INTERMEDIATE STATISTICAL METHODS

Christian Akrong Hesse
Methodist University College Ghana
20 PUBLICATIONS **44** CITATIONS

SEE PROFILE

John Benjamin Ofosu
Methodist University College Ghana
15 PUBLICATIONS **41** CITATIONS

SEE PROFILE

INTERMEDIATE STATISTICAL METHODS

J. B. Ofosu, BSc, PhD, FSS
Professor of Statistics
Methodist University College Ghana

Frank Otchere, BSc, MA, MPhil
Assistant Research Fellow
Institute of Statistical, Social and
Economic Research, University of Ghana

C. A. Hesse, BSc, MPhil
Lecturer in Statistics
Methodist University College Ghana
+233 244 648 757

Published by : Excellent Publishing & Printing
Printed by : Excellent Publishing & Printing
Cover design by : Daniel Akrong

ISBN: 978-9988-0-7816-4

DISTRIBUTED BY: EPP BOOKS SERVICES
Behind La Trade Fair Centre
P.O. Box TF 490
Trade Fair
Accra – Ghana

email: info@eppbooksservices
website: www.eppbookservices.com
Tel: 0302-784849

PREFACE

This book began many years ago as lecture notes for students at King Saud University in Saudi Arabia and later at the Methodist University College Ghana. Students in the third and fourth years of their undergraduate study, as well as those taking MSc courses, have used draft versions of the chapters, which have been subsequently revised and enriched.

Although some previous knowledge of basic statistical methods is assumed, yet, the coverage in this book has been made very comprehensive to provide quick revision to the necessary basic concepts wherever required. The book contains twelve chapters. Chapter 1 provides useful means of revising the normal, t,χ^2 and F distributions, which are used frequently in the subsequent chapters. Chapters 2, 3, 4 and 5 present the basic ideas of point estimation. Readers are advised to approach these topics gradually, and allow them time to digest before attempting complicated exercises. Chapter 6 covers interval estimation while Chapters 7 and 8 cover hypothesis testing and the likelihood ratio test, respectively. Chapters 9 and 10 deal with single-factor and multifactor experiments, respectively. The notions of randomization, blocking, factorial designs, and interactions are emphasized. Chapter 11 presents the concept of correlation between variables and introduces the principles of regression in the form of simple linear regression. Chapter 12 then presents multiple linear regression, with extensive coverage on applications of matrix algebra in multiple regression.

The major pedagogical features of the book are threefold:

1. The emphasis on concepts and methods through both ***real data*** and ***realistic-type*** applications throughout.
2. Each chapter has an extensive collection of exercises, including end-of-section exercises that emphasize the material in the section. Many of these exercises are from published sources, including past examination questions from King Saud University,

in Saudi Arabia, and Methodist University College Ghana. Answers to all the exercises are given at the end of the book.

3. References cited in each chapter are listed at the end of the chapter.

We are indebted to a large number of people in the production of this book. We particularly appreciate students and professional practitioners who provided feedback, often in the form of penetrating questions that led to rewriting or expansion of the material in the book.

We thank King Saud University and Methodist University College Ghana, for permission to use their past examination questions in Statistics

We have discussed some technical issues with a number of people, including Dr J. Opoku, Dr S. Asare, Dr A. Young, Dr M. Aboukalam, Dr M. Hendi, Dr M. Montasser, and Professors O. A. Y. Jackson, A. M. Aboummoh, A. Al-Zaid, S. Khan, and M. Khalid. To them, we are most grateful.

We are also indebted to Professor Abdullah Al-Shiha of King Saud University, for his permission to publish the statistical tables he used the Minitab software package to prepare. These tables are given in the Appendix.

Last, but not least, we wish to express our appreciation to the editorial staff of E P P Books Services – Ms Wendy Linda Wordie and Mr. Gabriel Nii Abbey – for their continued encouragement.

J. B. Ofosu
F. Otchere
C. A. Hesse

Accra,
January, 2014

CONTENTS

1 SAMPLING DISTRIBUTIONS 1

1.1 Two basic concepts: - population and sample.. 1

1.2 Statistics and sampling distributions.. 7

1.3 Sampling distribution of the mean 9

1.4 Sampling distribution of the difference of means.. 17

1.5 The chi-square distribution.. 20

1.6 The t-distribution.. 26

1.7 The F-distribution.. 29

1.8 Sampling distribution of S^2 33

1.9 Sampling distribution of the sample proportion.. 40

2 POINT ESTIMATION 45

2.1 The problem of point estimation.. 45

2.2 Methods for finding point estimators.. 46

2.3 Comparison of estimators.. 58

3 MINIMUM VARIANCE UNBIASED ESTIMATORS.. 81

3.1 Introduction.. 81

3.2 Lower bound for variance.. 81

3.3 Attainment of the Cramér-Rao lower bound 89

3.4 The exponential family of distributions 96

3.5 Cramér-Rao lower bound, sufficiency and the exponential family 99

3.6 Sufficiency, completeness and minimum variance unbiased estimators 100

4 PROPERTIES OF THE METHOD OF MAXIMUM LIKELIHOOD.. 118

4.1 Introduction.. 118

4.2 Formal statements of main properties.. 123

4.3 Practical aspects.. 129

5 BAYES ESTIMATORS.. 135

5.1 Introduction.. 135

5.2 Posterior distribution.. 136

5.3 Bayesian point estimation………………………………………………………………………... 140

5.4 Discussion……………………………………………………………………………………… 150

6 INTERVAL ESTIMATION………………………………………...…………..…………... 152

6.1 Introduction ……………………………..……………………………………………………… 152

6.2 Pivotal quantity method for constructing confidence intervals………………………………………. 153

6.3 Determination of pivotal quantities……………………………………………………………………. 169

6.4 Large-sample confidence intervals……………………………………………………………….. 162

6.5 Sampling from a normal population………………………………………………………………. 165

6.6 Sampling from two independent normal populations……………………………………………….. 169

6.7 Sampling from Bernoulli populations…………………………………………………………….. 174

7 TESTS OF HYPOTHESES…………………………………………...……………..…………... 178

7.1 Introduction ……………………………..……………………………………………………… 178

7.2 The most powerful test………………………………………………………………………… 189

7.3 Uniformly most powerful test…………………………………………………………………….. 199

7.4 Desirable properties of tests……………………………………………………………………… 218

8 THE LIKELIHOOD RATIO TEST….……………………………...…………..…………... 224

8.1 Introduction ……………………………..……………………………………………………… 224

8.2 Definitions and simple examples……………………………………………………………….. 224

8.3 The asymptotic distribution of the likelihood ratio statistic λ(X)………………………………………. 232

9 DESIGN AND ANALYSIS OF SINGLE-FACTOR EXPERIMENTS...…………..…………... 237

9.1 The nature of experimental investigations…..…...…………………………………………………… 237

9.2 Completely randomized single-factor experiments……………………………………………………. 239

9.3 Interval estimation of parameters………………………………………………………………………. 249

9.4 Multiple comparison methods………………………………………………………………………… 254

9.5 Residual analysis and model checking……………………………………………………………… 258

9.6 Randomized complete block design……………………………………………………………………. 268

9.7 The Latin square design……………………………………………………………………………… 285

10 FACTORIAL EXPERIMENTS…...…………………………………..……………..…………... 305

10.1 Introduction ………………………………..……………………………………………………… 305

10.2 Two-factor factorial experiments……………………………………………………………………… 307

10.3 The 2^k factorial design……………………………………………………………………………….. 319

11 SIMPLE LINEAR REGRESSION AND CORRELATION…………..……………..…………... 328

11.1 Introduction ………………………………..……………………………………………………… 328

11.2 The regression model……………………………………………………………………………… 339

11.3 Estimation of regression parameters……………………………………………………………… 331

11.4 Properties of least squares and maximum likelihood estimators………………………………………. 345

11.5 Inferences concerning linear regression coefficients……………………………………………………… 349

11.6 Evaluation of a regression equation………………………………………………………………. 358

11.7 Confidence and prediction interval estimation……………………………………………………. 364

11.8 Assessing the adequacy of a regression model…………………………………………………. 368

11.9 Transformations to a straight line………………………………………………………………… 374

11.10 Correlation……………………………………………………………………………………… 381

12 MULTIPLE LINEAR REGRESSION…………………………….……………..……………..…………... 391

12.1 Introduction ………………………………..……………………………………………………… 391

12.2 The multiple linear regression model…………………………………………………………… 391

12.3 Least squares estimation of parameters……………...……………………………………….…. 392

12.4 The matrix approach to multiple linear regression……………………………………………… 397

12.5 Properties of least squares estimators and estimation of σ^2 ………………………………………... 402

12.6 Hypothesis testing in multiple linear regression…………………………………………………... 405

12.7 Confidence intervals in multiple linear regression………………………………………………... 411

12.8 Polynomial regression………………………………………………………………………………… 414

APPENDIX……………………………………………………………………………………………….. 424

Table A.1 Cumulative Binomial Distribution……………………………………………………………. 425

Table A.2 Cumulative Poisson Distribution……………………………………………………………... 428

Table A.3 Area under the standard normal curve………………………………………………………... 430

Table A.4 Percentage points of the *t*-distribution……………………………………………………….. 432

Table A.5 Percentage points of the χ^2 distribution………………………………………………………. 434

Table A.6 Percentage points of the F-distribution……………………………………………………………….. 436
Table A.7 Upper Percentage Points of the Studentized Range Distribution……………………………….. 444
Table A.8 Tanh x……………………………………………………………………………………… 445

ANSWERS TO EXERCISES……………………………………………………………………………… **447**

CHAPTER ONE

Sampling Distributions

1.1 Two basic concepts: – population and sample

1.1.1 Population

Statistics is an increasingly important subject which is useful in many types of investigation. It is concerned with the collection, analysis and interpretation of data in which random variation is present. In collecting data concerning characteristics of a group of individuals or subjects, such as the number of people in Ghana who own cars, or the heights of students in a school, it is often impossible or impractical to observe the entire group, especially if it is large. Instead of examining the entire group, called the "population", one examines a small part of the group called a "sample".

In Statistics, we define a population as the set of objects we wish to study. The objects may be people, things, or data (numbers) associated with people, things, or events. For example, the set of plants of a certain type is a population. The set of reactions of all patients suffering from a certain disease, is a population. The set of levels of productivity of all employees under a certain programme, is a population.

A population may be ***finite*** or ***infinite***. If a population of values consists of a fixed number of these values, the population is said to be finite, otherwise, it is infinite. An infinite population consists of an endless succession of values. In practice, the term infinite population is used to refer to a population that cannot be enumerated in a reasonable period of time.

1.1.2 The distribution of a population

For a given population, there is a probability distribution which describes the distribution of values in the population. The random variable X, which has the given probability distribution, is called the ***population random variable***. It is common practice to refer to a population in terms of its corresponding probability distribution. For example, if a population random variable has the normal distribution, then we say that the population has the normal distribution; if the distribution is Poisson, then we say that the population has the Poisson distribution.

The parameters of a population random variable are called ***population parameters.***

Example 1.1

Let X denote the random variable associated with a certain population.

(a) If the p.d.f. of X is given by

$$f(x) = \frac{1}{\theta}e^{-x/\theta}, \quad x \geq 0, \ \theta > 0,$$

then the population has the exponential distribution with mean θ. The parameter θ is a parameter of the distribution. It is also a parameter of the population.

(b) If $P(X = x) = \mu^x e^{-x}/x!, \ x = 0, 1, 2, \ldots,$ then the population has the Poisson distribution with mean μ. The parameter μ is a parameter of the population.

1.1.3 Sampling from a population

If a sample is representative of a population, important conclusions about the population can often be inferred from analysis of the sample. The object of sampling is to form conclusions about a population based on a sample drawn from that population. The following are some of the reasons for sampling.

1. **If it is too costly and time consuming to consider the entire population.**
   For example, it would be costly and time consuming for the Ghana ***Broadcasting*** Corporation Audience Research Unit to ask all viewers – the population of viewers – for their reaction to a certain programme. The Unit, therefore, only interviews a group of viewers – a sample of viewers. Having assessed the reaction of the sample of viewers, the Unit infers from this the views of the entire viewing public.

2. **If the test is destructive.**
   For example, suppose you want to know the average length of life of a batch of car batteries. Since testing of a battery obviously destroys it, we have to consider a sample of batteries, not the complete batch.

3. **If sufficiently accurate results can be obtained from a sample.**
   For example, if the government is to carry out a survey to find the percentage of unemployment in Ghana and a precise figure is not required, then a sample, not the entire population, should be used.

1.1.4 Random samples from a population

There are many ways in which a sample can be taken from a population. For example, if I, as a teacher addressing a class of university students, wanted a sample of five students from my class, perhaps the quickest way would be to take five students sitting in the front row. Is this likely to reflect the properties of the population (the class) adequately? If I were interested in the average height of the students in the class, for example, this would be rather doubtful. Thus, if we wish to make inferences about a population as a whole, we should find out methods of taking a representative sample.

One way in which a representative sample can be obtained is by a process called ***random sampling***, which we now define.

Definition 1.1

Let X be a random variable corresponding to a population. A random sample of size n of the random variable X is a collection of n random variables $X_1, X_2, ..., X_n$, such that

(i) $X_1, X_2, ..., X_n$ are independent,

(ii) $P(X_i \le u) = P(X \le u)$ for all u, $(i = 1, 2, \ldots, n)$.

That is, each X_i has the same distribution as X.

In its simplest form, random sampling means that every member of the population has an equal chance of appearing in the sample, independently of the other members that happen to fall in the sample. More informally, a random sample of size n of a random variable X corresponds to n repeated measurements on X, made under essentially identical conditions. In order for X_1 and X_2 to have the same distribution, the conditions under which X_1 is observed must be the same as those under which X_2 is observed. Of course, experimental conditions can never be identically duplicated. However, those conditions which are different should have little or no effect on the outcome of the experiment.

As a consequence of Definition 1.1, if X is a continuous random variable with p.d.f. f, and if $X_1, X_2,..., X_n$ is a random sample of size n of X, then g, the joint p.d.f. of $X_1, X_2,..., X_n$ can be written as

$$g(x_1, x_2,..., x_n) = f(x_1)f(x_2)...f(x_n). \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (1.1.1)$$

If X is a discrete random variable and $p(x_i) = P(X = x_i)$, $i = 1, 2, \ldots, n$, then

$$P(X_1 = x_1, X_2 = x_2, ..., X_n = x_n) = p(x_1)p(x_2)...p(x_n). \quad \ldots\ldots\ldots \ldots \quad (1.1.2)$$

Example 1.2

Suppose a population random variable X, is the number of cars which enter a certain car park between 9 a.m. and 10 a.m. on Fridays. In order to obtain a random sample of size

n of X, we would presumably choose n Fridays at random and record the values $X_1, X_2, ..., X_n$. We would have to be certain that all the Fridays are "typical" Fridays. For instance, we might not want to include a particular Friday if it happens to coincide with Good Friday.

Sampling from a finite population

Consider a population consisting of N objects represented by 1, 2, …, N. Suppose we select a sample of size n ($n < N$) from this population. Let X_i denote the population value when the i^{th} item is chosen (i = 1, 2, …, n). The distribution of the random variables $X_1, X_2, ..., X_n$ obviously depends on how we go about the sampling. If we sample with replacement, each time choosing an object at random, then the random variables $X_1, X_2, ..., X_n$ are independent and identically distributed. That is, if X is the population random variable, then for each X_i,

$$P(X = j) \;=\; P(X_i = j) \;=\; \frac{1}{N}, \quad i = 1,\ 2,\ ...,\ \text{n}. \quad \dots\dots\dots\dots\dots\dots \quad (1.1.3)$$

Thus, if we sample at random with replacement from a finite population, then the conditions of Definition (1.1) are satisfied, and $X_1, X_2, ..., X_n$ is a random sample of size n from the population. On the other hand, if we sample at random from a finite population without replacement, then the random variables $X_1, X_2, ..., X_n$ are not independent. The conditions of Definition (1.1) are therefore not satisfied and so $X_1, X_2, ..., X_n$ is not a random sample from the population. ***For many practical purposes, sampling from a finite population which is very large, can be considered as sampling from an infinite population***.

One technique for obtaining a random sample from a finite population is to assign numbers to each member of the population, write these numbers on small pieces of paper, place them in a box and then draw numbers from the box, being careful to mix thoroughly before each draw. This method can be replaced by using a table of random numbers specially constructed for such purposes.

When the population size is large, the use of random numbers can become very laborious and at times practically impossible. For instance, if a sample of five cartons of canned fish is to be chosen at random for inspection from among the many thousands stored in a warehouse, one can hardly expect to number all the cartons, make a selection with the use of random numbers, and then pull out the ones that were chosen. In a situation like this, one really has very little choice but to make the selection relatively haphazard, hoping that this will not seriously violate the assumption of randomness which is basic to most statistical theory.

When dealing with infinite populations, the situation is somewhat different since we cannot physically number the elements of the population; but efforts should be made to approach the conditions of randomness by the use of artificial devices. For example, in selecting a sample from a production line, we may be able to approximate conditions of randomness by choosing one unit each half hour; when tossing a coin, we can try to flip it in such a way that neither side is intentionally favoured; and so forth. The proper use of artificial or mechanical device for selecting random samples is always preferable to human judgement, as it is extremely difficult to avoid unconscious biases when making almost any kind of selection.

Even with the careful choice of artificial devices, it is easy to commit gross error in the selection of a random sample. For example, suppose we have the task of selecting logs being fed into a sawmill by a constant-speed conveyor belt, for the purposes of obtaining a random sample of their lengths. One sampling device, which at first sight would seem to be random, consists of measuring the logs which pass a given point at the end of a certain number of 5-minute intervals. However, it can be seen that the longer logs require more time to pass the given point. Thus, the sample favours the longer logs and is therefore not random.

