that the expression of correlation co-efficient is symmetric in x and y. So, we write the correlation co-efficient as 'r' only ignoring the suffix.

An expression of r, useful for calculation, is given as

$$r = \frac{\sum_{i=1}^{n} x_i y_i - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)/n}{\sqrt{\left\{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}\right\}\left\{\sum_{i=1}^{n} y_i^2 - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n}\right\}}} \ldots\ldots(2)$$

If in a given problem, r comes out to be positive, our conclusion will be that y generally increases as x increases. In case r comes out to be negative,we would conclude that y generally decreases when x increases. If r is equal to zero, the conclusion is that as one increases (or decreases), the other remains unchanged.

The following properties of the correlation co-efficient are worth noting :

1. If all the points in a scatter diagram lie on a straight line with positive scope, then $r = + 1$. Similarly, if all the points in a scatter diagram lie on a straight line with negative slope, then $r =–1$. More specifically, if x and y satisfy $ax + by + c = 0$., 'a', 'b' and 'c' being constants, then the value of r either 1 or – 1, depending on the slope of the line.

2. Let us consider n pairs $(x_i, y_i), i = 1,2,...,n,$ of observations. Define

$u=\dfrac{x-a}{b}, v=\dfrac{y-c}{d}$, $(b \neq 0,\ d \neq 0)$. Then

$$r_{xy}=\frac{bd}{|b||d|}r_{uv}$$

Thus, if 'b' and 'd' are of the same sign, then $r_{xy}=r_{uv}$, as in this case $bd=|b||d|$. On the other hand, if 'b' and 'd' are of the opposite sign, then $bd = -|b|\ |d|$, and $r_{xy}=-r_{uv}$.

3. The numerical value of r (*i.e.*, $|\ r\ |$) measures the strength of the linear relationship and the sign of r indicates the direction of the relationship, *i.e.*, increasing or decreasing.

4. The correlation co-efficient r satisfies $|r| \leq 1$.

Proof: Let $(x_1,y_1),(x_2,y_2),\ldots,(x_n,y_n)$ be n pairs of observations. Define u and v as

$$u_i=\frac{x_i-\bar{x}}{s_x}, v_i=\frac{y_i-\bar{y}}{s_y},$$

$(i=1,2,3,\ldots,n)$

Then $\sum\limits_{i=1}^{n}(u_i+v_i)^2 \geq o$, ($u_i$ and v_i being real numbers)

or, $\sum\limits_{i=1}^{n}u_i^2+\sum\limits_{i=1}^{n}v_i^2+2\sum\limits_{i=1}^{n}u_iv_i \geq 0$,

or, $\dfrac{\sum\limits_{i=1}^{n}(x_i-\bar{x})^2}{s_x^2}+\dfrac{\sum\limits_{i=1}^{n}(y_i-\bar{y})^2}{s_y^2}+2\dfrac{\sum\limits_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{s_xs_y} \geq 0$

or, $n+n+2n.r_{xy} \geq 0$

or, $r_{xy} \geq -1\ldots\ldots(1)$

Again, considering $\sum_{i=1}^{n}(u_i - v_i)^2 \geq 0$,

we will get $n + n - 2n.r_{xy} \geq 0$

or, $-2nr_{xy} \geq -2n$

or, $r_{xy} \leq 1......(2)$

Combining (1) and (2), we have

$-1 \leq r_{xy} \leq 1$,

or, $|r_{xy}| \leq 1$.

6.4 Mathematical relationship between random variable, Regression Equation

When two variables are found to be associated, as revealed by the scatter diagram, one can think of a mathematical relationship between these variables. This may take the form of a straight line, quadratic or cubic curve, or some other complicated mathematical model. However, we will keep them to be of the simplest type, say, a polynomial of certain degree.

Another important aspect of fitting a line or curve from data is the choice of independent or dependent variables. Unlike mathematics, the selection of independent (or dependent) variables in statistics is not that clear cut. For example, in studying demand and price of a commodity, we are mostly inclined to take price as an independent variable, and demand as a dependent variable. In doing so, our logic may be that price determines demand of a commodity. But it may so happen, in some situations, that demand for a commodity will manipulate its price.

The case of regression involving two variables can be extended to the case of $n(\geq 3)$ variables. However, one needs to consider only the variables which are really meaningful and effective in explaining the data.

6.4.1 Curve fitting by the method of least squares

Let us consider *n* sets of observations, namely, $(Y, X_{1i}, X_{2i}, ... X_{ki}), (i = 1, 2, ..., n)$, involving k + 1 variables, where Y is the dependent variable and $X_1, X_2, ..., X_k$ are the independent variables.

As a mathematical relationship, we consider the regression plane

$$Y = a_0 + a_1 X_{1i} + a_2 X_{2i} + ... + a_k X_{ki,} \left(i = 1, 2, ..., n\right)$$

The constants $a_0, a_1, ..., a_k$ are estimated by the method of least squares, namely, by minimizing

$$S^2 = \sum_{i=1}^{n} \left(Y_i - \overline{Y}_i\right)^2$$, the error sum of squares,

where $\hat{Y}_i = \hat{a}_0 + \hat{a}_1 X_{1i} + ... + \hat{a}_k X_{ki}$, is the estimate.

On solving $O = \frac{\partial s^2}{\partial \hat{a}_j}, \left(i = 1, 2, ... n : j = 1, 2, ..., k\right)$

we have $O = \sum_{i=1}^{n} 2\left(Y - \hat{a}_0 - \hat{a}_1 X_{1i} - .... - \hat{a}_j X_{ji} - ... \hat{a}_k X_{ki}\right).\left(-X_{ji}\right)$

or, $O = \sum_{i=1}^{n} \left(X_{ji}\right)\left(Y - \hat{a}_0 - \hat{a}_1 X_{1i} - .... - \hat{a}_j X_{ji} - ... \hat{a}_k X_{ki}\right)$,

(for j = 1,2,3...,k), giving k equations ...(1)

Another equation is obtained from

$$\frac{\partial s^2}{\partial \hat{a}_0} = 0$$

which gives $\sum_{i=1}^{n} Y_i = n\hat{a}_0 + \hat{a}_1 \sum_{i=1}^{n} X_{1i} + ... + \hat{a}_k \sum_{i=1}^{n} X_{ki}$ ...(2)

The (k+1) equations in (1) and (2) determine the values of $\hat{a}_0, \hat{a}_1, ..., \hat{a}_k$ which are called the least squares estimates (LSE) of $a_0, a_1 ..., a_k$.

6.5 Regression equation—the case of variables

Given n pairs of observations of the form (x_i, y_i), we may be interested in deriving a functional relationship between x and y. The simplest type of relationship between x and y is a straight line given by

y = a + bx, where 'a' and 'b' are constants which are also called parameters.

These parameters are estimated from the given data by applying the method of least squares.

Thus, we minimize the error sum of squares

$s = \sum_{i=1}^{n} (y_i - a - bx_i)^2$ with respect to 'a' and 'b'.

So, by solving

$$\frac{\partial s}{\partial a} = 0 \text{ and } \frac{\partial s}{\partial b} = 0$$

we get $\sum_{i=1}^{n} y_i = na + b_i \sum_{i=1}^{n} x_i$

and $\sum_{i=1}^{n} x_i y_i = a \sum_{i=1}^{n} x_i + b \sum_{i=1}^{n} x_1^2$

These are called normal equations.

On solving, we get

$$b = \frac{n\sum_{i=1}^{n} x_i y_i - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{n\sum_{i=1}^{n} x_1^2 - \left(\sum_{i=1}^{n} x_i\right)^2}$$

$$= \frac{\operatorname{cov}(x,y)}{s_x^2}$$

$$= r\frac{s_y}{s_x},$$

where r is the correlation co-efficient between x and y, and s_x and s_y are the s.d's of x and y respectively.

Also, $a = \bar{y} - b\bar{x}, \bar{x}$ and $\bar{y}$ being the sample means.

Then the fitted regression line is

$$y = (\bar{y} - b\bar{x}) + bx$$

$$= \bar{y} + b(x - \bar{x}).$$

The constant b, sometimes written as by_x is called the regression co-efficient of y on x.

We can also fit a line of the form

$$x = a' + b'y = \bar{x} + b'(y - \bar{y}),$$

Which will be called regression equation of x on y, and b' will be denoted by b_{xy}.

Note that $b_{xy} = r\frac{s_y}{s_y}$, so that

$$b_{xy} \times b_{yx} = r\frac{s_x}{s_y} \times r\frac{s_y}{s_x}$$

or, $r = \pm\sqrt{b_{xy}.b_{yx}}$

Thus, r is the geometric mean of the two regression co-efficients. It takes the sign of either b_{xy} or b_{yn}, as b_{xy} and b_{yx} are of the same sign.

It can be shown that r^2 measures the proportion of total variability in the data that is explained by the regression equation. For instance, if r = 0.8, then 64% of the data variability is explained by the regression line.

Another important property of the regression lines is that they intersect at the point $(\bar{x}, \bar{y})$.

6.5.1 Regression equation considering three variables

The case of regression equation involving one explanatory variable can be extended to the situation where there are 'k' explanatory variables. If the explanatory variables are selected suitably, the predicting capability of the regression equation increases considerably. The simplest case of regression equation with, say, k = 2 explanatory variables is one where the equation is linear. Such a line is called a 'regression plane'. Thus, with three variables x_1, x_2 and x_3

we have

$$x_1 = b_0 + b_{12.3}x_2 + b_{13.2}x_3,$$

where $b_0, b_{12.3}$ and $b_{13.2}$ are estimated by the method of least squares. We discuss one numerical example as shown below.

Example 1.

The following table gives the number of children (x_1), educational level of the mother (x_2) and total monthly expenditure (x_3) of 15 nuclear families. Here only those families are considered for which the mothers have crossed the child-bearing age, and their educational level is measured in terms of number of years spent in school, college and university.