Example 1.3

The life length of an electron tube manufactured by a certain process at a certain factory is studied by choosing n tubes and measuring their life lengths, say $T_1, T_2, ..., T_n$. We may consider $(T_1, T_2, ..., T_n)$ as a random sample from a population of all possible life lengths of electron tubes manufactured in the specified way.

Example 1.4

The total weight of cocoa produced in Ghana in the year 1950 could be defined as a random variable X_1. During successive years, random variables $X_2, X_3, ..., X_n$ could be defined analogously. Again, we may consider $(X_1, X_2, ..., X_n)$ as a sample of size n, obtained from the population of possible yearly production of cocoa in Ghana. It is reasonable to assume that the X_i are independent and identically distributed random variables.

In all applications to be discussed in this book, we shall assume that the sampling has been done in such a manner that Definition (1.1) is satisfied.

Example 1.5

In a study of the distribution of weekly grocery allowances of households in a certain suburb, a consumer research group randomly selected 10 households in the area and found their weekly grocery allowances (in Ghana cedis) to be 75, 48, 60, 42, 62, 70, 80, 90, 81, and 75, respectively. In this example, what is the population, what is the sample, and what is the set of observed values of the random variable?

Solution

The population is the set of all weekly grocery allowances in the suburb.

If X_i is the weekly allowance of the i^{th} household which is contacted in the survey, then $X_1, X_2, ..., X_{10}$ is the random sample of size 10 from the population.

The set of observed values of the sample is $\{75, 48, 60, 42, 62, 70, 80, 90, 81, 75\}$.

Exercise 1(a)

1. Let $A = \{2, 9, 5\}$.
   (a) List all distinct samples of size 2 that can be obtained under sampling without replacement from the set A.
   (b) List all distinct samples of size 2 that can be obtained under sampling with replacement from the set A.
   (c) List all samples of size 2 that can be obtained under sampling without replacement from the set A.
2. An inspector examines every tenth piece coming off an assembly line. List some of the conditions under which this method might not yield a random sample.
3. Explain why the following will not lead to random samples from the desired populations.
   (a) To determine what the average person spends in a month, a market researcher interviews passengers on a luxury cruise.
   (b) To determine public sentiment about certain import restrictions, an interviewer asks voters: "Do you feel that this unfair practice should be stopped?"
4. It is known that a certain process makes vacuum tubes whose lifetimes are exponentially distributed with mean 300 hours. If we randomly select 4 tubes by this process, what is the probability that all 4 last:
   (a) more than 350 hours, (b) less than 300 hours, (c) between 400 and 500 hours?
5. Suppose that X is uniformly distributed over the interval (0, 2). If we take a random

sample of n observations, what is the joint probability density function of the sample?

6. A box contains 3 white and 2 red balls. Suppose we draw balls with replacement, from the box until a red ball is chosen. Let X be the number of draws required. If we take a random sample of 6 observations of X, what is the joint probability mass function of the sample?

7. Let $X_1, X_2, ..., X_{10}$ be a random sample from an exponential population with mean $1/\lambda$.
   (a) What is the joint density function of the sample, as a function of λ?
   (b) What value of λ maximizes the density in (a)?
   (c) If λ is unknown, would you say that the value derived in (b) is a good guess for its unknown value? Explain.

1.2 Statistics and sampling distributions

As we have suggested at the beginning of this chapter, we shall use the information obtained from a sample for the purpose of making inferences about the probability distribution of the random variable being sampled. Since the parameters which characterize a probability distribution are numbers, it is only natural that we would want to compute certain pertinent numerical characteristics obtainable from the sample values, which might help us in some way to make appropriate statements about these parameters which are often unknown. In this section, we shall discuss some of these characteristics. We first give a definition.

Definition 1.2

Let $X_1, X_2, ..., X_n$ be a random sample of a random variable X. Any function of $X_1, X_2, ..., X_n$, which does not depend on unknown parameters, is called a statistic.

A statistic then, is a quantity whose observed value can be computed, once the sample has been taken. Statistics form the bases of all statistical methods of making inferences about populations.

Given a random sample $X_1, X_2, ..., X_n$ of a random variable X, the following statistics are of interest.

(i) The k^{th} sample moment is

$$m'_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k, \quad k = 1, 2, 3, \ldots$$

(ii) The sample mean, $\overline{X}$, is the first sample moment. This is given by $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$.

(iii) The sample variance is $S^2 = \frac{1}{n-1}\sum_{i=1}^{n} X_i - \overline{X}\ ^2$.

The sample standard deviation is *S*. (We shall explain later why we divide by (*n* – 1) rather than by the more obvious choice *n*; see the remark at the end of Theorem 1.17.)

According to Definition 1.2, a statistic is a random variable; the particular values it will equal to vary from sample to sample. It is very important to keep this in mind. It will therefore be meaningful to consider the probability distribution of a statistic, its expectation, and its variance. The probability distribution of a statistic is called its **sampling distribution**. Strictly speaking, the distribution of any random variable is a sampling distribution, but the term is usually reserved for the distribution of a statistic. The sampling distribution of a statistic of a given size describes the way in which the statistic will vary from one sample to another. Sampling distributions are therefore useful in evaluating probability statements about statistics.

Construction of sampling distributions

Sampling distributions may be constructed empirically when sampling from discrete, or a finite population. To construct a sampling distribution we proceed as follows:

1. From a finite population of size *N*, randomly draw all possible samples of size *n*.
2. Compute the statistic of interest for each sample.
3. List in one column, the different distinct observed values of the statistic, and in another column, list the corresponding frequency of occurrence of each distinct observed value of the statistic.

The actual construction of a sampling distribution is a formidable task if the population size is large and is an impossible task if the population is infinite. In such cases, sampling distributions may be approximated by taking a large number of samples of a given size.

Important characteristics of sampling distributions

We usually are interested in knowing three things about a given sampling distribution: its mean, its variance, and its functional form (how it looks when graphed). The standard deviation of the sampling distribution of a statistic is called the ***standard error*** of the statistic.

In the sections below, we shall study the sampling distributions of frequently used statistics.

1.3 Sampling distribution of the mean

An important sampling distribution is the distribution of the sample mean. The following example illustrates how to construct the sampling distribution of the mean.

Example 1.6

Consider a population of size $N = 4$, consisting of the elements 6, 8, 10 and 12. The population mean is

$$\mu = \tfrac{1}{4}(6 + 8 + 10 + 12) = 9,$$

and the population variance is

$$\sigma^2 = \frac{1}{4}\sum_{i=1}^{4}(x_i - \mu)^2 = \frac{1}{4}\ (6-9)^2 + (8-9)^2 + (10-9)^2 + (12-9)^2 \quad = \tfrac{1}{4}(9+1+1+9) = 5.$$

It should be noted that μ and σ are parameters and are usually unknown.

Let us draw a sample of size $n = 2$ from this population, with replacement. These samples, along with their means, are shown in Table 1.1. Notice that there are 16 samples. In general, if we take a sample of size n, with replacement, from a population of size N, then the number of possible samples is equal to N^n.

***Table 1.1*: Samples of size $n = 2$ drawn with replacement, from a population with elements 6, 8, 10 and 12**

		Second draw			
		6	8	10	12
First draw	6	6, 6 (6)	6, 8 (7)	6, 10 (8)	6, 12 (9)
	8	8, 6 (7)	8, 8 (8)	8, 10 (9)	8, 12 (10)
	10	10, 6 (8)	10, 8 (9)	10, 10 (10)	10, 12 (11)
	12	12, 6 (9)	12, 8 (10)	12, 10 (11)	12, 12 (12)

Note: In Table 1.1, sample means are in brackets.

To find the sampling distribution of $\bar{X}$, we determine the frequency of occurrence of values of $\bar{X}$. This is given in Table 1.2. Notice that the data in Table 1.2 satisfy the requirements of a probability distribution.

Table 1.2: Sampling distribution of $\bar{X}$ computed from the samples in Table 1.1

$\bar{x}_i$	6	7	8	9	10	11	12	Total
Frequency (f_i)	1	2	3	4	3	2	1	16
Relative frequency	1/16	2/16	3/16	4/16	3/16	2/16	1/16	16/16

It was stated earlier that we are interested in the functional form of a sampling distribution, its mean, and its variance. We now consider these characteristics for the sampling distribution of $\bar{X}$. Figures 1.1 and 1.2 give a comparison of the distributions of the population and $\bar{X}$.

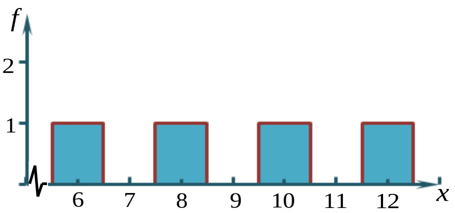

Fig. 1.1: The distribution of a population

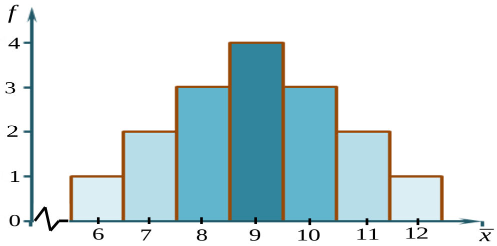

Fig. 1.2: Sampling distribution of $\bar{X}$

Notice the difference between the histogram of the population and the histogram of the sampling distribution of $\bar{X}$. The former is uniformly distributed, whereas the latter gradually rises to a peak and then drops off with perfect symmetry.

We now compute the mean $\mu_{\bar{X}}$ and the variance $\sigma_{\bar{X}}^2$ of the sampling distribution of $\bar{X}$.

$$\mu_{\bar{X}} = \frac{1}{16}\sum_{i=1}^{16} \bar{x}_i f_i = \frac{1}{16}(6 \times 1 + 7 \times 2 + 8 \times 3 + 9 \times 4 + 10 \times 3 + 11 \times 2 + 12 \times 1)$$

$$= \frac{1}{16}(6 + 14 + 24 + 36 + 30 + 22 + 12) = \frac{144}{16} = 9.$$

If can be seen that the mean of the sampling distribution of $\bar{X}$ is equal to the population mean. That is, $\mu_{\bar{X}} = \mu$.

$$\sigma_{\bar{X}}^2 = \frac{1}{16}\sum_{i=1}^{16} f_i \;\; \bar{x}_i - \mu_{\bar{X}} \;\; ^2$$

$$= \frac{1}{16}\Big[1(6-9)^2 + 2(7-9)^2 + 3(8-9)^2 + 4(9-9)^2 + 3(10-9)^2 + 2(11-9)^2 + 1(12-9)^2\Big]$$

$$= \frac{1}{16}(9 + 8 + 3 + 0 + 3 + 8 + 9) = \frac{40}{16} = \frac{5}{2} = 2.5.$$

If can be seen that $\sigma_{\bar{X}}^2 = \sigma^2/n$. We now state and prove these results.

Theorem 1.1

Let $\bar{X}$ denote the mean of a random sample of size n from a population with mean μ and variance σ^2. Then, (a) $E(\bar{X}) = \mu$, (b) $V(\bar{X}) = \sigma^2/n$.

Proof

(a)
$$E(\bar{X}) = E\Big[\frac{1}{n}\ X_1 + X_2 + \dots + X_n\ \Big]$$
$$= \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{1}{n}\sum_{i=1}^{n}\mu = \frac{1}{n}(n\mu) = \mu.$$

(b)
$$V(\bar{X}) = V\Big[\frac{1}{n}\ X_1 + X_2 + \dots + X_n\ \Big] = \frac{1}{n^2}V\ X_1 + X_2 + \dots + X_n\ .$$

Since $X_1, X_2, \dots, X_n$ are independent,

$$V(\bar{X}) = \frac{1}{n^2}\sum_{i=1}^{n} V(X_i)$$
$$= \frac{1}{n^2}\sum_{i=1}^{n}\sigma^2 = \frac{1}{n^2}\ n\sigma^2 = \frac{1}{n}\sigma^2.$$

The quantity $\sigma/\sqrt{n}$ is called the ***standard error of the mean***.

Theorem 1.1 gives two important results concerning the mean of the sampling distribution of $\bar{X}$.

(i) It is equal to the mean of the population from which the samples are drawn, provided the population mean exists. This is independent of the size of the sample.

(ii) Theorem 1.1(b) shows that the standard error of the mean is a decreasing function of *n*. This is a very important result for statistical methods; the larger the sample size, the less variable is $\bar{X}$ from sample to sample. This confirms the intuitive idea that the more observations are taken, the more accurate the mean of the observations will be.

Consider, for example, the following set of 12 numbers:

$$-5,\ -4,\ -3,\ -2,\ -1,\ 0,\ 1,\ 2,\ 3,\ 4,\ 5,\ 6.$$

If we take the average of these numbers, two at a time, in the order listed, we obtain the following set of averages:

$$-4.5,\ -2.5,\ -0.5,\ 1.5,\ 3.5,\ 5.5.$$

If we average the original set of numbers, three at a time, we obtain the following averages

$$-4,\ -1,\ 2,\ 5.$$

If we now average the numbers, four at a time, we obtain

$$-3.5,\ 0.5,\ 4.5.$$

It can be verified that the variance of each of these sets of averages is less than that of the previous set.

The reader must remember Theorem 1.1, as it is one of the most important results in Statistics. It should be pointed out that Theorem 1.1 is independent of the underlying distribution from which the sample was taken, and holds whether it is normal, Poisson or any other distribution with finite variance.

The distribution of $\bar{X}$ when the population is normally distributed

The sampling distribution of $\bar{X}$ depends on the underlying distribution of the population from which the sample was taken. The following theorem gives the distribution of $\bar{X}$ in the case where the population is normally distributed.

Theorem 1.2

Let $X_1, X_2, ..., X_n$ be a random sample of size n from a population which is $N(\mu,\ \sigma^2)$. If $\bar{X} = \frac{1}{n}\sum_{j=1}^{n} X_j$, then $\bar{X}$ is $N(\mu,\ \sigma^2/n)$. That is $\bar{X} - \mu \Big/ \sigma/\sqrt{n})$ is $N(0,\ 1)$.

Proof

We first find the moment generating function of $\bar{X}$ and then apply the uniqueness theorem of moment generating functions. Since $X_1, X_2, ..., X_n$ are independent,

$$M_{\bar{X}}(t) \;=\; \prod_{j=1}^{n} M_{\frac{1}{n}X_j}(t) \;=\; \prod_{j=1}^{n} M_{X_j}\;\; t/n\; .$$

But if X is $N(\mu,\, \sigma^2)$, then $M_X(t) = \exp\;\; \mu t + \frac{1}{2}\sigma^2 t^2\;\;$, (see Ofosu and Hesse (2011b), Theorem 6.19). It follows that

$$\begin{aligned} M_{\bar{X}}(t) &= \prod_{j=1}^{n} \exp\;\; \mu t/n + \tfrac{1}{2}\sigma^2\;\; t/n\;\; ^2 \\ &= e^{\mu t \,+\, \frac{1}{2}\sigma^2 t^2/n} = e^{\mu t \,+\, \frac{1}{2}t^2\;\; \sigma/\sqrt{n}\;\; ^2} \;=\; M_Y(t), \end{aligned}$$

where Y is $N\;\; \mu,\, \sigma^2/n\;\;$. By the uniqueness theorem, $\bar{X}$ is $N\;\; \mu,\, \sigma^2/n\;\;$. Thus, $\;\bar{X} - \mu\; \big/\; \sigma/\sqrt{n})$ is $N(0,\, 1)$.

The distribution of $\bar{X}$ from non-normally distributed populations

The following theorem gives the distribution of $\bar{X}$, in the case where the population is not normally distributed.

Theorem 1.3: (The Central Limit Theorem)

> If $\bar{X}$ is the mean of a random sample of size n taken from a population with mean μ and finite variance σ^2, then $Z = \;\bar{X} - \mu\; \big/\; \sigma/\sqrt{n}\;$ is a random variable whose distribution function approaches that of the standard normal distribution as $n \to \infty$.

The proof of Theorem 1.3 can be found from Ofosu and Hesse (2011b, page 263) and Rao (1973).

The logical question that arises at this point is: How large does the sample have to be before we can apply the central limit theorem? There is no one answer, since the size of the sample needed depends on the extent of non-normality present in the population. One rule of thumb states that, in most practical situations, a sample of size 30 is satisfactory. In general, the approximation to normality of the sampling distribution of $\bar{X}$ becomes better and better as the sample size increases.

Example 1.7

The lifetime X days, of a television picture tube is a random variable with p.d.f.

$$f(x) \;=\; \lambda^2 x e^{-\lambda x}, \quad x \geq 0,$$

where $\lambda = 1/50$. In an experiment, 50 television tubes are placed on a life test, yielding the observed values $X_1, X_2, ..., X_{50}$. Find the probability that the sample mean lies between 90 and 120 days.