Table : 6.1. Number of children, educational level of the mother and total monthly expenditure of 15 families :

Sl. No. of family	Number of children (x_1)	Educational level of mother (x_2)	Total monthly expenditure in Rs. (x_3)
1	2	14	4350
2	1	17	5450
3	1	15	4500
4	6	12	3750
5	2	12	4220
6	7	7	3150
7	1	17	6750
8	3	10	3600
9	3	12	3225
10	4	10	4875
11	5	9	2250
12	4	8	2700
13	6	11	4050
14	0	16	7225
15	5	8	1800

Our aim is to fit a regression plane of x_1 on x_2 and x_3, and examine the effects of mother's educational level and farnily's standard of living (as given by the total monthly expenditure) on the number of clildren born to the family.

Let the regression plane of x_1 on x_2 and x_3 be given by

$$x_1 = b_0 + b_{12.3}\, x_2 + b_{13.2} x_3, .$$

The least-square estimates of the parameters $b_0, b_{12.3}$ and $b_{13.2}$ can be shown to be

$$b_{12.3} = \frac{s_1}{s_2} \times \frac{r_{12} - r_{13} r_{23}}{1 - r_{23}}$$

$$b_{13.2} = \frac{s_1}{s_3} \times \frac{r_{13} - r_{12}r_{23}}{1 - r_{23}}$$

and $x_1 = b_0 + b_{12.3}\, x_2 + b_{13.2} x_3,$

where s_1, s_2, s_3 are the standard deviations of x_1, x_2 and x_3 respectively.

For the data in table 6.1, we have the following :

$\bar{x}_1 = 3.33, \quad s_1 = 2.0869,$

$\bar{x}_2 = 11.87, \; s_2 = 3.2014,$

$\bar{x}_3 = 4126.33, \quad s_3 = 1464.09,$

and $r_{12} = -0.8217, \; r_{13} = -0.7068$ and $r_{23} = 0.8342$.

Finally, $b_{12.3} = -0.4974, \; b_{13.2} = -0.0001$ and $b_0 = 9.6468$.

So, the fitted regression plane of x_1 on x_2 and x_3 is given by

$x_1 = 9.6468 - 0.4974\, x_2 - 0.0001\, x_3$.

The multiple correlation co-efficient can be found to be $r_{1.23} = 0.8226$.

Thus, $r^2_{12.3}$ being 0.6767, the fitted regression plane explains 67.67 per cent of the total variability in the observed values of x_1. Our conclusion, therefore, is that the performance of mother's educational level (x_2) and family's total monthly expenditure (x_3) as explanatory variables is satisfactory, though there may be other explanatory variables which we have not considered.

As an application of the above fitted regression plane, we can predict the number of children x_1 in a family with, say, x_2 = 15 and x_3 = 5000, as follows :

$$\begin{aligned} x_1 &= 9.6468 - 0.4974 \times 15 - 0.0001 \times 5000 \\ &= 1.6858 \\ &= 2 \end{aligned}$$

(**Source :** Statistics for the Social Sciences by Gun, A.M and Aich, A.B. (2002), World Press, Kolkata-700 073, p. 53-54.)

6.6 Partial and Multiple Correlation

Let us consider the case of three variables x_1, x_2 and x_3.Then the regression line, which is actually a plane, called regression plane, of x_1 on x_2 and x_3 is given by $x_1 = b_0 + b_{12}x_2 + b_{13}x_3$, when b_{12}, b_{13} are co-efficients, called partial regression co-efficients and written by $b_{12.3}$ and $b_{13.2}$ respectively. Hers, $b_{12.3}$ measures the change in $x_1$for unit change in x_2, x_3 being fixed. Similarly, $b_{13.2}$ measures the change in x_1 for unit change in x_3, x_2 being fixed.

Multiple correlation co-efficient—

The multiple correlation co-efficient with three or more variables is a generalization of the simple correlation co-efficient involving two variables only. The multiple correlation co-efficient is defined to be the usual correlation co-efficient between the observed x_1, and its estimate obtained from the regression equation (here, regression plane). In case of three variables x_1, x_2 and x_3, this correlation is denoted by $r_{1.23}$.

Mathematically, let x_1 and $\hat{x}_1$ be respectively the observed x_1 and its estimate obtained as

$$\hat{x}_1 = b_0 + b_2x_2 + b_3x_3$$

Then $r_{1.23} = \text{corr}.(x_1, \hat{x}_1)$

$$= \frac{\text{cov}(x_1, \hat{x}_1)}{s.d.(x_1).s.d.(\hat{x}_1)}$$

On simplification, we have

$$r_{1.23} = \sqrt{\frac{r_{12}^2 + r_{13}^2 - 2r_{12}r_{13}r_{23}}{1 - r_{23}^2}},$$

where r_{12} is the simple correlation co-efficient between $x_1$and x_2, r_{13} that between x_1 and x_3, and, finally, r_{23} between x_2 and x_3.

Contrary to the case of a simple correlation co-efficient, a multiple correlation co-efficient is never negative, and lies in between 0 and 1, *i.e.*, $0 \le r_{1.23} \le 1$.

An important interpretation of the multiple correlation co-efficient is that it measures the degree of association between the observed x_1 and its estimate $\hat{x}_1$ obtained from the regression plane.

Thus, larger value of $r_{1.23}$ indicates that the regression plane is adequate as a predicting formula for x_1.

Partial correlation co-efficient :

Let there be three variables x_1, x_2 and x_3. Here, we are interested in knowing the degree of association between, say, x_1 and x_2, when the influence of x_3 is eleminated from both. The resulting correlation co-efficient, called partial correlation co-efficient, will be denoted by $r_{12.3}$. The formula for $r_{12.3}$ is given by

$$r_{12.3} = \frac{r_{12} - r_{13}r_{23}}{\sqrt{\left(1-r_{13}^2\right)\left(1-r_{23}^2\right)}},$$

where r_{12}, r_{13} and r_{23} are the total correlation co-efficients.

6.7 Summary

In this chapter, we have studied the important concept of association and regression. The idea of correlation co-efficient is discussed in detail, along with its various properties. The scatter diagram, as discussed here, though simple in nature, plays a vital role in identifying the nature and extent of association between variables.The idea of LSE is also given.

The regression equation is stated and the parameters are estimated by least squares method (LSE). The significance of regression co-efficients is elaborated. The relationship between regression co-effiicients and correlation co-efficient is established.

The case of three variables is also taken up. A numerical example in this context is also given. Finally, partial and multiple correlation co-efficient are discussed.

6.8 Exercises

1. What do you mean by 'scatter diagram'? What are their uses?
2. Define correlation co-efficient. Write its three important properties.
3. Prove that $|r| \leq 1$. When does the equality hold?
4. Interpret the cases (a) $r = 1$, (b) $r = -1$, (c) $r = 0$.
5. What are regression lines? How do you estimate the regression co-efficients?
6. What is multiple correlation co-efficient? What about its range of variation?
7. What are partial regression co-efficients? How would you interpret them? Can $b_{12.3}$ and b_{12} be equal?
8. What are partial correlation co-efficient? Can $r_{12.3}$ and r_{12} be equal?

6.9 Further Readings

1. Goon, A. M. and Aicj, A.B.—Statistics for the Social Sciences, World Press, Kolkata, 2002.

2. Goon, A.M. Gupta, M.K. and Dasgupta, B—Fundamental of Statistics (Vol-1), World Press, Kolkata, 2000.

3. Goon, A.M. Gupta, M.K and Dasgupta, B—An Outline of Statistical Theory (Vol.-1), World Press, Kolkata, 2010.

4. Hogg., R.V and Craig, A.—Introducation to Mathematical Statistics, Macmillan Publishing Co. Delhi, 1978.

5. Rohatgi, V.K. and Saleh, A.K. —An Introduction to Probability and Statistics, John Wiley, 2005.

6. Casella, George and Berger, Roger—Statistical Inference, Delhi, 2002.

7. Bhat, B.R.—Modern Probability Theory, New Delhi, 2000.

8. Weiss, Neil A.—Introductory Statistics, Pearson, Education, 2007.

9. Chatterjee, S. and Simonoff J.S.— Handbook of Regression Analysis, Wiley, NY.

10. Cacoullos, T.—Exercises in Probability, Narosa, New Delhi.

APPENDIX

Tables for Normal, χ^2, t and F-distribution. These tables are used to find critical values in respect of a given test, and to find confidence interval for given level of significance.