Solution

We are required to find $P\ \ 90 < \bar{X} < 120\ \ $. We first find the moment generating function of X and use it to find $\mu = E(X)$ and $\sigma^2 = V(X)$. Now,

$$M_X(t) \;=\; \int_0^\infty e^{xt}\lambda^2 x e^{-\lambda x}dx \;=\; \int_0^\infty \lambda^2 x e^{-x(\lambda - t)}dx.$$

Let $y = x(\lambda - t)$. Then,

$$\begin{aligned}
M_X(t) &= \lambda^2 \int_0^\infty \frac{y}{\lambda - t\ \ ^2} e^{-y}dy \;=\; \frac{\lambda^2}{(\lambda - t)^2}\int_0^\infty y^{2-1}e^{-y}dy \\
&= \ 1 - t/\lambda\ ^{-2}\,\Gamma(2) \;=\; \ 1 - t/\lambda\ ^{-2}, \quad t < \lambda. \\
M'_X(t) &= -2\ \ 1 - t/\lambda\ ^{-3}\ \ -1/\lambda\ \;=\; 2\ \ 1 - t/\lambda\ ^{-3}\big/\lambda \\
M''_X(t) &= -2(3)\ \ 1 - t/\lambda\ ^{-4}\ \ -1/\lambda^2\ \;=\; 6\ \ 1 - t/\lambda\ ^{-4}\ \ 1/\lambda^2\ .
\end{aligned}$$

It follows that

$$\begin{aligned}
\mu &= E(X) = M'_X(0) = 2/\lambda = 100, \\
E(X^2) &= M''_X(0) = 6/\lambda^2, \\
\sigma^2 &= V(X) = E(X^2) - \ E(X)\ ^2 = \frac{6}{\lambda^2} - \frac{4}{\lambda^2} = \frac{2}{\lambda^2} = 5\,000.
\end{aligned}$$

Since the sample size is large, $\ \bar{X} - 100\ \big/\sqrt{5\,000/50} = (\bar{X} - 100)/10$ is approximately $N(0, 1)$. Thus,

$$\begin{aligned}
P\ \ 90 < \bar{X} < 120\ &= P\ \ -1 < \frac{\bar{X} - 100}{10} < 2\ \ = P(-1 < Z < 2), \quad \text{where } Z \text{ is } N(0, 1) \\
&= P(Z < 2) - P(Z < -1) \\
&= 0.9772 - 0.1587 = 0.8185.
\end{aligned}$$

The finite population correction factor

Theorem 1.1 is valid when sampling is either with replacement from a finite population or the samples are drawn from an infinite population. In general, we do not sample with replacement, and in most practical situations, it is necessary to sample from a finite population. In the following example, we consider the sampling distribution of $\bar{X}$ when we sample without replacement from a finite population.

Example 1.8

Consider Example 1.6. Suppose we draw a sample of size 2 without replacement. The sample means that result are those above the principal diagonal of Table 1.1, which are the same as those below the principal diagonal, if we ignore the order in which the observations were drawn. It can be seen that there are 6 possible samples. [In general, if we draw a sample of size *n* from a finite population of size *N* without replacement, and ignore the order in which the sample values are drawn, then the number of possible samples is $\begin{matrix} N \\ n \end{matrix}\Big]$.

The mean of the 6 samples is

$$\mu_{\bar{X}} = \frac{1}{6}\; 7+8+9+9+10+11 \;= 9 = \mu.$$

It can be seen that the mean of the sampling distribution of $\bar{X}$ is equal to the population mean.

The variance of this sampling distribution is

$$\sigma^2_{\bar{X}} = \frac{1}{6}\sum_i \; x_i - \mu_{\bar{X}} \;^2 = \frac{1}{6}\; (-2)^2+(-1)^2+0^2+0^2+1^2+2^2 \; = \frac{5}{3}.$$

It can be seen that the variance of the sampling distribution of $\bar{X}$ is not equal to σ^2/n. However, it can be verified that

$$\sigma^2_{\bar{X}} = \frac{\sigma^2}{n}\;\frac{N-n}{N-1} = \frac{5}{2}\times\frac{4-2}{4-1} = \frac{5}{3}.$$

We may generalize these results in the following theorem.

Theorem 1.4

Let $\bar{X}$ denote the mean of a sample of size *n* taken without replacement from a finite population of size *N*. If the population has mean μ and variance σ^2, then $E(\bar{X}) = \mu$ and $V(\bar{X}) = \frac{\sigma^2}{n}\;\frac{N-n}{N-1}$.

For a proof of this theorem, see Lindgren (1993). The factor $N-n \,/\, N-1$ is often called the **finite population correction factor** and can be ignored when the sample size is small in comparison with the population size. When the population is much larger than the sample, the difference between σ^2/n and $\sigma^2/n\ \ (N-n)/(N-1)$ will be negligible. For example, suppose a population consists of 10 000 observations and a sample of size 20 is selected from this population. The finite population correction factor would be equal to $10\,000-20 \;/\; 10\,000-1 \;=0.9981$. To multiply σ^2/n by 0.9981 is almost equivalent to multiplying it by 1.

Many statisticians do not use the finite population correction factor unless the sample contains more than 5% of the observations in the population. That is, the finite population correction factor is usually ignored when $n/N \le 0.05$.

Example 1.9

In a certain large human population, cranial lengths are approximately normally distributed with mean 185.6 mm and standard deviation 12.7 mm. Find the probability that a sample of size 10 from this population has a mean less that 190 mm.

Solution

We assume that the population is large relative to the sample and so the finite population correction factor can be ignored. Since the population is approximately normally distributed, we assume that, for all practical purposes, $\bar{X}$ is approximately $N\ \ 185.6,\ (12.7)^2/10$. We are required to find $P\ \ \bar{X} < 190$. Now,

$$P\ \ \bar{X} < 190 \quad = \quad P\left(\frac{\bar{X}\ -\ 185.6}{12.7/\sqrt{10}} < \frac{190\ -\ 185.6}{12.7/\sqrt{10}}\right) \quad = \quad P(Z < 1.10) \quad = \quad 0.8643.$$

Exercise 1(b)

1. In a certain survey, it was found that the mean serum cholesterol level for Saudi males aged 25 – 40 years was 180 mg/dl. The standard deviation was approximately 43 mg/dl. Consider the sampling distribution of the sample mean based on a random sample of size 60 from this population of males.
   (a) What is the mean of the sampling distribution?
   (b) Find the standard error of the mean.
   (c) Find the probability that the sample mean serum cholesterol level will be
      (i) between 170 mg/dl and 195 mg/dl,
      (ii) below 175 mg/dl, (iii) greater than 190 mg/dl.

2. A population consists of the following values: 7, 9, 3, 1, 5. Construct the sampling distribution of $\bar{X}$ based on samples of size two selected without replacement. Find the mean and the variance of the sampling distribution.

3. What is the value of the finite population correction factor in the formula for $\sigma^2_{\bar{X}}$ when (a) $n = 5$ and $N = 200$; (b) $n = 100$ and $N = 5\,000$?

4. When we sample from an infinite population, what happens to the standard error of the mean if the sample size is
   (a) increased from 50 to 200, (b) increased from 400 to 900,
   (c) decreased from 225 to 25, (d) decreased from 640 to 400?

5. Let $\bar{X}$ denote the mean of a random sample of size $n = 49$ from a Poisson distribution with mean 4.5. Find $P\ \bar{X} < 5\ .$

6. Let $\bar{T}$ denote the mean of a random sample of size 50 from a population with p.d.f.

$$f(t) = \frac{1}{50}e^{-t/50}, \quad t \geq 0.$$

Find: (a) $P\ 40 < \bar{T} < 60\ ,$ (b) $P\ \bar{T} < 60 | \bar{T} > 40\ ,$ (c) $P\ \bar{T} < 45\ ,$

(d) $P\ \bar{T} > 55\ .$

1.4 Sampling distribution of the difference of means

1.4.1 Introduction

Frequently, the interest in an investigation is focused on two populations. Specifically, an investigator may wish to know something about the difference between two population means. For example, a researcher may wish to know if it is reasonable to conclude that two population means are different. In another situation, the researcher may be interested in the magnitude of the difference between two population means. A medical research team, for example, may want to know whether or not the mean serum cholesterol level is higher in a population of sedentary office workers than in a population of labourers. If the researchers are able to conclude that the population means are different, they may wish to know by how much they differ. A knowledge of the sampling distribution of the difference between two means is useful in investigations of this type.

1.4.2 Sampling distribution of the difference of means

Suppose we have two independent populations, the first with mean μ_1 and variance σ_1^2, and the second with mean μ_2 and variance σ_2^2. Let $\bar{X}_1$ denote the mean of a random sample of size n_1 from the first population and let $\bar{X}_2$ denote the mean of an independent sample of size n_2 from the second population. The following theorem gives the sampling distribution of the statistic $\bar{X}_1 - \bar{X}_2$.

Theorem 1.5

Suppose independent random samples of sizes n_1 and n_2 are drawn from two populations with means μ_1 and μ_2, and variances σ_1^2 and σ_2^2, respectively. Let

$$Z = \frac{\bar{X}_1 - \bar{X}_2 \quad - \quad \mu_1 - \mu_2}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}}. \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \qquad (1.4.1)$$

(a) If the two populations are normally distributed, then Z is N (0, 1).

(b) For large values of n_1 and n_2 $\; n_1 > 30,\ n_2 > 30$, Z is approximately N (0, 1).

Proof

(a) By Theorem 1.2, $\bar{X}_1$ is $N \;\mu_1,\ \sigma_1^2/n_1$ and $\bar{X}_2$ is $N \;\mu_2,\ \sigma_2^2/n_2$. Using the fact that linear combinations of independent normal random variables follow a normal distribution, we can say that the sampling distribution of $\bar{X}_1 - \bar{X}_2$ is $N \;\mu_1 - \mu_2,\ \sigma_1^2/n_1 + \sigma_2^2/n_2$. It follows that Z is N(0, 1).

(b) This result follows from the central limit theorem and part (a) of this theorem.

The quantity Z, given in Equation (1.4.1), can be used to find confidence intervals for $\mu_1 - \mu_2$ (see Ofosu and Hesse (2011a, page 158).

Example 1.10

Electric light bulbs of manufacturer *A* have a mean lifetime of 1 500 hours with a standard deviation of 100 hours, while those of manufacturer *B* have a mean lifetime of 1 200 hours with a standard deviation of 80 hours. What is the probability that a random sample of 125 bulbs from manufacturer *A* will have a mean lifetime that is at least 312 hours more than that of a random sample of 100 bulbs from manufacturer *B*?

Solution

Let $\bar{X}_A$ and $\bar{X}_B$ denote the mean lifetimes of samples *A* and *B*, respectively. We wish to find $P\ \bar{X}_A - \bar{X}_B > 312$. Now, since the sample sizes are large, $\bar{X}_A - \bar{X}_B$ is approximately normally distributed with mean (1 500 – 1 200) = 300 hours and standard deviation $\sqrt{100^2/125 \ + \ 80^2/100} \ = \ 12$ hours. Hence,

$$\begin{aligned} P\ \bar{X}_A - \bar{X}_B > 312 \quad &= \ P\left[\frac{\bar{X}_A - \bar{X}_B - 300}{12} > \frac{312 - 300}{12}\right] \\ &= \ P\ \ Z\ > 1\ ,\ \text{where } Z \text{ is } N\ (0,\ 1) \\ &= \ 1 - P\ \ Z\ \le 1 \quad = \ 1 - 0.8413 \\ &= \ 0.1587. \end{aligned}$$

Exercise 1(c)

1. Given two normally distributed populations with equal means and variances of $\sigma_1^2 = 100$ and $\sigma_2^2 = 80$, what is the probability that samples of size $n_1 = 25$ and $n_2 = 16$ will yield a value of $\bar{X}_1 - \bar{X}_2$ greater than 8?

2. For a population of 18-year-old boys and 18-year-old girls, the means and standard deviations, respectively, of their sub-scapular skin-fold thickness values are as follows: boys, 9.7 and 6.0; girls, 15.6 and 9.5. Simple random samples of 40 boys and 35 girls are selected from the populations. What is the probability that the difference between sample means $\bar{X}_{\text{girls}} - \bar{X}_{\text{boys}}$ will be greater than 10?

3. In a study of annual family expenditures for general health care, two populations were surveyed with the following results:

Population 1: $n_1 = 40$, $\bar{x}_1 =$ GH¢346.00,
Population 2: $n_2 = 35$, $\bar{x}_2 =$ GH¢300.00.

If it is known that the population variances are $\sigma_1^2 = 2\,800$ and $\sigma_2^2 = 3\,250$, what is the probability of obtaining sample results $\bar{X}_1 - \bar{X}_2$ as large as those shown if there is no difference in the means of the two populations?

4. Given two normally distributed populations with equal means and variances of $\sigma_1^2 = 240$ and $\sigma_2^2 = 350$, what is the probability that samples of size $n_1 = 40$ and $n_2 = 35$ will yield a value of $\bar{X}_1 - \bar{X}_2$ greater than 12?

5. A random sample of size $n_1 = 16$ is selected from a normal population with a mean of 75 and a standard deviation of 8. A second random sample of size $n_2 = 9$ is taken from another normal population with mean 70 and standard deviation 12. Let $\bar{X}_1$ and $\bar{X}_2$ be the two sample means. Find:

   (a) $P\ \bar{X}_1 - \bar{X}_2 > 4$, (b) $P\ 3.5 \le \bar{X}_1 - \bar{X}_2 \le 5.5$, (c) $P\ 3 \le \bar{X}_1 - \bar{X}_2 < 5$,

   (d) $P\ \bar{X}_1 - \bar{X}_2 < 9$.

6. The effective life of a component used in a jet-turbine aircraft engine is a random variable with mean 5 000 hours and standard deviation 40 hours. The distribution of effective life is fairly close to a normal distribution. The engine manufacturer introduces an improvement into the manufacturing process for this component that increases the mean life to 5 050 hours and decreases the standard deviation to 30 hours. Suppose that a random sample of $n_1 = 16$ components is selected from the "old" process and a random sample of $n_2 = 25$ components is selected from the "improved" process. Find the probability that the difference in the two sample means $\bar{X}_1 - \bar{X}_2$ is at least 25 hours. Assume that the old and improved processes can be regarded as independent populations.

1.5 The chi-square distribution

The chi-square distribution will occur in several inference problems which we shall study.

Definition 1.3

A continuous random variable X has a chi-square distribution with v degrees of freedom if its p.d.f. is given by

$$f_X(x) = \begin{cases} \frac{1}{2^{v/2}\Gamma\frac{v}{2}} x^{v/2-1}e^{-x/2}, & x > 0, \\ 0, & \text{elsewhere,} \end{cases} \qquad \text{.............................} \quad (1.5.1)$$

where v is a positive integer.

Several chi-square distributions are shown in Fig. 1.3. It can be seen that the chi-square random variable is non-negative and its probability distribution is skewed to the right. However, as v increases, the distribution becomes more symmetric. As $v \to \infty$, the limiting form of the chi-square distribution is the normal distribution.

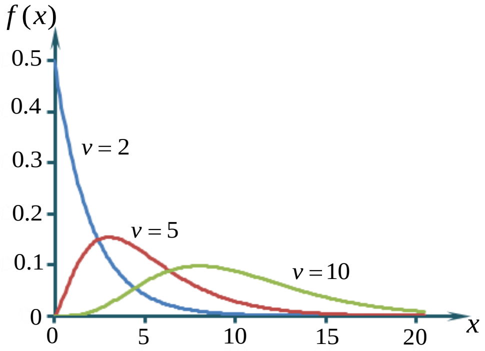

Fig. 1.3: ***The p.d.f's of some chi-sauare distributions***

The percentage points of the chi-square distribution are given in Table A.5 in the Appendix. Define $\chi^2_{\alpha,v}$ as the upper α percentage point of the χ^2 distribution with v degrees of freedom, or the value of the chi-square random variable with v degrees of freedom such that the area to the right is equal to α (see Fig. 1.4).

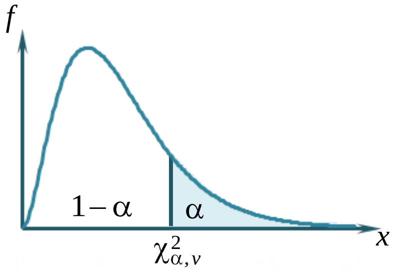

Fig. 1.4: ***The percentage point, $\chi^2_{\alpha,v}$, of the χ^2 distribution with v degrees of freedom***

To illustrate the use of Table A.5, note that the areas α are the column headings and the degrees of freedom, v, are given in the left column. For example, $\chi^2_{0.05,14} = 23.68$. This value is called the upper 5% percentage point of the chi-square distribution with 14 degrees of freedom.

The chi-square distribution has numerous applications in Statistics and some of these applications are discussed by Ofosu and Hesse (2011a).

Properties of the chi-square distribution

Theorem 1.6: (The moment generating function and characteristic function of the χ^2 distribution)

If X has the chi-square distribution with ν degrees of freedom, then its moment generating function and characteristic function are given by

$$M_X(t) = 1-2t^{-\nu/2}, \qquad \phi_X(t) = (1-2it)^{-\nu/2}, \qquad t<\tfrac{1}{2}.$$

Proof

$$M_X(t) = \int_0^\infty \frac{e^{xt}x^{\nu/2-1}e^{-x/2}}{2^{\nu/2}\Gamma\frac{\nu}{2}}\,dx = \frac{1}{2^{\nu/2}\Gamma\frac{\nu}{2}}\int_0^\infty x^{\nu/2-1}e^{-x(1/2-t)}dx.$$

Substituting $y = x\ \frac{1}{2}-t$ and simplifying, we obtain

$$M_X(t) = \frac{1}{2^{\nu/2}\Gamma\frac{\nu}{2}\ \frac{1}{2}-t^{\nu/2}}\int_0^\infty y^{\nu/2-1}e^{-y}dy = \frac{\Gamma\frac{\nu}{2}}{2^{\nu/2}\Gamma\frac{\nu}{2}\ \frac{1}{2}-t^{\nu/2}}$$

$$= 1-2t^{-\nu/2}, \quad t<\tfrac{1}{2}. \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (1.5.2)$$

$$\phi_X(t) = M_X(it) = 1-2it^{-\nu/2}, \quad t<\tfrac{1}{2}. \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (1.5.3)$$

Theorem 1.7: (The mean and the variance of the χ^2 distribution)

If X has the χ^2 distribution with n degrees of freedom, then

$$E(X) = n \text{ and } V(X) = 2n.$$

Proof

$M_X(t) = 1-2t^{-\frac{1}{2}n}$, (see Theorem 1.6).

Hence,

$$M'_X(t) = n\ 1-2t^{-\frac{1}{2}n-1},$$

$$M''_X(t) = n(n+2)\ 1-2t^{-\frac{1}{2}n-2},$$

$$E(X) = M'_X\ 0 = n,\ E(X^2) = M''_X\ 0 = (n+2)n,$$

and

$$V(X) = E(X^2) - E(X)^2 = n^2+2n-n^2 = 2n.$$

Theorem 1.8: (The additivity theorem of the chi-square distribution)

Let $X_1, X_2, ..., X_k$ be independent chi-square random variables with $v_1, v_2, ..., v_k$ degrees of freedom, respectively. Then, $X = X_1 + X_2 + ... + X_k$ has the chi-square distribution with degree of freedom equal to $v = \sum_{i=1}^{k} v_i$.