TABLE I : STANDARD NORMAL DISTRIBUTION

$P[z \leq k]$ for specified values of k

k	0.0 +	0.5 +	1.0 +	1.5 +	2.0 +	2.5 +	3.0 +	3.5 +
0.00	5000	6915	8413	9332	9772	$9^{2}379$	$9^{2}865$	$9^{3}77$
0.01	5040	6950	8438	9345	9778	$9^{2}396$	$9^{2}869$	$9^{3}78$
0.02	5080	6985	8461	9357	9783	$9^{2}413$	$9^{2}874$	$9^{3}78$
0.03	5120	7019	8485	9370	9788	$9^{2}430$	$9^{2}878$	$9^{3}79$
0.04	5160	7054	8508	9382	9793	$9^{2}446$	$9^{2}882$	$9^{3}80$
0.05	5199	7088	8531	9394	9798	$9^{2}461$	$9^{2}886$	$9^{3}81$
0.06	5239	7123	8554	9406	9803	$9^{2}477$	$9^{2}889$	$9^{3}81$
0.07	5279	7157	8577	9418	9808	$9^{2}492$	$9^{2}893$	$9^{3}82$
0.08	5319	7190	8599	9429	9812	$9^{2}506$	$9^{2}897$	$9^{3}83$
0.09	5359	7224	8621	9441	9817	$9^{2}520$	$9^{2}900$	$9^{3}83$
0.10	5398	7257	8643	9452	9821	$9^{2}534$	$9^{3}03$	$9^{3}84$
0.11	5438	7291	8665	9463	9826	$9^{2}547$	$9^{3}06$	$9^{3}85$
0.12	5478	7324	8686	9474	9830	$9^{2}560$	$9^{3}10$	$9^{3}85$
0.13	5517	7357	8708	9484	9834	$9^{2}573$	$9^{3}13$	$9^{3}86$
0.14	5557	7389	8729	9495	9838	$9^{2}585$	$9^{3}16$	$9^{3}86$
0.15	5596	7422	8749	9505	9842	$9^{2}598$	$9^{3}18$	$9^{3}87$
0.16	5636	7454	8770	9515	9846	$9^{2}609$	$9^{3}21$	$9^{3}87$
0.17	5675	7486	8790	9525	9850	$9^{2}621$	$9^{3}24$	$9^{3}88$
0.18	5714	7517	8810	9535	9854	$9^{2}632$	$9^{3}26$	$9^{3}88$
0.19	5753	7549	8830	9545	9857	$9^{2}643$	$9^{3}29$	$9^{3}89$
0.20	5793	7580	8849	9554	9861	$9^{2}653$	$9^{3}31$	$9^{3}89$
0.21	5832	7611	8869	9564	9864	$9^{2}664$	$9^{3}34$	$9^{3}90$
0.22	5871	7642	8888	9573	9868	$9^{2}674$	$9^{3}36$	$9^{3}90$
0.23	5910	7673	8907	9582	9871	$9^{2}683$	$9^{3}38$	$9^{4}04$
0.24	5948	7704	8925	9591	9875	$9^{2}693$	$9^{3}40$	$9^{4}08$

0.25	5987	7738	8944	9599	9878	$9^{2}702$	$9^{3}42$	$9^{4}12$
0.26	6026	7764	8962	9608	9881	$9^{2}711$	$9^{3}44$	$9^{4}15$
0.27	6064	7794	8980	9616	9884	$9^{2}720$	$9^{3}46$	$9^{4}18$
0.28	6103	7823	8997	9625	9887	$9^{2}728$	$9^{3}48$	$9^{4}22$
0.29	6141	7852	9015	9633	9890	$9^{2}736$	$9^{3}50$	$9^{4}25$
0.30	6179	7881	9032	9641	9893	$9^{2}744$	$9^{3}52$	$9^{4}28$
0.31	6217	7910	9049	9649	9896	$9^{2}752$	$9^{3}53$	$9^{4}31$
0.32	6255	7939	9066	9656	9898	$9^{2}760$	$9^{3}55$	$9^{4}33$
0.33	6293	7967	9082	9664	9901	$9^{2}767$	$9^{3}57$	$9^{4}36$
0.34	6331	7995	9099	9671	9904	$9^{2}774$	$9^{3}58$	$9^{4}39$
0.35	6368	8023	9115	9678	9906	$9^{2}781$	$9^{3}60$	$9^{4}41$
0.36	6406	8051	9131	9686	9909	$9^{2}788$	$9^{3}61$	$9^{4}43$
0.37	6443	8078	9147	9693	9911	$9^{2}795$	$9^{3}62$	$9^{4}46$
0.38	6480	8106	9162	9699	9913	$9^{2}801$	$9^{3}64$	$9^{4}48$
0.39	6517	8133	9177	9706	9916	$9^{2}807$	$9^{3}65$	$9^{4}50$
0.40	6554	8159	9192	9713	9918	$9^{2}813$	$9^{3}66$	$9^{4}52$
0.41	6591	8186	9207	9719	9920	$9^{2}819$	$9^{3}68$	$9^{4}54$
0.42	6628	8212	9222	9726	9922	$9^{2}825$	$9^{3}69$	$9^{4}56$
0.43	6664	8238	9236	9732	9925	$9^{2}831$	$9^{3}70$	$9^{4}58$
0.44	6700	8264	9251	9738	9927	$9^{2}836$	$9^{3}71$	$9^{4}59$
0.45	6736	8289	9265	9744	9929	$9^{2}841$	$9^{3}72$	$9^{4}61$
0.46	6772	8315	9279	9750	9931	$9^{2}846$	$9^{3}73$	$9^{4}63$
0.47	6808	8340	9292	9756	9932	$9^{2}851$	$9^{3}74$	$9^{4}64$
0.48	6844	8365	9306	9761	9934	$9^{2}856$	$9^{3}75$	$9^{4}66$
0.49	6879	8389	9319	9767	9936	$9^{2}861$	$9^{3}76$	$9^{4}67$

Notes : (1) The specified value of k is obtainable by adding the number in the marginal column with the number in the marginal row.

(2) Decimal points in the body of the table are omitted, and repeated 9's are indicated by powers; e.g., $9^{2}379$ stands for 0.99379 and 6915 for 0.6915.

DISTRIBUTION OF STANDARD NORMAL VARIABLE

Values of z_α				
α	0.05	0.025	0.01	0.005
z_α	1.645	1.960	2.326	2.576

TABLE II : χ^2 DISTRIBUTION

Values of $\chi^2_{\alpha : \nu}$

α / ν	0.99	0.98	0.95	0.90	0.80	0.70	0.50	0.30	0.20	0.10	0.05	0.02	0.01
1	0.0^3157	0.0^3628	0.0^2393	0.0158	0.0642	0.148	0.455	1.074	1.642	2.706	3.841	5.412	6.635
2	0.0201	0.0404	0.103	0.211	0.446	0.713	1.386	2.408	3.219	4.605	5.991	7.824	9.210
3	0.115	0.185	0.352	0.584	1.005	1.424	2.366	3.665	4.642	6.251	7.815	9.837	11.345
4	0.297	0.429	0.711	1.064	1.649	2.195	3.357	5.878	5.989	7.779	9.488	11.668	13.277
5	0.554	0.752	1.145	1.610	2.343	3.000	4.351	6.064	7.289	9.236	11.070	13.388	15.086
6	0.872	1.134	1.635	2.204	3.070	3.828	5.348	7.231	8.558	10.645	12.592	15.033	16.812
7	1.239	1.564	2.167	2.833	3.822	4.671	6.346	8.383	9.803	12.017	14.067	16.622	18.475
8	1.646	2.032	2.733	3.490	4.594	5.527	7.344	9.524	11.030	13.362	15.507	18.168	20.090
9	2.088	2.532	3.325	4.168	5.380	6.393	8.343	10.656	12.242	14.684	16.919	19.679	21.666
10	2.358	3.059	3.940	4.865	6.179	7.267	9.342	11.781	13.442	15.987	18.307	21.161	23.209
11	3.053	3.609	4.575	5.578	6.989	8.148	10.341	12.899	14.631	17.275	19.675	22.618	24.725
12	3.571	4.178	5.226	6.304	7.807	9.034	11.340	14.011	15.821	18.549	21.026	24.054	26.217
13	4.107	4.765	5.892	7.042	8.634	9.926	12.340	15.119	16.985	19.812	22.362	25.472	27.688
14	4.660	5.368	6.571	7.790	9.467	10.821	13.339	16.222	18.151	21.064	23.685	26.873	29.141
15	5.229	5.985	7.261	8.547	10.307	11.721	14.339	17.322	19.311	22.307	24.996	28.259	30.578
16	5.812	6.614	7.962	9.312	11.152	12.624	15.338	18.418	20.465	23.542	26.296	29.633	32.000
17	6.408	7.255	8.672	10.085	12.002	13.531	16.338	19.511	21.615	24.769	27.587	30.995	33.409
18	7.015	7.906	9.390	10.865	12.857	14.440	17.338	20.601	22.760	25.989	28.869	32.346	34.805
19	7.633	8.567	10.117	11.651	13.716	15.352	18.338	21.689	23.900	27.204	30.144	33.687	36.191
20	8.260	9.237	10.851	12.443	14.578	16.266	19.337	22.775	25.038	28.412	31.410	35.020	37.566
21	8.897	9.915	11.591	13.240	15.445	17.182	20.337	23.858	26.171	29.615	32.671	36.343	38.932
22	9.542	10.600	12.338	14.041	16.314	18.101	21.337	24.939	27.301	30.813	33.924	37.659	40.289
23	10.196	11.293	13.091	14.848	17.487	19.021	22.337	26.018	28.429	32.007	35.172	38.968	41.638
24	10.856	11.992	13.348	15.659	18.062	19.943	23.337	27.096	29.553	33.196	36.415	40.270	42.980
25	11.524	12.697	14.611	16.473	18.940	20.867	24.337	28.172	30.675	34.382	37.652	41.566	44.314
26	12.198	13.409	15.379	17.292	19.820	21.792	25.336	29.246	31.795	35.563	38.885	42.856	45.642
27	12.879	14.125	16.151	18.114	20.703	22.719	26.336	30.319	32.912	36.741	40.113	44.140	46.963
28	13.565	14.847	16.928	18.939	21.588	23.647	27.336	31.391	34.027	37.916	41.337	45.419	48.278
29	14.256	15.574	17.708	19.768	22.475	24.577	28.336	32.461	35.439	39.087	42.557	46.693	49.588
30	14.953	16.306	18.493	20.599	23.364	25.508	29.336	33.530	36.250	40.256	43.773	47.962	50.892

Note : For values of v greater than 30, the quantity $\sqrt{2\chi^2} - \sqrt{2\nu - 1}$ may be taken to be distributed as approximately a standard normal variable.