Proof

Since $X_1, X_2, ..., X_k$ are independent,

$$M_X(t) \;=\; \prod_{j=1}^{k} M_{X_j}(t) \;=\; \prod_{j=1}^{k} (1 - 2t)^{-\frac{1}{2}v_j} \;=\; (1 - 2t)^{-\frac{1}{2}\sum_{j=1}^{k} v_j} \;=\; M_Y(t),$$

where Y has the chi-square distribution with $v = \sum_{j=1}^{k} v_j$ degrees of freedom. By the uniqueness theorem, X has the chi-square distribution with $v = \sum_{j=1}^{k} v_j$ degrees of freedom.

Relationship to the normal distribution

The following theorem gives the relationship between the chi-square and the normal distributions.

Theorem 1.9

If Z is $N(0, 1)$, then Z^2 has the chi-square distribution with 1 degree of freedom.

Proof

We first find the moment generating function of Z^2.

$$M_{Z^2}(t) \;=\; \int_{-\infty}^{\infty} e^{z^2 t}\left(\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}\right) dz \;=\; \frac{2}{\sqrt{2\pi}} \int_0^{\infty} e^{-\frac{1}{2}z^2(1-2t)} \; dz.$$

Substituting $y = \frac{1}{2}z^2(1 - 2t)$ and simplifying, we obtain

$$M_{Z^2}(t) \;=\; \frac{1}{\pi(1 - 2t)^{1/2}} \int_0^{\infty} y^{-\frac{1}{2}} e^{-y}\, dy, \qquad t < \frac{1}{2}$$

$$= \frac{1}{\pi(1 - 2t)^{1/2}} \int_0^{\infty} y^{\frac{1}{2}-1} e^{-y}\, dy \;=\; \frac{\Gamma\ \frac{1}{2}}{\sqrt{\pi}(1 - 2t)^{1/2}}$$

$$= \frac{\sqrt{\pi}}{\sqrt{\pi}(1-2t)^{1/2}} = 1 - 2t\ ^{-1/2}, \qquad t < \tfrac{1}{2}. \ \ldots\ldots\ldots\ldots\ldots \qquad (1.5.4)$$

But this is the moment generating function of the χ^2 distribution with 1 degree of freedom (see Theorem 1.6). Therefore, by the uniqueness theorem, Z^2 has the χ^2 distribution with 1 degree of freedom.

The following theorem shows how a chi-square random variable with n degrees of freedom can be obtained from n independent standard normal random variables.

Theorem 1.10

If $Z_1, Z_2, \ldots, Z_n$ are independent standard normal random variables, then $Y = \sum_{i=1}^{n} Z_i^2$ has the chi-square distribution with n degrees of freedom.

Proof

This theorem follows from Theorems 1.8 and 1.9.

Corollary 1.1

Let $X_1, X_2, \ldots, X_n$ be a random sample of size n from a population which is $N(\mu, \sigma^2)$. Then, $\sum_{i=1}^{n} \ X_i - \mu \ /\sigma \ ^2$ has the chi-square distribution with n degrees of freedom.

Proof

Let $Z_i = \ X_i - \mu \ /\sigma, \quad i = 1, 2, \ldots, n.$ Then $Z_1, Z_2, \ldots, Z_n$ are independent standard normal random variables. The result therefore follows from Theorem 1.10.

The normal approximation to the chi-square distribution

The following theorem gives the normal approximation to the χ^2 distribution.

Theorem 1.11: (The normal approximation to the chi-square distribution)

If X_n has the chi-square distribution with n degrees of freedom, then $Y_n = \ X_n - n \ /\sqrt{2n}$ is asymptotically distributed as $N(0, 1)$ as $n \to \infty$.

Proof

Y_n can be expressed in the form

$$Y_n = \left(\sum_{i=1}^{n} W_i - n\right)\Big/\sqrt{2n} = \left(\frac{1}{n}\sum_{i=1}^{n} W_i - 1\right)n\Big/\sqrt{2n} = \bar{W} - 1 \Big/ \sqrt{2}/\sqrt{n}\ ,$$

where $W_1,\ W_2,\ \ldots,\ W_n$ are independent identically distributed (*i.i.d.*) random variables whose common distribution is chi-square with one degree of freedom. Theorem 1.11 therefore follows from Theorems 1.3 and 1.7. Note that $E(W_i) = 1$ and $V(W_i) = 2,\quad (i = 1,\ 2,\ \ldots,\ n)$.

It has been found that a better normal approximation to the chi-square distribution is obtained by using the following theorem.

Theorem 1.12

If χ_n^2 has the chi-square distribution with *n* degrees of freedom, then

$$T_n = \sqrt{2\chi_n^2} - \sqrt{2n-1} \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \qquad (1.5.5)$$

is asymptotically distributed as $N(0,\ 1)$ as $n \to \infty$.

A proof of Theorem 1.12 is given by Severo and Zelen (1960).

Remarks

Let z_α be defined by the equation

$$P(Z \le z_\alpha) = \alpha,\ \text{where } Z \text{ is } N(0,\ 1). \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \qquad (1.5.6)$$

Using Theorem 1.12, we obtain:

$$z_{1-\alpha} = \sqrt{2\chi_{\alpha,v}^2} - \sqrt{2v-1} \qquad \text{or} \qquad \chi_{\alpha,v}^2 = \tfrac{1}{2}\ z_{1-\alpha} + \sqrt{2v-1}\ ^2.$$

This approximation, which is reasonably good for $v > 30$, can be used to generate a chi-square deviate from a standard normal deviate on a computer (see Severo and Zelen, 1960).

As an illustration, when $\nu = 70$ and $\alpha = 0.05$, $z_{1-0.05} = z_{0.95} = 1.645$. Therefore,

$$\chi_{0.05,70}^2 = \tfrac{1}{2}\ 1.645 + \sqrt{139}\ ^2 = 90.25.$$

The exact value is 90.53.

Example 1.11

Let $X_1, X_2, ..., X_{80}$ be a random sample of size 80 from a population which is $N(10,\ \sigma^2)$ and let $Y^2 = \sum_{i=1}^{80} X_i - 10\ ^2 \big/ 80$. Find $P\ Y^2/\sigma^2 < 1.454$

(a) exactly, (b) by using a normal approximation.

Solution

(a) $P\ Y^2/\sigma^2 < 1.454 \ = \ P(W < 116.32)$, where $W = \frac{80Y^2}{\sigma^2} = \sum_{i=1}^{80} \frac{X_i - 10}{\sigma}\ ^2$.

By Corollary 1.1, W has the χ^2 distribution with 80 degrees of freedom. Using Table A.5, in the Appendix, we obtain $P\ Y^2/\sigma^2 < 1.454 \ = \ 0.995$.

(b) By Theorem 1.11, $W - 80 \big/ \sqrt{2\times 80}$ is approximately $N(0,\ 1)$. Thus,

$$P(W < 116.32) \ = \ P\ Z < \frac{116.32 - 80}{\sqrt{2 \times 80}}\ , \quad \text{where } Z \text{ is } N(0,\ 1)$$

$$= \ P(Z < 2.87) \ = \ 0.9979.$$

Alternatively, using Theorem 1.12, we obtain

$$P(W < 116.32) \ = \ P\ \sqrt{2W} - \sqrt{159} \ < \ \sqrt{232.64} - \sqrt{159}$$

$$= P(Z < 2.64) \ = \ 0.9959.$$

It can be seen that the normal approximation is very close. Moreover, Theorem 1.12 gives a better approximation than Theorem 1.11.

1.6 The *t*-distribution

Definition 1.4

A random variable T has the *t*-distribution with *v* degrees of freedom if its p.d.f. is given by

$$f_T(x) = \frac{\Gamma\ \frac{1}{2}(v+1)}{\sqrt{v\pi}\ \Gamma\ \frac{1}{2}(v)}\ 1 + \frac{t^2}{v}\ ^{-\frac{1}{2}(v+1)}, \qquad -\infty < t < \infty. \ \ldots\ldots\ldots\ldots\ldots \qquad (1.6.1)$$

The *t*-distribution was first published in 1908 in a paper by W. S. Gosset when he was working for an Irish brewery. Gosset published under the pseudonym student. The distribution is therefore often known as the Student's t-distribution. The distribution has the following properties:

1. It has mean 0.
2. It is symmetrical about the mean.

3. As the number of degrees of freedom $v \to \infty$, the limiting form of the t-distribution is the standard normal distribution.

The t-distribution is a family of distributions, since there is a different distribution for each value of v. The t-distribution, like the standard normal distribution, has been extensively tabulated. Table A.4, at the end of the book, contains selected values of α and v, where $t_{\alpha,v}$ is a value of the t-distribution with v degrees of freedom above which we find an area (or probability) α. Thus, $t_{\alpha,v}$ is the upper-tail α percentage point of the t-distribution with v degrees of freedom. In the table, the left hand column contains values of v, the column headings are areas α in the right-hand tail of the t-distribution, and the entries are values of $t_{\alpha,v}$.

It is not necessary to tabulate values of $t_{\alpha,v}$ for α greater than 0.50, since, by the symmetry of the t-distribution, $t_{1-\alpha,v} = -t_{\alpha,v}$. That is, the t-value having an area of $1 - \alpha$ to the right (and therefore an area of α to the left), is equal to the negative of the t-value that has an area of α in the right tail of the distribution (see Fig. 1.5).

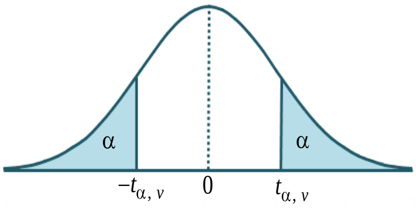

Fig. 1.5: ***Percentage points of the t-distribution***

An important feature of the t-distribution is that it does not depend on any unknown population parameters.

Example 1.12

The t-value with $v = 14$ degrees of freedom that leaves an area of 0.025 to the left, and therefore an area of 0.975 to the right, is $t_{0.975,14} = -t_{0.025,14} = -2.145$.

Example 1.13

Find $P(-t_{0.025,\,10} < T < t_{0.05,\,10})$.

Solution

$$P(-t_{0.025,10} < T < t_{0.05,10}) \;=\; P(t_{0.975,\,10} < T < t_{0.05,\,10}).$$

Since $t_{0.05,10}$ leaves an area of 0.05 to the right, and $t_{0.975,10}$ leaves an area of 0.975 to the right,

$$P(-t_{0.025,10} < T < t_{0.05,10}) \;=\; 0.975 - 0.05 \;=\; 0.925.$$

Example 1.14

Find k such that $P(k < T < -1.761) = 0.045$, where T has the t-distribution with 14 degrees of freedom.

Solution

From Table A.4, we note that 1.761 corresponds to $t_{0.05,\,14}$. Therefore, $-1.761 = -t_{0.05,\,14}$. Let $k = -t_{\alpha,14.}$ The given equation can therefore be expressed in the form $P(-t_{\alpha,14} < T < -t_{0.05,14}) = 0.045$. But, $-t_{0.05,14}$ has an area of 0.05 to the left and $-t_{\alpha,14}$ has an area of α to the left. Hence,

$$0.05 - \alpha \;=\; 0.045, \quad \text{(see Fig. 1.6).}$$

or $$\alpha \;=\; 0.05 - 0.045 = 0.005.$$

Hence,

$$k = -t_{0.005,14} = -2.977, \text{ on using Table A.4.}$$

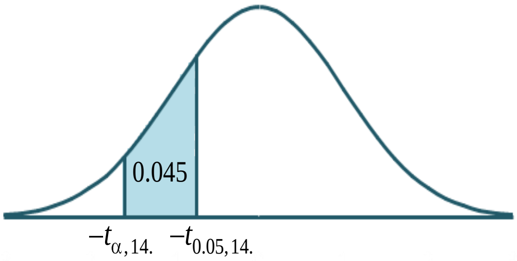

Fig. 1.6: ***The t-values for Example 1.14***

The relationship between the normal, the chi-square and the t-distributions

The following theorem gives the relationship between the normal, the chi-square and the t-distributions.

Theorem 1.13

> Let Z be $N(0, 1)$ and X an independent chi-square random variable with v degrees of freedom. If $T = Z\big/\sqrt{X/v}$, then T has the t-distribution with v degrees of freedom.

The reader is asked to prove this theorem in Exercise 1(d), Question 8.

It should be observed that a t-distribution is completely determined by the parameter v, the number of degrees of freedom of the random variable that has the chi-square distribution.

Example 1.15

Let $Z_1, Z_2, ..., Z_{20}$ be independent standard normal random variables. State, giving reasons, the distributions of

(a) $Y = \sum_{i=1}^{15} Z_i^2$, (b) $Z_{16}\big/ \; Y/15^{\;1/2}$, (c) $Z_1\Big/\left(\frac{1}{19}\sum_{i=2}^{20} Z_i^2\right)^{1/2}$.

Solution

(a) Y is the sum of the squares of 15 independent standard normal random variables. Hence, by Theorem 1.10, Y has the χ^2 distribution with 15 degrees of freedom.

(b) Z_{16} is $N(0,\ 1)$ and Y is an independent χ^2 random variable with 15 degrees of freedom. Hence, $Z_{16}\big/\sqrt{Y/15}$ has the *t*-distribution with 15 degrees of freedom (see Theorem 1.13).

(c) Z_1 is $N(0,\ 1)$ and $\sum_{i=2}^{20} Z_i^2$ is an independent chi-square random variable with 19 degrees of freedom. Hence, $Z_1\Big/\left(\frac{1}{19}\sum_{i=2}^{20} Z_i^2\right)^{1/2}$ has the *t*-distribution with 19 degrees of freedom.

1.7 The *F*-distribution

Another distribution of considerable importance is the F-distribution, which we now define.

Definition 1.5

A random variable X has the *F*-distribution with v_1 and v_2 degrees of freedom if its p.d.f. is given by

$$f_X(x) = \frac{\Gamma\ (v_1\ +\ v_2)/2\quad v_1/v_2\ ^{v_1/2}\ x^{\frac{1}{2}v_1-1}}{\Gamma\ v_1/2\ \Gamma\ v_2/2\ \ 1+xv_1/v_2\ ^{\frac{1}{2}\ v_1+v_2}},\qquad x \ge 0.\quad \ldots\ldots\ldots \qquad (1.7.1)$$

It should be observed that an *F*-distribution is completely determined by the two parameters v_1 and v_2. The order in which the degrees of freedom are given is important since the density of the F-distribution is not symmetrical in v_1 and v_2. The number of degrees of freedom of the numerator in the ratio v_1/v_2 that appears in Equation (1.7.1) is always quoted first.

The *F*-distribution has been extensively tabulated. Table A.6, in the Appendix, contains selected values of $F_{\alpha,\, v_1,\, v_2}$ for various values of α, v_1 and v_2, where $F_{\alpha,\, v_1,\, v_2}$ is a value of the *F*-distribution with v_1 and v_2 degrees of freedom above which we find an area α.

Theorem 1.14

If X has an *F*-distribution with m and n degrees of freedom, then $1/X$ has an *F*-distribution with n and m degrees of freedom.

The reader is asked to prove this result in Exercise 1(d), Question 12.

Remarks

Theorem 1.14 allows us to tabulate the F-distribution for the upper tail only. Table A.6 gives the upper α percentage point of the F-distribution. The lower $1-\alpha$ percentage point can be obtained from the upper α percentage point by using the following equation

$$F_{1-\alpha,\, v_1,\, v_2} = \frac{1}{F_{\alpha,\, v_2,\, v_1}}. \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (1.7.2)$$

For example,

$$F_{0.95,\, 15,\, 10} = \frac{1}{F_{0.05,\, 10,\, 15}} = \frac{1}{2.54} = 0.3937.$$

Example 1.16

If F has the F-distribution with 5 and 10 degrees of freedom, find a and b such that $P(F \leq a) = 0.05$ and $P(F \leq b) = 0.95,$ and accordingly $P(a \leq F \leq b) = 0.90.$

Solution

$$P(F \leq a) = P\ \tfrac{1}{F} \geq \tfrac{1}{a}\ .$$

Therefore,

$$P(F \leq a) = 0.05 \quad \Rightarrow \quad P\ \tfrac{1}{F} \geq \tfrac{1}{a} = 0.05.$$

Now $1/F$ has the *F*-distribution with 10 and 5 degrees of freedom. Hence, from Table A.6,

$$\frac{1}{a} = 4.74, \quad \text{giving} \quad a = \frac{1}{4.74} = 0.211.$$

Furthermore,

$$P(F \le b) = 0.95 \quad \Rightarrow \quad P(F > b) = 0.05 \quad \Rightarrow \quad b = 3.33.$$

Thus,

$$P \;\; a \le F \le b \;\; = P(F \le b) - P(F \le a) = 0.95 - 0.05 = 0.90.$$

The relationship between the chi-square and the F-distributions

The following theorem gives the relationship between the chi-square distribution and the *F*-distribution.

Theorem 1.15

Let X be a chi-square random variable with v_1 degrees of freedom and let Y be a chi-square random variable with v_2 degrees of freedom. If X and Y are independent, then $X/v_1 \;\big/\; Y/v_2$ has the F-distribution with v_1 and v_2 degrees of freedom.

The reader is asked to prove this theorem in Exercise 1(d), Question 10.

Notice that the number of degrees of freedom of the chi-square random variable that appears in the numerator, is always quoted first.

The relationship between the t and the F distributions

The following theorem gives the relationship between the *t* and the *F*-distributions.

Theorem 1.16

Let Z be $N(0,\ 1)$ and X an independent χ^2 random variable with *n* degrees of freedom. By Theorem 1.13, $T = Z\big/\sqrt{X/n}$ has the *t*-distribution with *n* degrees of freedom. By Theorems 1.9 and 1.15, $T^2 = Z^2\big/\; X/n$ has the *F*-distribution with 1 and n degrees of freedom.