TABLE III : t-DISTRIBUTION : $P[t \leq k]$ for different values of $k \geq 0$ and on differing degrees of freedom v

k	$v = 1$	2	3	4	5	6	7	8	9	10
0	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500
0.1	.532	.535	.537	.537	.538	.538	.538	.539	.539	.539
.2	.563	.570	.573	.574	.575	.576	.576	.577	.577	.577
.3	.593	.604	.608	.610	.612	.613	.614	.614	.614	.615
.4	.621	.636	.642	.645	.647	.648	.649	.650	.651	.651
.5	.648	.667	.674	.678	.681	.683	.684	.685	.685^3	.686
.6	.672	.695	.705	.710	.713	.715	.716	.717	.718	.719
.7	.694	.722	.733	.739	.742	.745	.747	.748	.749	.750
.8	.715	.746	.759	.766	.770	.773	.775	.777	.778	.779
.9	.733	.768	.783	.790	.795	.799	.801	.803	.804	.805
1.0	.750	.789	.804	.813	.818	.822	.825	.827	.828	.830
1.1	.765	.807	.824	.833	.839	.843	.846	.848	.850	.851
1.2	.779	.823	.842	.852	.858	.862	.865	.868	.870	.871
1.3	.791	.838	.858	.868	.875	.879	.883	.885	.887	.889
1.4	.803	.852	.872	.883	.890	.894	.898	.900	.902	.904
1.5	.813	.864	.885	.896	.903	.908	.911	.914	.916	.918
1.6	.822	.875	.896	.908	.915	.920	.923	.926	.928	.930
1.7	.831	.884	.906	.918	.925	.930	.933	.936	.938	.940
1.8	.839	.893	.915	.927	.934	.939	.943	.945	.947	.949
1.9	.846	.901	.923	.935	.942	.947	.950	.953	.955	.957
2.0	.852	.908	.930	.942	.949	.954	.957	.960	.962	.963
2.1	.858	.915	.937	.948	.955	.960	.963	.965	.967	.969
2.2	.864	.921	.942	.954	.960	.965	.968	.970	.972	.974
2.3	.869	.926	.947	.958	.965	.969	.972	.975	.976	.978
2.4	.874	.931	.952	.963	.969	.973	.976	.978	.980	.981
2.5	.879	.935	.956	.967	.973	.977	.979	.981	.983	.984
2.6	.883	.939	.960	.970	.976	.980	.982	.984	.986	.987
2.7	.887	.943	.963	.973	.979	.982	.985	.986	.988	.989
2.8	.891	.946	.966	.976	.981	.984	.987	.988	.990	.991
2.9	.894	.949	.969	.978	.983	.986	.988	.990	.991	.992
3.0	.898	.952	.971	.980	.985	.988	.990	.991	.992	.993

3.1	.901	.955	.973	.985	.987	.989	.991	.993	.994	.994
3.2	.904	.957	.975	.983	.988	.991	.992	.994	.995	.995
3.3	.906	.960	.997	.985	.989	.992	.993	.995	.995	.996
3.4	.909	.962	.979	.986	.990	.993	.994	.995	.996	.997
3.5	.911	.964	.980	.988	.991	.994	.995	.906	.997	.997
3.6	.914	.965	.982	.989	.992	.994	.996	.996	.997	.998
3.7	.916	.967	.983	.990	.993	.995	.996	.997	.997	.998
3.8	.918	.969	.984	.990	.994	.995	.997	.997	.998	.998
3.9	.920	.970	.985	.991	.994	.996	.997	.998	.998	.998
4.0	.922	.971	.986	.992	.995	.996	.997	.998	.998	.999
4.1	.924	.973	.987	.993	.995	.997	.998	.998	.999	.999
4.2	.926	.974	.988	.993	.996	.997	.998	.998	.999	.999
4.3	.927	.975	.988	.994	.996	.997	.998	.999	.999	.999
4.4	.929	.976	.989	.994	.996	.998	.998	.999	.999	.999
4.5	.930	.977	.990	.995	.997	.998	.999	.999	.999	.999
4.6	.932	.978	.990	.995	.997	.998	.999	.999	.999	.999
4.7	.933	.979	.991	.995	.997	.998	.999	.999	.999	1.000
4.8	.935	.980	.991	.996	.998	.998	.999	.999	.999	
4.9	.936	.980	.992	.996	.998	.999	.999	.999	1.000	
5.0	.937	.981	.992	.996	.998	.999	.999	.999		
5.1	.938	.982	.993	.996	.998	.999	.999	.999		
5.2	.939	.982	.993	.997	.998	.999	.999	1.000		
5.3	.941	.983	.993	.997	.998	.999	.999			
5.4	942	.984	.994	.997	.998	.999	.999			
5.5	.943	.984	.994	.997	.999	.999	.999			
5.6	.944	.985	.994	.997	.999	.999	1.000			
5.7	.945	.985	.995	.998	.999	.999				
5.8	.946	.986	.995	.998	.999	.999				
5.9	.947	.986	.995	.998	.999	.999				
6.0	0.947	0.987	0.995	0.998	0.999	0.999				