It can be seen that, if Y has the t-distribution with *n* degrees of freedom, then Y^2 has the *F*-distribution with 1 and n degrees of freedom.

Exercise 1(d)

1. The random variable X has the continuous uniform distribution over the interval (0, 1).
   (a) Show that $-2\ln X$ has the chi-square distribution with 2 degrees of freedom.
   (b) If $X_1, X_2, ..., X_{10}$ is a random sample of size 10 of X, find the distribution of $\sum_{i=1}^{10} -2\ln X_i$.

2. If $U_1, U_2, ..., U_{13}$ are independent standard normal random variables, find k_1, k_2, and k_3, such that:
   (a) $P\left(\sum_{i=4}^{13} U_i^2 > k_1\right) = 0.75,$ (b) $P\left\{\left(\sum_{i=1}^{6} U_i^2\right)\Big/\left(\sum_{i=7}^{12} U_i^2\right) < k_2\right\} = 0.95,$
   (c) $P\left\{U_1\Big/\left(\frac{1}{9}\sum_{i=2}^{10} U_i^2\right)^{\frac{1}{2}} \le k_3\right\} = 0.025.$
   Give reasons for your answers.

3. If $Z_1, Z_2, ..., Z_{15}$ are independent standard normal random variables, find c_1, c_2 and c_3 such that
   (a) $P\left(\sum_{i=1}^{10} Z_i^2 \le c_1\right) = 0.25,$ (b) $P\left(Z_1\Big/\left(\frac{1}{14}\sum_{i=2}^{15} Z_i^2\right)^{1/2} \le c_2\right) = 0.85,$
   (c) $P\left\{\left(\sum_{i=1}^{10} Z_i^2\right)\Big/\left(\sum_{i=11}^{15} Z_i^2\right) \ge c_3\right\} = 0.025.$

4. Use a table of the *t*-distribution to find the following:
   (a) $t_{0.025,\,14}$, (b) $-t_{0.10,\,10}$, (c) $t_{0.995,\,7}$.

5. If T has the t-distribution with v degrees of freedom, find
   (a) $P(T < 2.365)$, when $v = 7$, (b) $P(T > 1.318)$, when $v = 24$.

6. Let $X_1, X_2, ..., X_{60}$ be a random sample from a population which is $N(10, \sigma^2)$ and let $Y^2 = \sum_{i=1}^{60} X_i - 10\ ^2 /60$. Find $P\ Y^2/\sigma^2 < 1.24$.
   (a) exactly, (b) by using a normal approximation.

7. Let $Z_1, Z_2, ..., Z_{25}$ be independent standard normal random variables. State, giving reasons, the distributions of:

   (a) $X = \left(\frac{1}{10}\sum_{i=1}^{10} Z_i^2\right) \Big/ \left(\frac{1}{15}\sum_{i=11}^{25} Z_i^2\right)$, (b) $Y = Z_1^2 \Big/ \left(\frac{1}{19}\sum_{i=2}^{20} Z_i^2\right)$, (c) Z_1^2 / Z_5^2.

8. Prove Theorem 1.13. [Hint: The p.d.f. of T can be obtained by using the change-of-variable technique (see Hogg et al., 2005 and Chapter 8 of Ofosu and Hesse, 2011b).

9. Let $X_1, X_2, ..., X_n$ be normally and independently distributed with means μ_i and variances $\sigma_i^2, (i = 1, 2, \ldots, n)$. Find the moment generating function of $X_i - \mu_i \ /\sigma_i \ ^2$. Deduce the distribution of $U = \sum_{i=1}^{n} \left(\frac{X_i - \mu_i}{\sigma_i}\right)^2$.

10. Prove Theorem 1.15.

11. If X has the F-distribution with m and n degrees of freedom, prove that $E(X) = n/(n-2)$, for $n > 2$.

    $\left[\text{Hint: Let } X = \ U/m \ / \ V/n \ . \text{ Then } E(X) = \frac{n}{m} E(U) E \ \frac{1}{V} \ \right]$.

12. Prove Theorem 1.14.

13. Let $Y = 2\lambda \sum_{i=1}^{n} X_i$, where $X_1, X_2, ..., X_n$ are i.i.d. exponential random variables with mean $1/\lambda$. Show that Y has the chi-square distribution with $2n$ degrees of freedom.

1.8 Sampling distribution of S^2

Let $X_1, X_2, ..., X_n$ be a random sample of size n of a random variable with mean μ and variance σ^2, and let

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

where $\bar{X}$ is the sample mean. In this section, we consider the sampling distribution of S^2.

Theorem 1.17

$E(S^2) \ = \ \sigma^2.$

Proof

$$\begin{aligned}\sum_{i=1}^{n} X_i - \bar{X}\ ^2 &= \sum_{i=1}^{n} X_i \ - \mu \ + \mu \ - \bar{X}\ ^2 \\ &= \sum_{i=1}^{n} X_i \ - \mu\ ^2 \ + \ 2\ \mu \ - \bar{X} \sum_{i=1}^{n} X_i \ - \mu \ + \sum_{i=1}^{n} \mu \ - \bar{X}\ ^2 \\ &= \sum_{i=1}^{n} X_i \ - \mu\ ^2 \ - \ 2n\ \mu \ - \bar{X}\ ^2 \ + \ n\ \mu \ - \bar{X}\ ^2 \\ &= \sum_{i=1}^{n} X_i \ - \mu\ ^2 \ - \ n\ \mu \ - \bar{X}\ ^2 .\end{aligned}$$

Therefore,

$$\begin{aligned}E(S^2) &= \frac{1}{n-1}\left\{\sum_{i=1}^{n} E\ \ X_i \ - \mu\ ^2 \ - \ nE\ \ \mu \ - \bar{X}\ ^2\right\} \\ &= \frac{1}{n-1}\left\{\sum_{i=1}^{n} V(X_i) \ - \ nV(\bar{X})\right\} \ = \frac{1}{n-1}\ \ n\sigma^2 \ - \ n\sigma^2/n \quad = \ \sigma^2 .\end{aligned}$$

Remark

In defining S^2, we divide by $(n-1)$ instead of n so that the above result will hold. The significance of this result will be discussed in Section 2.3.

The distribution of S^2 depends on the distribution of the population from which the sample is taken. We now consider the case where the population is normally distributed.

Theorem 1.18

Let $X_1, X_2, ..., X_n$ be a random sample of size n from a normal population with mean μ and variance σ^2. Then:

(a) the sample mean $\bar{X}$ and the sample variance S^2 are independent,

(b) $(n-1)S^2/\sigma^2$ has the chi-square distribution with $(n-1)$ degrees of freedom.

Proof

(a) Let $\boldsymbol{Y} = \boldsymbol{AX}$, where

$$\boldsymbol{Y} = \begin{pmatrix} Y_1 \\ . \\ . \\ . \\ Y_n \end{pmatrix}, \qquad \boldsymbol{X} = \begin{pmatrix} X_1 \\ . \\ . \\ . \\ X_n \end{pmatrix},$$

and where $\boldsymbol{A} = \; a_{ij}$ is an $n \times n$ orthogonal matrix with

$$a_{1j} = \frac{1}{\sqrt{n}}, \qquad j = 1,\ 2,\ \ldots,\ n.$$

Thus, $Y_1 = \bar{X}\sqrt{n},$ $\qquad$ (1.8.1)

$$\sum_{i=1}^{n} Y_i^2 = \boldsymbol{Y'Y} = \boldsymbol{X'A'AX} = \boldsymbol{X'X} = \sum_{i=1}^{n} X_i^2, \tag{1.8.2}$$

since matrix $\boldsymbol{A}$ is orthogonal and so $\boldsymbol{A'A} = \boldsymbol{I},$ the $n \times n$ identity matrix.

Without loss of generality, we assume that $\mu = 0.$ The joint probability density function of the X_i is then given by

$$f(x_1,\ x_2,\ \ldots,\ x_n) = \frac{1}{\sigma\sqrt{2\pi}^{\;n}} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n} x_i^2\right). \tag{1.8.3}$$

Hence, g, the joint probability density function of the Y_i is

$$g(y_1,\ y_2,\ \ldots,\ y_n) = f(x_1,\ x_2,\ \ldots,\ x_n)\left|J\right|, \tag{1.8.4}$$

where

$$J = \frac{\partial \boldsymbol{X}}{\partial \boldsymbol{Y}} = \frac{1}{\frac{\partial Y}{\partial X}} = \frac{1}{\left|\boldsymbol{A}\right|}. \tag{1.8.5}$$

But $\left|\boldsymbol{A}\right| = \pm 1,$ since matrix $\boldsymbol{A}$ is orthogonal. Hence from Equations (1.8.2) to (1.8.5),

$$g(y_1,\ y_2,\ \ldots,\ y_n) = \frac{1}{\sigma\sqrt{2\pi}^{\;n}} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n} y_i^2\right) = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} \exp \; -\frac{1}{2\sigma^2} y_i^2 \;,$$

showing that $Y_1,\ Y_2,\ \ldots,\ Y_n$ are independent with each distributed as $N(0,\ \sigma^2)$. Now,

$$(n-1)S^2 = \sum_{i=1}^{n} \; X_i - \bar{X} \;^2 = \sum_{i=1}^{n} X_i^2 - n\bar{X}^2 = \sum_{i=1}^{n} Y_i^2 - Y_1^2,$$

on using (1.8.1) and (1.8.2). It follows that

$$(n-1)S^2 = \sum_{i=2}^{n} Y_i^2, \tag{1.8.6}$$

showing that S^2 and Y_1 are independent and hence by (1.8.1), S^2 and $\bar{X}$ are independent.

(b) From Equation (1.8.6), we obtain $\dfrac{(n-1)S^2}{\sigma^2} = \sum_{i=2}^{n} \dfrac{Y_i^2}{\sigma^2}.$

Therefore $(n-1)S^2/\sigma^2$ is equal to the sum of the squares of $(n-1)$ independent standard normal random variables and so by Theorem 1.10, it has the chi-square distribution with $(n-1)$ degrees of freedom.

Remarks

It is rather remarkable that S^2 and $\bar{X}$, both functions of the same random sample, end up being independent random variables when we take a random sample from a normal population. This theorem is one of the most important results in elementary sampling distribution theory, as we shall see in Theorems 1.19. 1.20 and 1.21. The quantity $(n-1)S^2/\sigma^2$ is used in inferences concerning the population variance σ^2 [see Ofosu and Hesse (2011a)].

Theorem 1.19: (Student's Theorem)

If $\bar{X}$ and S^2 are the mean and variance, respectively, of a random sample of size *n* taken from a normal population with mean μ and unknown variance σ^2, then

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}} \quad \text{..} \quad (1.8.7)$$

has the *t*-distribution with $(n-1)$ degrees of freedom.

Proof

In Theorem 1.2, we proved that $\bar{X} - \mu \Big/ \sigma/\sqrt{n}$ is $N(0, 1)$ and in Theorem 1.18, we proved that $\bar{X}$ and S^2 are independent and $(n-1)S^2/\sigma^2$ has the chi-square distribution with $(n-1)$ degrees of freedom. The result therefore follows from Theorem 1.13 upon writing T in Equation (1.8.7) as

$$T = \frac{\bar{X} - \mu \Big/ \sigma/\sqrt{n}}{\sqrt{n-1 \;\; S^2/(n-1)\sigma^2}}.$$

Remarks

Theorem 1.19 is often referred to as ***Student's Theorem***. The quantity

$$\bar{X}-\mu \Big/ S/\sqrt{n}$$

is used in inferences concerning the population mean μ. In particular, the quantity can be used to find confidence intervals for μ (see Ofosu and Hesse, 2011a).

If we take two independent samples from two normal populations with means μ_1 and μ_2, Theorem 1.5 gives the sampling distribution of $\bar{X}_1-\bar{X}_2$ when the population variances σ_1^2 and σ_2^2 are known. However, in many applications, σ_1^2 and σ_2^2 are unknown and in such cases the sampling distribution of $\bar{X}_1-\bar{X}_2$ is unknown unless it can be assumed that the unknown variances are equal. When this condition is satisfied, the following theorem gives the sampling distribution of $\bar{X}_1-\bar{X}_2$.

Theorem 1.20

If $\bar{X}_1$ and S_1^2 are the mean and variance respectively, of a random sample of size n_1 taken from a normal population with mean μ_1 and unknown variance σ^2, and if $\bar{X}_2$ and S_2^2 are the corresponding statistics for an independent random sample of size n_2 from a second normal population with mean μ_2 and unknown variance σ^2, then the quantity

$$T = \frac{\bar{X}_1-\bar{X}_2 \quad - \quad \mu_1-\mu_2}{S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}, \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (1.8.8)$$

where

$$S_p^2 = \frac{n_1-1\ S_1^2 + \ n_2-1\ S_2^2}{n_1+n_2-2},$$

has the t-distribution with $v = n_1+n_2-2$ degrees of freedom.

Proof

By Theorem 1.5,

$$Z = \frac{\bar{X}_1 - \bar{X}_2 \quad - \quad \mu_1 - \mu_2}{\sigma\sqrt{1/n_1 \ + \ 1/n_2}}$$

has the standard normal distribution. By Theorem 1.18(b), $n_1 - 1\ S_1^2/\sigma^2$ and $n_2 - 1\ S_2^2/\sigma^2$ are independent χ^2 random variables with $(n_1 - 1)$ and $(n_2 - 1)$ degrees of freedom, respectively. Hence, by Theorem 1.8, $U = \ (n_1 - 1)S_1^2 + (n_2 - 1)S_2^2 \ /\sigma^2$ has the χ^2 distribution with $(n_1 + n_2 - 2)$ degrees of freedom. Furthermore, by Theorem 1.18, U and Z are independent. The result therefore follows from Theorem 1.13 upon writing T in (1.8.8) as

$$T \ = \ \frac{Z}{\sqrt{U/\ n_1 + n_2 - 2}} \ = \ \frac{Z}{\sqrt{S_p^2/\sigma^2}}.$$

Remarks

Theorem 1.20 is valid if the samples are independent, the populations are normally distributed, and the population variances are not too different.

The quantity T in Equation (1.8.8) is used in inferences concerning $\mu_1 - \mu_2$ (see Ofosu and Hesse, 2011a, pages 159 and 187).

Theorem 1.21

Let S_1^2 and S_2^2 denote the variances of independent random samples of sizes n_1 and n_2 taken from two distinct normal populations with variances σ_1^2 and σ_2^2, respectively. Then,

$$F \ = \ \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \qquad (1.8.9)$$

has the F-distribution with $(n_1 - 1)$ and $(n_2 - 1)$ degrees of freedom.

Proof

The result follows from Theorem 1.15 and the fact that $(n_1 - 1)S_1^2/\sigma_1^2$ and $(n_2 - 1)S_2^2/\sigma_2^2$ are independent chi-square random variables with $(n_1 - 1)$ and $(n_2 - 1)$ degrees of freedom, respectively (see Theorem 1.18).

Remarks

The quantity F, in Equation (1.8.9), is used in inferences concerning the comparison of variances (see Ofosu and Hesse, 2011a, page 166). Notice again that in the F-distribution, the number of degrees of freedom associated with the χ^2 variable in the numerator is always stated first, followed by that associated with the χ^2 variable in the denominator.

Example 1.17

Let $X_1, X_2, ..., X_9;\ Y_1, Y_2, ..., Y_9$ be independent $N(\mu,\ \sigma^2)$ random variables and let

$$\bar{X} = \tfrac{1}{9}\sum_{i=1}^{9} X_i,\quad \bar{Y} = \tfrac{1}{9}\sum_{i=1}^{9} Y_i,\quad S_1^2 = \tfrac{1}{8}\sum_{i=1}^{9}\ X_i - \bar{X}\ ^2 \quad \text{and} \quad S_2^2 = \tfrac{1}{8}\sum_{i=1}^{9}\ Y_i - \bar{Y}\ ^2.$$

(a) Express $P\left[\ \bar{X} - \bar{Y}\ >\ \sqrt{S_1^2\ +\ S_2^2}\ \right]$ in terms of the *t*-distribution.

(b) Find $P\ \ S_1^2\ <\ S_2^2\ \ .$

Solution

(a) $8S_1^2/\sigma^2$ and $8S_2^2/\sigma^2$ are independent χ^2 random variables, each with 8 degrees of freedom. It follows that

$$U\ =\ 8\ \ S_1^2 + S_2^2\ \ /\sigma^2$$

has the χ^2 distribution with 16 degrees of freedom. Furthermore,

$$Z = 3\ \ \bar{X} - \bar{Y}\ \ /\sigma\sqrt{2}$$

is $N(0,\ 1)$. Moreover, U and Z are independent. It follows that

$$W\ =\ Z/\sqrt{U/16}$$

has the t-distribution with 16 degrees of freedom. But

$$\bar{X} - \bar{Y} > \sqrt{S_1^2 + S_2^2}\quad \Rightarrow\quad W > 3.$$

It follows that

$$P\ \ \bar{X} - \bar{Y} > \sqrt{S_1^2 + S_2^2}\ \ \ =\ \ P\ \ T > 3\ \ ,$$

where T has the *t*-distribution with 16 degrees of freedom.

(b) $$P\ \ S_1^2\ <\ S_2^2\ \ \ =\ P\left(\frac{8S_1^2}{\sigma^2} < \frac{8S_2^2}{\sigma^2}\right)\ \ =\ P\ \ X < Y\ \ ,$$

where X and Y are independent χ^2 random variables, each with 8 degrees of freedom. It follows that

$$P\ \ S_1^2 < S_2^2 \ \ = \int_0^\infty P\ \ X < y|Y = y\ \ f(y)\,dy,$$

where $f(y)$ is the p.d.f. of the chi-square distribution with 8 degrees of freedom. Since *X* and *Y* are independent,

$$P\ \ S_1^2 < S_2^2 \ \ = \int_0^\infty P\ \ X < y\ \ f(y)\,dy \ \ = \ \ \int_0^\infty F(y)\,f(y)\,dy \ \ = \ \int_0^1 z\,dz \ = \ \tfrac{1}{2},$$

where *F* is the distribution function of the χ^2 distribution with 8 degrees of freedom.

Exercise 1(e)

1. State the relationship between:

   (a) the normal and the χ^2 distributions, (b) the normal and the *t*-distributions,

   (c) the normal and the *F*-distributions, (d) the normal, χ^2 and the *t*-distributions,

   (e) the χ^2 and the *F*-distributions, (f) the t and the *F*-distributions.

2. If $X_1, X_2, ..., X_n;\ Y_1, Y_2, ..., Y_n$ are independent $N(\mu, \sigma^2)$ random variables and

   $$n\bar{X} = \sum_{i=1}^{n} X_i,\quad n\bar{Y} = \sum_{i=1}^{n} Y_i,\quad (n-1)S_1^2 = \sum_{i=1}^{n}\ X_i - \bar{X}\ ^2 \quad \text{and} \quad (n-1)S_2^2 = \sum_{i=1}^{n}\ Y_i - \bar{Y}\ ^2,$$

   express:

   (a) $P\ \ \bar{X} - \bar{Y} > \sigma\ \ ,$ (b) $P\left(\sum_{i=1}^{n}\ X_i - Y_i\ ^2 < n\right),$

   (c) $P\ \ \bar{X} - \bar{Y} > \sqrt{S_1^2 + S_2^2}\ \ ,$ (d) $P\ \ S_1^2 < S_2^2\ \ .$

   in terms of the normal, chi-square, or *t* distributions.

3. If F has an F-distribution with 12 and 20 degrees of freedom, find the values of *a* and *b* such that $P(F \leq a)\ =\ 0.025,\qquad P\ \ F \leq b\ \ =\ 0.975.$

1.9 Sampling distribution of the sample proportion

In the above sections, we have considered the sampling distributions of statistics computed from measured variables. In this section, we consider the sampling distributions of statistics that result from counts or frequency data.

1.9.1 Population proportion

If 5% of a population are smokers, then we say that the proportion of smokers in the population is $\frac{5}{100}$ or 0.05. This is an example of a population proportion. Many questions of interest relate to population proportions. What proportion of patients who receive a particular type of treatment recover? What proportion of some population has a certain disease? What proportion of a population are immune to a certain disease? We denote a population proportion by p. It can be seen that:

$$p = \frac{\text{number in the population with a given characteristic}}{\text{total number in the population}}.$$

1.9.2 Sample proportion

If we take a random sample of 250 patients from a hospital and 50 of the patients have blood group A, then the proportion of patients in the sample with blood group A, is $\frac{50}{250} = \frac{1}{5}$. This is a sample proportion. We use the sample proportion as an estimate of p. This estimate is denoted by $\hat{p}$. The corresponding statistic, $\hat{P}$, denotes the point estimator of p.

1.9.3 The distribution of $\hat{P}$

Suppose we select n students at random from a school and find that m of these students have blood group A. Then, the proportion of students in the sample with blood group A, is $\hat{P} = m/n$. It can be seen that, $n\hat{P}$ of the students in the sample, have blood group A. What is the distribution of $n\hat{P}$? It is clear that $n\hat{P}$ has the binomial distribution with parameters (n, p), where p is the proportion of students in the school with blood group A. Thus,

$$E(n\hat{P}) = np \quad \text{and} \quad V(n\hat{P}) = np(1-p), \text{ giving}$$

$$E(\hat{P}) = p \quad \text{and} \quad V(\hat{P}) = np(1-p)/n^2 = p(1-p)/n.$$

Therefore, $\mu_{\hat{P}} = p$ and $\sigma^2_{\hat{P}} = p(1-p)/n$.

What is the distribution of $\hat{P}$? The following theorem gives the distribution of $\hat{P}$ when n is large.

Theorem 1.22

When the sample size is large, the distribution of $\hat{P}$ is approximately $N\ \ p, p(1-p)/n$.

For a proof of Theorem 1.22, see Rao (1973) and Loève (1955). This normal approximation is valid if $np > 5$ and $n(1-p) > 5$.

Example 1.18

Suppose it is known that 90% of a certain population are non-smokers. If a random sample of size 200 is drawn from this population, find the probability that the sample proportion of non-smokers is less than 0.85.

Solution

Both np and $n(1-p)$ are greater than 5 $[np = 200 \times 0.9 = 180,\ n(1-p) = 200 \times 0.1 = 20]$ and so we can assume that the sampling distribution of $\hat{P}$ is approximately normally distributed with mean $\mu_{\hat{P}} = 0.90$ and $\sigma^2_{\hat{P}} = 0.90(0.1)/200 = 0.000\,45$. Hence,

$$P\ \ \hat{P} < 0.85 \quad = \quad P\left(\frac{\hat{P} - 0.9}{\sqrt{0.000\,45}} < \frac{0.85 - 0.90}{\sqrt{0.000\,45}}\right)$$

$$= \quad P\ \ Z < -2.36 \quad = \quad 0.0091, \quad \text{where } Z \text{ is } N(0,\ 1).$$

1.10 Sampling distribution of the difference between two sample proportions

Often there are two populations in which we are interested and we desire to assess the probability associated with a difference in proportions computed from samples drawn from each of these populations. The relevant sampling distribution is the distribution of the difference between the two sample proportions.

Example 1.19

Consider the problem where we wish to estimate the difference between two population proportions, p_1 and p_2. For example, we might let p_1 be the proportion of smokers with lung cancer and p_2 the proportion of non-smokers with lung cancer. Our problem, then, is to estimate the difference between these two proportions. First, we select independent random samples of sizes n_1 and n_2 from the two populations and then determine the sample proportions $\hat{P}_1$ and $\hat{P}_2$. What is the distribution of $\hat{P}_1 - \hat{P}_2$? The following theorem gives the large sample distribution of $\hat{P}_1 - \hat{P}_2$.

Theorem 1.23

If independent random samples of size n_1 and n_2 are drawn from two populations of dichotomous variables, where the proportions of observations with the characteristic of interest in the two populations are p_1 and p_2, respectively, the distribution of the difference between the sample proportions, $\hat{P}_1 - \hat{P}_2$, is approximately normal with mean

$$\mu_{\hat{P}_1-\hat{P}_2} = p_1 - p_2 \text{ and variance } \sigma^2_{\hat{P}_1-\hat{P}_2} = \frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}, \text{when}$$

n_1 and n_2 are large. Thus, when n_1 and n_2 are sufficiently large,

$$Z = \frac{\hat{P}_1 - \hat{P}_2 \quad - \quad p_1 - p_2}{\sqrt{\dfrac{p_1(1-p_1)}{n_1} + \dfrac{p_2(1-p_2)}{n_2}}} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (1.10.1)$$

is approximately $N(0,\ 1)$.

For a proof of Theorem 1.23, see Rao (1973) and Loève (1955). We consider n_1 and n_2 sufficiently large when $n_1 p_1$, $n_2 p_2$, $n_1(1-p_1)$ and $n_2(1-p_2)$ are all greater than 5.

The quantity Z in Equation (1.10.1) is used in inferences concerning population proportions (see Ofosu and Hesse (2011a), page 170).

Example 1.20

In school A, the proportion of girls is $p_1 = 0.50$ and in school B, the proportion of girls is $p_2 = 0.33$. What is the probability that random samples of size 100 drawn from each of the schools will yield a value $\hat{P}_1 - \hat{P}_2$ greater than 0.3?

Solution

$$\mu_{\hat{P}_1-\hat{P}_2} = 0.50 - 0.33 = 0.17, \text{ and } \sigma^2_{\hat{P}_1-\hat{P}_2} = \frac{0.33 \times 0.67}{100} + \frac{0.5 \times 0.5}{100} = 0.004\,711.$$

$n_1 p_1$, $n_2 p_2$, $n_1(1-p_1)$ and $n_2(1-p_2)$ are all greater than 5 and so we assume that $\hat{P}_1 - \hat{P}_2$ is approximately normally distributed with mean 0.17 and variance 0.004 711.

$$P \;\; \hat{P}_1 - \hat{P}_2 > 0.3 \;\; = P\left(\frac{\hat{P}_1 - \hat{P}_2 - 0.17}{\sqrt{0.004\,711}} > \frac{0.3 - 0.17}{\sqrt{0.004\,711}}\right) = P \;\; Z > 1.89 \;\; = 0.0294,$$

where Z is $N(0,\ 1)$.

Exercise 1(f)

1. Given a population in which $p = 0.6$ and a random sample of size 100 from this population, find:

   (a) $P\ \hat{P}\ \geq\ 0.65$, (b) $P\ \hat{P}\ \leq\ 0.58$, (c) $P\ 0.56 < \hat{P}\ <\ 0.63$.

2. In a survey conducted in 1989, 19% of the respondents, 16 years of age, stated that they had not heard of the AIDS virus, HIV. Find the probability that in a sample of size 175 from this population, at least 25% would not have heard of the AIDS virus, HIV.
3. In a certain population of males, 19.4% are known to be obese. If 150 males are selected from this population, calculate the probability that less than 15% will be obese.
4. In a certain population of teenagers, it is known that 10% of the boys are obese. If the same proportion of girls in the population are obese, what is the probability that a random sample of 250 boys and 200 girls will yield a value of $\hat{P}_1 - \hat{P}_2\ \geq\ 0.06$?
5. A certain survey revealed that 14% of males and 23.8% of females deviated from their desirable weight by 20% or more. Suppose we select a random sample of 120 males and an independent random sample of 130 females. Calculate the probability that the difference between the sample proportions, $\hat{P}_F - \hat{P}_M$, will be between 0.04 and 0.20.

References

Hogg, R. V., Mckean, J. W. and Craig, A. T. (2005). Introduction to Mathematical Statistics, 6th Edition. *Pearson Education Ltd. London.*

Lindgren, B. W. (1993) Statistical Theory, 4th Edition. *Chapman and Hall, London.*

Loève, M. (1955). Probability Theory. *D. Van Nostrand, Princeton, N. J.*

Ofosu, J. B. and Hesse, C. A. (2011a). Elementary Statistical Methods, 2nd Edition. *E P P Books Services, Accra.*

Ofosu, J. B. and Hesse, C. A. (2011b). Introduction to probability and probability distributions. *E P P Books Services, Accra.*

Rao, C. R. (1973). Linear Statistical Inference and its applications, 2nd Edition. *John Wiley and Sons, New York.*

Severo, N. C. and Zelen, M. (1960). Normal approximation to the chi-square and non-central *F* probability functions. *Biometrika*, **47,** 411 – 416.

CHAPTER TWO

Point Estimation

2.1 The problem of estimation

In Chapter 1, we learnt that a sample from a population is useful in making inferences about the population. Two important problems in statistical inference are estimation and tests of hypotheses. One type of estimation, namely, point estimation, is the subject of this chapter.

The problem of estimation, as it shall be considered herein, is loosely defined as follows: Assume that some characteristic of the elements in a population can be represented by a random variable X whose p.d.f. (or probability mass function) is $f(x;\,\boldsymbol{\theta})$, a function of known form depending on an unknown parameter $\boldsymbol{\theta} = (\theta_1, \theta_2, ..., \theta_m)$, which has m components. The allowable range of $\boldsymbol{\theta}$ is defined as the ***parameter space*** Ω. For example, if X is $N(\mu,\, \sigma^2)$, then $\boldsymbol{\theta} = (\mu,\, \sigma^2)$, and $\Omega = \{-\infty < \mu < \infty,\ 0 < \sigma^2 < \infty\}$. Let $\boldsymbol{X} = (X_1,\, X_2,\, ...,\, X_n)$ and $\boldsymbol{x} = (x_1,\, x_2,\, ...,\, x_n)$, a value of $\boldsymbol{X}$. Our aim is to find a statistic $t(\boldsymbol{X})$, whose value $t(\boldsymbol{x})$, would give an approximate value of $\boldsymbol{\theta}$. This statistic is called an estimator of $\boldsymbol{\theta}$ and $t(\boldsymbol{x})$ is called an ***estimate*** of $\boldsymbol{\theta}$.

There are two kinds of estimates in common use. An estimate of a population parameter, expressed by a single number, is called a ***point estimate***. An estimate of a population parameter, expressed by two numbers between which the parameter is expected to lie with a given level of confidence, is called an ***interval estimate*** (see Chapter 6).

As an example, suppose that $X_1,\, X_2,\, ...,\, X_n$ is a random sample of size n from a population with an unknown mean μ. The sample mean is a point estimator of μ. That is, $\hat{\mu} = \bar{X}$. After the sample has been taken, the numerical value $\bar{x}$, of $\bar{\boldsymbol{X}}$, is the point estimate of μ. Thus, if $n = 4$, $x_1 = 25$, $x_2 = 30$, $x_3 = 29$ and $x_4 = 31$, then the point estimate of μ is $\bar{x} = \frac{1}{4}(25 + 30 + 29 + 31) = 28.75$.

Point estimation admits two problems: the first is to find a statistic which can be used as an estimator; the second, is to select criteria and techniques to define and find a "best" estimator among many possible estimators. In Section 2.2, we introduce several methods of finding point estimators. One of these, and probably the most important, is the method of ***maximum likelihood***. Another method, called the method of ***least squares***, will be discussed in Chapter 11. In Section 2.3, several "optimum" properties of estimators are defined. These include unbiasedness, relative efficiency, consistency and sufficiency. It should be noted that since an estimator is a statistic, it is a random variable. The sampling distribution of this statistic can be used to measure the quality of the estimator.

2.2 Methods for finding point estimators

In this section, we consider several methods for finding point estimators.

2.2.1 The likelihood function

One of the best methods of obtaining a point estimate is the method of maximum likelihood. Before we introduce the maximum likelihood method, we discuss the concept of ***likelihood function***, which plays a fundamental role in statistical inference.

Definition 2.1 (The likelihood function)

Suppose that X is a random variable with p.d.f. (or probability mass function) $f(x;\, \theta_1, \theta_2, ..., \theta_m)$, where $\theta_1, \theta_2, ..., \theta_m$ are unknown parameters. Let $x_1, x_2, ..., x_n$ be the observed values in a random sample of size n of X. The likelihood function of the sample is

$$l(\theta_1, \theta_2, ..., \theta_m) = l(\boldsymbol{\theta}) = \prod_{i=1}^{n} f(x_i;\, \theta_1, \theta_2, ..., \theta_m) = \prod_{i=1}^{n} f(x_i;\, \boldsymbol{\theta}), \ldots\ldots \qquad (2.2.1)$$

where $\boldsymbol{\theta} = \theta_1, \theta_2, ..., \theta_m$.

Example 2.1

Three independent observations on a Poisson distribution with an unknown mean μ are 6, 9 and 11. What is the likelihood function?

Solution

The probability mass function of the Poisson distribution is

$$P(X = x) = \frac{\mu^x e^{-\mu}}{x!}, \quad x = 0, 1, 2, ...$$

The likelihood function is therefore given by

$$l(\mu) = \frac{\mu^6 e^{-\mu}}{6!} \frac{\mu^9 e^{-\mu}}{9!} \frac{\mu^{11} e^{-\mu}}{11!} = \frac{\mu^{26} e^{-3\mu}}{6!9!11!}.$$

It can be seen that the likelihood function is the joint probability of an observed sample, regarded as a function of the unknown parameters. The random variables are taken as fixed at their observed values.

Example 2.2

Let $x_1, x_2, ..., x_n$ be the observed values in a random sample of size n from a population which is $N(\mu, \sigma^2)$. The likelihood function is

$$\begin{aligned} l(\mu, \sigma) &= \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma^2}\left(x_i - \mu\right)^2\right) \\ &= \frac{1}{\sigma^n \left(2\pi\right)^{n/2}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(x_i-\mu\right)^2\right\}. \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (2.2.2) \end{aligned}$$

Simplification of the likelihood function

Two points simplify the task of plotting the likelihood function. First, likelihood functions are nearly always obtained by multiplying the probabilities of independent events, and the likelihood function usually simplifies by considering the logarithm. If $l(\boldsymbol{\theta})$ is a likelihood function, we write

$$L(\boldsymbol{\theta}) = \ln l(\boldsymbol{\theta}). \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (2.2.3)$$

For example, the log-likelihood for Example 2.1 is

$$L(\mu) = 26\ln\mu - 3\mu - \ln\left(6!9!11!\right). \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (2.2.4)$$

The second simplification is that, since likelihood functions have relative values only, for different values of the unknown parameters, terms in the likelihood or log-likelihood not involving the unknown parameters may be dropped. For example, in Equation (2.2.4), we can drop the last term and in Equation (2.2.2), we can express the likelihood function in the form

$$l(\mu, \sigma) = \frac{k}{\sigma^n}\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(x_i-\mu\right)^2\right\}, \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (2.2.5)$$

where k is a constant.

2.2.2 The method of maximum likelihood

We first define the maximum likelihood estimator.

Definition 2.2 (The maximum likelihood estimator)

Given the likelihood function $l(\boldsymbol{\theta})$, we choose as a point estimate of $\boldsymbol{\theta}$, that value $\hat{\boldsymbol{\theta}}$, which maximizes $l(\boldsymbol{\theta})$, such that $\boldsymbol{\theta} \in \Omega$.

Since the maxima of $l(\boldsymbol{\theta})$ and $\ln l(\boldsymbol{\theta})$ occur at the same value of $\boldsymbol{\theta}$, in finding the maximum likelihood estimate of $\boldsymbol{\theta}$, we usually maximize $L(\boldsymbol{\theta}) = \ln l(\boldsymbol{\theta})$ with respect to $\theta_1, \theta_2, ..., \theta_m$. Thus, if certain regularity conditions are satisfied (see Section 3.2), the maximum likelihood estimates of $\theta_1, \theta_2, ..., \theta_m$ are the roots of the m ***likelihood equations***

$$\frac{\partial}{\partial\theta_i} L(\theta_1, \theta_2, ..., \theta_m) = 0, \quad i = 1, 2, ..., m. \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (2.2.6)$$

It turns out that this method of estimation has many desirable properties, which we shall discuss in Chapter 4. The main justification depends on asymptotic properties, which hold as $n \to \infty$, but there are some very useful small-sample properties as well (see Chapter 4).