TABLE- IV : *t*–DISTRIBUTION

Values of $t_{\alpha:v}$

$^{\alpha}/_{v}$	0.05	0.025	0.01	0.005
1	6.314	12.706	31.821	63.657
2	2.920	4.303	6.965	9.925
3	2.353	3.182	4.541	5.841
4	2.132	2.776	3.747	4.604
5	2.015	2.571	3.365	4.032
6	1.943	2.447	3.143	3.707
7	1.895	2.365	2.998	3.499
8	1.860	2.306	2.896	3.355
9	1.833	2.262	2.821	3.250
10	1.812	2.228	2,764	3.169
11	1.796	2.201	2.718	3.106
12	1.782	2.179	2.681	3.055
13	1.771	2.160	2.650	3.012
14	1.761	2.145	2.624	2.977
15	1.753	2.131	2.602	2.947
16	1.746	2.120	2.583	2.921
17	1.740	2.110	2.567	2.898
18	1.734	2.101	2.552	2.878
19	1.729	2.093	2.539	2.861
20	1.725	2.086	2.528	2.845
21	1.721	2.080	2.518	2.831
22	1.717	2.074	2.508	2.819
23	1.714	2.969	2.500	2.807
24	1.711	2.064	2.492	2.797
25	1.708	2.060	2.485	2.787
26	1.706	2.056	2.479	2.779
27	1.703	2.052	2.473	2.771
28	1.701	2.048	2.467	2.763
29	1.699	2.045	2.462	2.756
30	1.697	2.042	2.457	2.750
40	1.684	2.021	2.423	2.704
60	1.671	2.000	2.390	2.660
120	1.658	1.980	2.358	2.617
∞	1.645	1.960	2.326	2.576*

*For very large v, $t_{\alpha\,;\,v}$ becomes approximately equal to z_{α} .

Table V

Percentage points of the *F*-distribution, Upper 5% points

n_2/n_1	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	161.4	199.5	215.7	224.6	230.2	234.0	236.8	238.9	240.5	241.9	243.9	245.9	248.0	249.1	250.1	251.1	252.2	253.3	254.3
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.41	19.43	19.45	19.45	19.46	19.47	19.48	19.49	19.50
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74	8.70	8.66	8.64	8.62	8.59	8.57	8.55	8.53
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91	5.86	5.80	5.77	5.75	5.72	5.69	5.66	5.63
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68	4.62	4.56	4.53	4.50	4.46	4.43	4.40	4.36
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00	3.94	3.87	3.84	3.81	3.77	3.74	3.70	3.67
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57	3.51	3.44	3.41	3.38	3.34	3.30	3.27	3.23
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.22	3.15	3.12	3.08	3.04	3.01	2.97	2.93
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07	3.01	2.94	2.90	2.86	2.83	2.79	2.75	2.71
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.85	2.77	2.74	2.70	2.66	2.62	2.58	2.54
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.79	2.72	2.65	2.61	2.57	2.53	2.49	2.45	2.40
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69	2.62	2.54	2.51	2.47	2.43	2.38	2.34	2.30
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.60	2.53	2.46	2.42	2.38	2.34	2.30	2.25	2.21
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.53	2.46	2.39	2.35	2.31	2.27	2.22	2.18	2.13
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.48	2.40	2.33	2.29	2.25	2.20	2.16	2.11	2.07

16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.42	2.35	2.28	2.24	2.19	2.15	2.11	2.06	1.01
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.38	2.31	2.23	2.19	2.15	2.10	2.06	2.01	1.96
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.34	2.27	2.19	2.15	2.11	2.06	2.02	1.97	1.92
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.31	2.23	2.16	2.11	2.07	2.03	1.98	1.93	1.88

Table V

F-distribution

n_2/n_1	1	2	3	4	5	6	7	8	9	10	11	12	15	20	24	30	40	60	∞
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.28	2.20	2.12	2.08	2.04	1.99	1.95	1.90	1.84
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.25	2.18	2.10	2.05	2.01	1.96	1.92	1.87	1.81
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.23	2.15	2.07	2.03	1.98	1.94	1.89	1.84	1.78
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.20	2.13	2.05	2.01	1.96	1.91	1.86	1.81	1.76
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.18	2.11	2.03	1.98	1.94	1.89	1.84	1.79	1.73
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.16	2.09	2.01	1.96	1.92	1.87	1.82	1.77	1.71
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.15	2.07	1.99	1.95	1.90	1.85	1.80	1.75	1.69
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20	2.13	2.06	1.97	1.93	1.88	1.84	1.79	1.73	1.67
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19	2.12	2.04	1.96	1.91	1.87	1.82	1.77	1.71	1.65
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22	2.18	2.10	2.03	1.94	1.90	1.85	1.81	1.75	1.70	1.64
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.09	2.01	1.93	1.89	1.84	1.79	1.74	1.68	1.62
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.00	1.92	1.84	1.79	1.74	1.69	1.64	1.58	1.51
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.92	1.84	1.75	1.70	1.65	1.59	1.53	1.47	1.39
120	3.92	3.07	2.68	2.45	2.29	2.17	2.09	2.02	1.96	1.91	1.83	1.75	1.66	1.61	1.55	1.50	1.43	1.35	1.25
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.75	1.67	1.57	1.52	1.46	1.39	1.32	1.22	1.00

Notes