Example 2.3

Suppose that the lifetime X days, of a component is a random variable with p.d.f.

$$f(x;\, \theta) = \frac{1}{\theta} e^{-x/\theta}, \quad x > 0.$$

If n such components are tested and $x_1, x_2, ..., x_n$ are their failure times, find the maximum likelihood estimator of θ.

Solution

The likelihood function is

$$l(\theta) = \frac{1}{\theta^n} e^{-\frac{1}{\theta}\sum_{i=1}^{n} x_i} = \frac{1}{\theta^n} e^{-n\bar{x}/\theta}.$$

$$L(\theta) = \ln l(\theta) = -n \ln\theta - n\bar{x}/\theta.$$

$$\frac{\partial}{\partial\theta} L(\theta) = -\frac{n}{\theta} + \frac{n\bar{x}}{\theta^2} \quad \text{and} \quad \frac{\partial^2}{\partial\theta^2} L(\theta) = \frac{n}{\theta^2} - \frac{2n\bar{x}}{\theta^3}.$$

The likelihood equation is

$$-\frac{n}{\theta} + \frac{n\bar{x}}{\theta^2} = 0.$$

This gives $\theta = \bar{x}$. Notice that, when $\theta = \bar{x}$, $\frac{\partial^2}{\partial\theta^2} L(\theta) < 0$, so $\theta = \bar{x}$ maximizes $L(\theta)$. The maximum likelihood estimate of θ is therefore given by $\hat{\theta} = \bar{x}$. The corresponding statistic, $\hat{\theta} = \bar{X}$, is the maximum likelihood estimator of θ.

Suppose that for $n = 5$, we obtained the following failure times (in days): 30, 33.5, 31, 29.5 and 36. The maximum likelihood estimate of θ, based on this sample would be

$$\hat{\theta} = \frac{1}{5}(30 + 33.5 + 31 + 29.5 + 36) = 32.$$

Remarks

If a population is discrete, the likelihood function is given by

$$l(\theta) \;=\; P(X_1 = x_1,\, X_2 = x_2,\, \ldots,\, X_n = x_n).$$

This is the probability of observing the sample, as a function of θ. If a population is continuous, $l(\theta)$ is proportional to the probability of observing a sample outcome in the neighbourhood of the values actually observed. Thus, the method of maximum likelihood selects the value of θ for which the probability of obtaining the observed data is a maximum. In other words, the maximum likelihood method selects the most likely value of θ based on the sample [see Barnard et al. (1962)].

Example 2.4

A random sample of size n is taken from a population which is $N(\mu,\, \sigma^2)$. Find the maximum likelihood estimators of μ and σ^2.

Solution

If the observed values are $x_1,\, x_2,\, \ldots,\, x_n$, then the likelihood function is

$$l(\mu,\, \sigma) \;=\; \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} \exp\; -\frac{1}{2\sigma^2}\;\; x_i \;-\; \mu\;\; ^2 \;=\; \frac{1}{2\pi\sigma^2}\;\;^{n/2} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\;\; x_i - \mu\;\; ^2\right\}.$$

The logarithm of the likelihood function is

$$L(\mu,\, \sigma) \;=\; -\frac{1}{2} n \ln\,(2\pi) - \frac{1}{2} n \ln \sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}\;\; x_i - \mu\;\; ^2.$$

Therefore,

$$\frac{\partial L}{\partial \mu} \;=\; \frac{1}{\sigma^2}\sum_{i=1}^{n}\;\; x_i - \mu \quad \text{and} \quad \frac{\partial L}{\partial \sigma^2} \;=\; -\frac{n}{2\sigma^2} \;+\; \frac{1}{2\sigma^4}\sum_{i=1}^{n}\;\; x_i - \mu\;\; ^2.$$

On equating these partial derivatives to zero and solving the resulting equations for μ and σ^2, we obtain

$$\hat{\mu} \;=\; \frac{1}{n}\sum_{i=1}^{n} x_i \;=\; \bar{x} \quad \text{and} \quad \hat{\sigma}^2 \;=\; \frac{1}{n}\sum_{i=1}^{n}\;\; x_i - \bar{x}\;\; ^2.$$

A check on the second partial derivatives shows that these maximize $L(\mu,\, \sigma)$, so they are maximum likelihood estimates. The maximum likelihood estimators of μ and σ^2 are therefore given by

$$\hat{\mu} \;=\; \bar{X} \quad \text{and} \quad \hat{\sigma}^2 \;=\; \frac{1}{n}\sum_{i=1}^{n}\;\; X_i - \bar{X}\;\; ^2.$$

The method of maximum likelihood is often the estimation method that statisticians prefer, because it is usually easy to use and produces estimators with good statistical properties. However, sometimes complications are encountered. For instance, it is not always easy to maximize the likelihood function because the equation obtained from $dL(\theta)/d\theta = 0$ may be difficult to solve. Furthermore, it may not always be possible to directly use calculus methods to determine the maximum of $L(\theta)$. These are illustrated in the following two examples

Example 2.5

Let $X_1, X_2, ..., X_n$ be a random sample of size *n* of the random variable *X* with p.d.f.

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

Find the maximum likelihood estimator of θ.

Solution

The likelihood function of a random sample of size *n* is

$$l(\theta) = \prod_{i=1}^{n} \frac{1}{\theta} = \begin{cases} \frac{1}{\theta^n}, & 0 < x_i \le \theta \text{ for } i = 1, 2, ..., n, \\ 0, & \text{otherwise.} \end{cases}$$

Since θ exceeds all the observations if and only if it exceeds the largest x_i, the likelihood function can be written as

$$l(\theta) = \begin{cases} \frac{1}{\theta^n}, & \theta \ge \max(x_1, x_2, ..., x_n), \\ 0, & \text{otherwise.} \end{cases} \qquad (2.2.7)$$

Fig. 2.1 shows a graph of $l(\theta)$, where $x_{(n)} = \max\limits_{1 \le i \le n}(x_i)$.

Calculus methods cannot be used to find the value of θ which maximizes $l(\theta)$, because the maximum value of $l(\theta)$ occurs at a point of discontinuity of $l(\theta)$. It can be seen that $l(\theta)$ is a decreasing function of θ for $\theta \ge \max\limits_{1 \le i \le n}(x_i)$ and is zero elsewhere. Hence the maximum value of $l(\theta)$ occurs at the smallest value of θ. It follows that the maximum likelihood estimate of θ is given by $\hat{\theta} = x_{(n)} = \max\{x_i\}$. The corresponding statistic, $\hat{\theta} = X_{(n)} = \max\limits_{1 \le i \le n}(X_i)$, is the

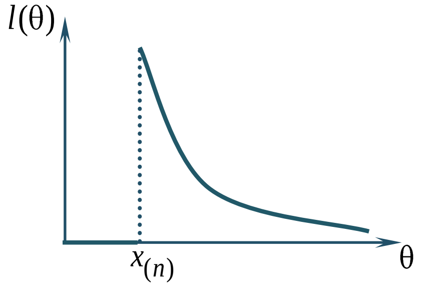

Fig. 2.1: ***The likelihood function for Example 2.5***

maximum likelihood estimator of θ.

If for $n=5$, we obtain the values 2.3, 4.5, 2, 5.1, 3.6, then $\hat{\theta}=5.1$.

Example 2.6

The random variable X has the gamma distribution with p.d.f.

$$f_X(x)=\begin{cases}\dfrac{x^{\alpha-1}e^{-x/\beta}}{\beta^{\alpha}\Gamma(\alpha)}, & x\geq 0,\ \alpha>0,\ \beta>0,\\ 0, & \text{elsewhere.}\end{cases}$$

Find the maximum likelihood estimators of α and β, based on a random sample of size n.

Solution

Let $x_1, x_2, \ldots, x_n$ denote the observed values of the random sample. The likelihood function is

$$l(\alpha,\ \beta) = \prod_{i=1}^{n}\frac{x_i^{\alpha-1}e^{-x_i/\beta}}{\beta^{\alpha}\Gamma(\alpha)} = \left\{\frac{1}{\beta^{\alpha}\Gamma(\alpha)}\right\}^{n}\left(\prod_{i=1}^{n}x_i\right)^{\alpha-1} e^{-\left(\sum\limits_{i=1}^{n}x_i\right)\Big/\beta}.$$

$$L(\alpha,\ \beta) = \ln l(\alpha,\ \beta) = -n\alpha\ln\beta - n\ln\Gamma(\alpha) + (\alpha-1)\sum_{i=1}^{n}\ln x_i - \left(\sum_{i=1}^{n}x_i\right)\Big/\beta.$$

$$\frac{\partial L}{\partial\alpha} = -n\ln\beta - n\frac{d}{d\alpha}\ln\Gamma(\alpha) + \sum_{i=1}^{n}\ln x_i.$$

$$\frac{\partial L}{\partial\beta} = -\frac{n\alpha}{\beta} + \frac{n\overline{x}}{\beta^2}.$$

The maximum likelihood estimates of α and β are the roots of the equations

$$-n\ln\beta - n\frac{d}{d\alpha}\ln\Gamma(\alpha) + \sum_{i=1}^{n}\ln x_i = 0. \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (2.2.8)$$

$$-\frac{n\alpha}{\beta} + \frac{n\overline{x}}{\beta^2} = 0. \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (2.2.9)$$

Unfortunately, no explicit solution to these equations exist. This is a problem we sometimes meet in applications of the maximum likelihood method. Now, from Equation (2.2.9), $\beta=\overline{x}/\alpha$.

Substituting this in (2.2.8), we obtain

$$-n\ln\ \overline{x}/\alpha\ - n\frac{d}{d\alpha}\ln\Gamma(\alpha) + \sum_{i=1}^{n}\ln x_i = 0. \ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (2.2.10)$$

Equation (2.2.10) can be solved numerically for α (see Chapter 4).

Example 2.6 shows that we cannot always obtain maximum likelihood estimators by solving the likelihood equation

$$\frac{\partial}{\partial\theta}L(\theta) = 0.$$

An excellent survey on the maximum likelihood method of estimation is given by Norden (1972, 1973) and Sprott and Kalbefleisch (1969).

Example 2.7

Suppose that each trial of an experiment can result in k mutually exclusive outcomes, $A_1, A_2, \ldots, A_k$, with corresponding probabilities $p_1, p_2, \ldots, p_k$, where $\sum_{i=1}^{k} p_i = 1$. If n trials of this experiment give f_i outcomes of the event A_i, (i = 1, 2, …, k), find the maximum likelihood estimators of $p_1, p_2, \ldots, p_k$.

Solution

The likelihood function is

$$l(p_1, p_2, \ldots, p_k) = \frac{n!}{f_1!\, f_2! \ldots f_k!} p_1^{f_1} p_2^{f_2} \ldots p_k^{f_k}.$$

Thus,

$$\ln l = c + \sum_{i=1}^{k} f_i \ln p_i, \quad \text{where } c = \ln\left(\frac{n!}{f_1!\, f_2! \ldots f_k!}\right).$$

In maximizing $\ln l$ with respect to $p_1, p_2, \ldots, p_k$, the probability vector $(p_1, p_2, \ldots, p_k)$, is restricted by the condition $\sum_{i=1}^{k} p_i = 1$. Using the Lagrange method, we maximize

$$h(p_1, p_2, \ldots, p_k) = \ln l - \lambda \sum_{i=1}^{k} p_i.$$

$$\frac{\partial h}{\partial p_i} = \frac{f_i}{p_i} - \lambda, \quad i = 1, 2, \ldots, k.$$

Setting these partial derivatives to zero, we obtain

$$\frac{f_i}{p_i} = \lambda, \quad i = 1, 2, \ldots, k \text{ or } p_i = \frac{f_i}{\lambda}, \quad i = 1, 2, \ldots, k.$$

But $\sum_{i=1}^{k} p_i = 1$. Therefore, $\frac{1}{\lambda}\sum_{i=1}^{k} f_i = 1$, or $1 = n/\lambda$, which gives $\lambda = n$.

Thus, $p_i = f_i/n, \quad i=1, 2, \ldots, k$. This means that $\hat{p}_i = f_i/n$, the relative frequency of A_i. The maximum likelihood estimator of p_i is therefore the relative frequency of A_i.

2.2.3 The method of moments

Let $x_1, x_2, ..., x_n$ be the observed values of a random sample of size *n* of the random variable *X*, whose p.d.f. or probability mass function $f(x; \theta_1, \theta_2, ..., \theta_m)$ depends on the unknown parameters $\theta_1, \theta_2, ..., \theta_m$. It is assumed that the functional form of *f* is known. The k^{th} population moment is $\mu'_k = E(X^k)$. In general, μ'_k will be a known function of the *m* parameters $\theta_1, \theta_2, ..., \theta_m$. The k^{th} sample moment is

$$M'_k = \frac{1}{n}\sum_{i=1}^{n} x_i^k .$$

The method of moments estimates of $\theta_1, \theta_2, ..., \theta_m$ are the roots of the *m* equations

$$\mu'_k = M'_k, \quad k=1, 2, ..., m. \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (2.2.11)$$

These equations reduce, for *m* = 1, to

$$E(X) = \bar{x}$$

and for $m = 2$, to

$$E(X) = \bar{x}, \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (2.2.12)$$

and $$E(X^2) = \frac{1}{n}\sum_{i=1}^{n} x_i^2 . \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (2.2.13)$$

Equation (2.2.13) is equivalent to

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

$$= \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \bar{x}^2 = \frac{1}{n}\sum_{i=1}^{n} \ x_i - \bar{x}\ ^2 . \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (2.2.14)$$

Remarks

The method of moments estimators of $\theta_1, \theta_2, ..., \theta_m$ are based on the strong law of large numbers, which states that if $X_1, X_2, ..., X_n$ is a random sample of size *n* of the random variable *X*, then $\frac{1}{n}\sum_{i=1}^{n} X_i^k \xrightarrow{a.s.} E(X^k)$. Thus, if the k^{th} population moment is finite, then the k^{th} sample moment $\xrightarrow{a.s.}$ the k^{th} population moment (see Ofosu and Hesse (2011), page 260).

Example 2.8

In Example 2.3, find the method of moments estimate of θ.

Solution

We first find $E(X)$. This is given by

$$E(X) = \int_0^\infty \tfrac{x}{\theta} e^{-x/\theta} dx. \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (2.2.15)$$

Integrating by parts, we obtain

$$E(X) = \left[-xe^{-x/\theta}\right]_0^\infty + \int_0^\infty e^{-x/\theta} dx = 0 + \left[-\theta e^{-x/\theta}\right]_0^\infty = \theta.$$

The moment estimate of θ is given by $\hat{\theta} = \bar{x}$. The moment estimator of θ is therefore given by $\hat{\theta} = \bar{X}$.

Note: $E(X)$ can also be obtained by substituting $y = x/\theta$ in Equation (2.2.15). This gives

$$E(X) = \int_0^\infty \tfrac{x}{\theta} e^{-x/\theta} dx = \int_0^\infty \theta y e^{-y} dy = \theta \int_0^\infty y^{2-1} e^{-y} dy = \theta\Gamma(2) = \theta,$$

as before.

It can be seen that, in Examples 2.3 and 2.8, the method of moments gives the same estimator as the maximum likelihood method. We now consider an example in which the two methods give different estimators.

Example 2.9

In Example 2.5, find the method of moments estimator of θ.

Solution

Here, $E(X) = \int_0^\theta \tfrac{x}{\theta}\, dx = \left[\frac{x^2}{2\theta}\right]_0^\theta = \tfrac{1}{2}\theta.$

The moment estimate of θ is therefore given by

$$\tfrac{1}{2}\tilde{\theta} = \bar{x} \quad \text{or} \quad \tilde{\theta} = 2\bar{x}.$$

Thus, the moment estimator of θ is given by

$$\tilde{\theta} = 2\bar{X}. \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (2.2.16)$$

This estimator is different from the maximum likelihood estimator of θ (see Example 2.5).

Example 2.10

The random variable *X* has the geometric distribution with probability mass function

$$P(X=x) = pq^{x-1}, \quad x = 1, 2, 3, \ldots, p+q=1.$$

Find the moment estimator of *p* based on a random sample of size *n* of *X*.

Solution

We first find $E(X)$.

$$\begin{aligned} E(X) &= \sum_{x=1}^{\infty} xq^{x-1}p = p\sum_{x=1}^{\infty} xq^{x-1} \\ &= p\sum_{x=1}^{\infty}\frac{d}{dq}q^x = p\frac{d}{dq}\left(\sum_{x=1}^{\infty} q^x\right) = p\frac{d}{dq}\frac{q}{1-q} \\ &= p\left\{\frac{(1-q)(1) - q(-1)}{(1-q)^2}\right\} = \frac{p}{(1-q)^2} = \frac{p}{p^2} = \frac{1}{p}. \end{aligned}$$

The moment estimate of p is therefore given by (see Equation (2.2.12))

$$\frac{1}{\hat{p}} = \bar{x} \quad \text{or} \quad \hat{p} = \frac{1}{\bar{x}}.$$

The corresponding statistic, $\hat{p} = \frac{1}{\bar{X}}$, is the moment estimator of p.

Example 2.11

In Example 2.6, we found that application of the maximum likelihood method gave two equations that were impossible to solve except by numerical methods. We now determine the method of moments estimates of the parameters α and β. We first find $E(X)$. This is given by

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

Let $y = x/\beta$. Then, $x = \beta y$, $dx = \beta dy$, and

$$E(X) = \int_0^{\infty}\frac{(\beta y)^{\alpha}e^{-y}}{\beta^{\alpha}\Gamma\ \alpha}\beta dy = \int_0^{\infty}\frac{\beta y^{\alpha+1-1}e^{-y}}{\Gamma\ \alpha}dy = \frac{\beta\Gamma(\alpha+1)}{\Gamma(\alpha)}.$$

Using the result $\Gamma(\alpha+1) = \alpha\Gamma(\alpha)$, we obtain

$$E(X) = \frac{\beta\alpha\Gamma(\alpha)}{\Gamma(\alpha)} = \alpha\beta. \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (2.2.17)$$

We next find $E(X^2)$. This is given by

$$E(X^2) \;=\; \int_0^\infty \frac{x^{\alpha+1}e^{-x/\beta}}{\beta^\alpha \Gamma\ \alpha}\,dx.$$

Substituting $y = x/\beta$, we obtain

$$\begin{aligned}
E(X^2) \;&=\; \int_0^\infty \frac{(\beta y)^{\alpha+1}e^{-y}}{\beta^\alpha \Gamma\ \alpha}\beta dy \;=\; \beta^2 \int_0^\infty \frac{y^{\alpha+1}e^{-y}}{\Gamma\ \alpha}\,dy \\
&=\; \beta^2 \int_0^\infty \frac{y^{\alpha+2-1}e^{-y}}{\Gamma\ \alpha}\,dy \;=\; \frac{\beta^2\Gamma(\alpha+2)}{\Gamma(\alpha)} \;=\; \frac{\beta^2}{\Gamma(\alpha)}\Gamma(\alpha+1+1) \\
&=\; \frac{\beta^2}{\Gamma(\alpha)}(\alpha+1)\Gamma(\alpha+1) \;=\; \frac{\beta^2(\alpha+1)\alpha\Gamma(\alpha)}{\Gamma(\alpha)} \;=\; \alpha^2\beta^2+\alpha\beta^2.
\end{aligned}$$

Thus,

$$V(X) \;=\; \alpha^2\beta^2+\alpha\beta^2-\alpha^2\beta^2 \;=\; \alpha\beta^2. \qquad (2.2.18)$$

Therefore, the method of moments estimates of α and β are given by the roots of the equations (see Equations (2.2.12), (2.2.14), (2.2.17) and (2.2.18))

$$\hat{\alpha}\hat{\beta} \;=\; \bar{x}, \qquad (2.2.19)$$

and
$$\hat{\alpha}\hat{\beta}^2 \;=\; \frac{1}{n}\sum_{i=1}^{n} x_i-\bar{x}\;^{2}. \qquad (2.2.20)$$

Squaring each side of Equation (2.2.19), we obtain

$$\hat{\alpha}^2\hat{\beta}^2 \;=\; \bar{x}^2. \qquad (2.2.21)$$

From Equations (2.2.20) and (2.2.21), by division, we obtain

$$\hat{\alpha} \;=\; \frac{n\bar{x}^2}{\sum_{i=1}^{n} x_i-\bar{x}\;^{2}},$$

and
$$\hat{\beta} \;=\; \frac{\bar{x}}{\hat{\alpha}} \;=\; \frac{\sum_{i=1}^{n} x_i\;-\bar{x}\;^{2}}{n\bar{x}}.$$

Remarks

Not only does the method of moments often enable us to estimate parameters that are difficult to estimate by means of the maximum likelihood method, but it is highly useful for estimating properties of distributions for which no explicit density functions are available. For example, we may wish to estimate the mean and variance of a random variable whose distribution is unknown. If we assume that the population possesses the first two moments, we can use the first two sample moments to estimate the first two population moments.

Example 2.12

Let $X_1, X_2, \ldots, X_n$ be a random sample of size n of the random variable X with p.d.f.

$$f(x;\,\theta) \;=\; \frac{1}{\pi\left[1+(x-\theta)^2\right]}, \qquad -\infty < x < \infty,$$

where θ is an unknown parameter. Why can't we apply the method of moments to estimate θ?

Solution

It can be shown that the characteristic function of X is $\exp(-|t|+it\theta)$. This is not differentiable at the point $t=0$ and so $E(X)$ is not finite. We cannot therefore equate $E(X)$ to $\overline{x}$ to obtain the moment estimate of θ.

2.2.4 Other methods for finding point estimators

There are several other methods for finding point estimators of parameters. Among these are: (a) the method of least squares, which is discussed in Chapter 11, (b) the Bayes method, which is discussed in Chapter 5, (c) the minimum-chi-square and the minimum distance methods, which are discussed in Mood et al. (1974).

Exercise 2(a)

1. Let $X_1, X_2, \ldots, X_n$ be a random sample of size n from the Poisson distribution with probability mass function

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

Find: (a) the maximum likelihood estimator of λ, (b) the moment estimator of λ.

2. Let $X_1, X_2, \ldots, X_n$ be a random sample of size n from the geometric distribution with parameter p. That is,

$$P(X = x) \;=\; p(1-p)^{x-1}, \qquad x = 1,\ 2,\ \ldots\ .$$

Find: (a) the maximum likelihood estimator of p, *(b)* the moment estimator of p.

3. Let $X_1, X_2, \ldots, X_n$ be a random sample of size n from the beta distribution with

p.d.f. $$f(x;\,\theta) \;=\; \begin{cases} \theta x^{\theta-1}, & 0 < x < 1,\ \theta > 0, \\ 0, & \text{otherwise.} \end{cases}$$

Find: (a) the maximum likelihood estimator of θ, (b)the moment estimator of θ.

4. Let $x_1, x_2, ..., x_n$ be the observed values of a random sample of size n from a gamma distribution with p.d.f.

$$f(x;\ \alpha,\ \lambda) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1}e^{-\lambda x}, \qquad x>0,\ \lambda>0,\ \alpha>0.$$

(a) Find the likelihood function and the log likelihood function.

(b) Find the equations that define the maximum likelihood estimates of α and λ. Can they be solved explicitly?

(c) Find the maximum likelihood estimate of $\mu = \alpha/\lambda$.

(d) Show that the moment estimates of α and λ are given by

$$\hat{\alpha} = \frac{n\bar{x}^2}{\sum_{i=1}^{n} x_i - \bar{x}\ ^2}, \qquad \hat{\lambda} = \frac{n\bar{x}}{\sum_{i=1}^{n} x_i - \bar{x}\ ^2}.$$

5. Let $X_1, X_2, ..., X_n$ be a random sample of size n from a distribution with p.d.f.

$$f(x;\ \theta) = \begin{array}{ll} e^{-(x-\theta)}, & 0<\theta<x<\infty, \\ 0, & \text{otherwise.} \end{array}$$

Find: (a) the maximum likelihood estimator of θ, (b) the moment estimator of θ.

2.3 Comparison of estimators

In many situations, it is possible to find several statistics which could be used as estimators of an unknown population parameter (see, for example, Examples 2.5 and 2.9). A choice between these estimators is based on the sampling distributions of the estimators. In this section, we consider some of the desirable properties required of an estimator.

2.3.1 Unbiasedness

An estimator of an unknown parameter should be “close”, in some sense, to the true value of the parameter. One desirable property of an estimator is that of unbiasedness.

Definition 2.3 (Unbiased estimator)

An estimator $\hat{\theta}$ of θ, based on a sample of size n, is said to be unbiased if $E(\hat{\theta}) = \theta$ for all possible values of θ and for all n.

If $\hat{\theta}$ is an unbiased estimator of θ, it means that $\hat{\theta}$ takes, “on average”, the value of θ, the quantity it is designed to estimate.

Definition 2.4 (The bias of an estimator)

If an estimator $\hat{\theta}$ is not unbiased for θ, then the difference $E(\hat{\theta})-\theta$ is called the bias of the estimator.

Example 2.13

Let $X_1, X_2, ..., X_n$ be a random sample of size *n* from a population with an unknown mean μ and an unknown variance σ^2. Let $\bar{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$ and $S^2=\frac{1}{n-1}\sum_{i=1}^{n}\ X_i-\bar{X}\ ^{2}$.

(a) In Theorem 1.1, we proved that $E(\bar{X})=\mu$. Therefore, the sample mean $\bar{X}$ is an unbiased estimator of the population mean μ.

(b) In Theorem 1.17, we proved that $E(S^2)=\sigma^2$. S^2 is therefore an unbiased estimator of σ^2.

(c) $V(S)=E(S^2)-\ E(S)\ ^{2}=\sigma^2-\ E(S)\ ^{2}$. Since $V(S)>0$, $E(S)<\sigma$, showing that S is a biased estimator of σ.

(d) If μ is known, then $\frac{1}{n}\sum_{i=1}^{n}\ X_i-\mu\ ^{2}$ is an unbiased estimator of σ^2.

Example 2.14

Refer to Examples 2.5, and 2.9. Show that the moment estimator of θ is unbiased, but the maximum likelihood estimator of θ is biased.

Solution

The moment estimator of θ is given by $\tilde{\theta}=2\bar{X}$ (see Example 2.9).

$$E(\tilde{\theta})\ =\ 2E(\bar{X})\ =\ 2E(X)\ =\ 2\int_0^{\theta}\frac{x}{\theta}\,dx\ =\ \left[\frac{x^2}{\theta}\right]_0^{\theta}\ =\ \theta. \quad \dots\dots\dots\dots\dots\dots \quad (2.3.1)$$

$\tilde{\theta}$ is therefore an unbiased estimator of θ.

The maximum likelihood estimator of θ is given by $\hat{\theta}=X_{(n)}$ (see Example 2.5). The distribution function of $X_{(n)}$ is given by

$$G_{X_{(n)}}(x) = P(X_{(n)} \le x) = \prod_{i=1}^{n} P(X_i \le x), \text{ by independence}$$

$$= \left(\int_0^x \frac{1}{\theta}\,dy\right)^n = \begin{cases} \left(\frac{x}{\theta}\right)^n, & 0 < x \le \theta, \\ 0, & \text{otherwise.} \end{cases}$$

The p.d.f. of $X_{(n)}$ is therefore given by

$$g_{X_{(n)}}(x) = \begin{cases} nx^{n-1}/\theta^n, & 0 < x \le \theta, \\ 0, & \text{otherwise.} \end{cases}$$

Thus, $E(\hat{\theta}) = \int_0^\theta \frac{nx^n}{\theta^n}\,dx = \frac{n\theta}{n+1}$. .... (2.3.2)

$E(\hat{\theta}) \neq \theta$ and so $\hat{\theta}$ is not an unbiased estimator of θ.

2.3.2 Mean square error and relative efficiency

Unbiasedness, by itself, is not enough to ensure that an estimator is "good". Suppose $\hat{\theta}_A$ and $\hat{\theta}_B$ are unbiased estimators of θ. This means that the distributions of these estimators are centered at the true value of θ. However, the variances of the estimators may be different. Fig. 2.2 illustrates the situation. Since $\hat{\theta}_A$ has a smaller variance than $\hat{\theta}_B$, the estimator $\hat{\theta}_A$ is more likely to produce an estimate close to the true value of θ than the estimator $\hat{\theta}_B$. Thus, if $\hat{\theta}_A$ and $\hat{\theta}_B$ are both unbiased estimators of θ based on the same sample size, we would prefer $\hat{\theta}_A$ to $\hat{\theta}_B$ if

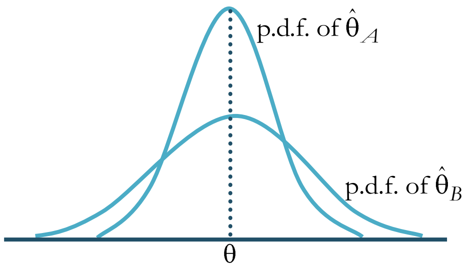

Fig. 2.2: ***Sampling distributions of two unbiased estimators, $\hat{\theta}_A$ and $\hat{\theta}_B$***

$$V(\hat{\theta}_A) < V(\hat{\theta}_B), \quad \text{for all } \theta \in \Omega. \quad \text{...} \quad (2.3.3)$$

If (2.3.3) is satisfied, we say that $\hat{\theta}_A$ is more efficient than $\hat{\theta}_B$.

Mean square error

Sometimes, it is necessary to use a biased estimator. If an estimator is biased, its variance is not used for comparison. The relevant measure is the mean square error which we now define.

Definition 2.5 **(Mean square error)**

The mean square error of an estimator $\hat{\theta}$ of the parameter θ is defined as

$$MSE(\hat{\theta}) = E\left[(\hat{\theta}-\theta)^2\right]. \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (2.3.4)$$

The mean square error can be expressed in the following form:

$$\begin{aligned} MSE(\hat{\theta}) &= E\left[\ \hat{\theta} - E(\hat{\theta}) + E(\hat{\theta}) - \theta\ ^2\right] \\ &= E\left[\ \hat{\theta}-E(\hat{\theta})\ ^2\right] + E\left[\ E(\hat{\theta})-\theta\ ^2\right] + 2E\left[\ \hat{\theta}-E(\hat{\theta}) \quad E(\hat{\theta})-\theta\ \right] \\ &= V(\hat{\theta}) + \ E(\hat{\theta})-\theta\ ^2 + 2\ E(\hat{\theta})-\theta \quad E(\hat{\theta})-E(\hat{\theta}) \\ &= V(\hat{\theta}) + \ E(\hat{\theta})-\theta\ ^2 + 0 \\ &= V(\hat{\theta}) + (\text{bias})^2. \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (2.3.5) \end{aligned}$$

The mean square error is an important criterion for comparing two estimators. An estimator $\hat{\theta}_A$ is said to be more efficient than the estimator $\hat{\theta}_B$ if

$$MSE(\hat{\theta}_A) < MSE(\hat{\theta}_B). \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (2.3.6)$$

This result reduces to (2.3.3) if $\hat{\theta}_A$ and $\hat{\theta}_B$ are unbiased.

Relative efficiency

Two competing estimators of a parameter θ, can be compared by finding the relative efficiency of one of the estimators with respect to the other.

Definition 2.6 **(Relative efficiency)**

The relative efficiency of $\hat{\theta}_B$ with respect to $\hat{\theta}_A$ is defined by the ratio

$$eff\,(\hat{\theta}_B, \hat{\theta}_A) = \frac{MSE(\hat{\theta}_A)}{MSE(\hat{\theta}_B)} \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (2.3.7)$$

when this ratio does not depend on θ.

If $\hat{\theta}_B$ and $\hat{\theta}_A$ are unbiased estimators of θ, Equation (2.3.7) reduces to

$$eff\,(\hat{\theta}_B, \hat{\theta}_A) = \frac{V(\hat{\theta}_A)}{V(\hat{\theta}_B)}. \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (2.3.8)$$

If this ratio is less than 1, we would conclude that $\hat{\theta}_A$ is a more efficient estimator of θ than $\hat{\theta}_B$.

Example 2.15

Refer to Example 2.14. Find the mean square error of:

(a) $\tilde{\theta}$, the moment estimator of θ, (b) $\hat{\theta}$, the maximum likelihood estimator of θ.

Which is the better estimator of θ? Explain your choice.

Solution

(a)

$$
\begin{aligned}
MSE(\tilde{\theta}) &= V(\tilde{\theta}) + \quad E(\tilde{\theta})-\theta \ ^{2} \\
&= V(\tilde{\theta}) + 0, \quad \text{(see Equation (2.3.1) on page 59)} \\
&= V(2\bar{X}), \quad \text{(see Equation (2.2.16)on page 54)} \\
&= 4V(\bar{X}) \ = \ 4\sigma^2/n, \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (2.3.9)
\end{aligned}
$$

where $\sigma^2 = V(X)$. Now,

$$
\begin{aligned}
V(X) &= E(X^2) - \quad E(X) \ ^{2} \\
&= \int_0^\theta \frac{x^2}{\theta}dx - \quad \int_0^\theta \frac{x}{\theta}dx \ ^{2} \ = \ \left[\frac{x^3}{3\theta}\right]_0^\theta - \left\{\left[\frac{x^2}{2\theta}\right]_0^\theta\right\}^2 \\
&= \theta^2/12. \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (2.3.10)
\end{aligned}
$$

Using Equations (2.3.9) and (2.3.10), we obtain

$$
MSE(\tilde{\theta}) = \frac{4}{n}\left(\frac{\theta^2}{12}\right) = \frac{\theta^2}{3n}. \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (2.3.11)
$$

(b)

$$
\begin{aligned}
MSE(\hat{\theta}) &= V(\hat{\theta}) + \quad E(\hat{\theta})-\theta \ ^{2} \\
&= V(\hat{\theta}) + \quad \frac{n\theta}{n+1} - \theta \ ^{2}, \quad \text{(see Equation (2.3.2)on page 60)} \\
&= V(\hat{\theta}) + \quad \frac{-\theta}{n+1} \ ^{2}. \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (2.3.12) \\
V(\hat{\theta}) &= E\left[X_{(n)}^2\right] - \quad E\left[X_{(n)}\right] \ ^{2} \\
&= \int_0^\theta \frac{nx^{n+1}}{\theta^n}dx - \quad \frac{n\theta}{n+1} \ ^{2}, \quad \text{(see Equation (2.3.2)on page 60)}
\end{aligned}
$$

$$= \left[\frac{nx^{n+2}}{(n+2)\theta^n}\right]_0^\theta - \frac{n^2\theta^2}{(n+1)^2} = \frac{n\theta^2}{n+2} - \frac{n^2\theta^2}{(n+1)^2}$$

$$= \frac{n\theta^2}{(n+1)^2(n+2)}. \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (2.3.13)$$

Using Equations (2.3.12) and (2.3.13), we obtain

$$MSE(\hat{\theta}) = \frac{n\theta^2}{(n+1)^2(n+2)} + \frac{\theta^2}{(n+1)^2} = \frac{2\theta^2}{(n+1)(n+2)}. \quad \ldots\ldots\ldots\ldots \quad (2.3.14)$$

Using Equations (2.3.11) and (2.3.14), we obtain

$$eff(\tilde{\theta}, \hat{\theta}) = \frac{MSE(\hat{\theta})}{MSE(\tilde{\theta})} = \frac{6n}{(n+1)(n+2)}.$$

It can be seen that $eff(\tilde{\theta}, \hat{\theta}) < 1$ when $n \geq 2$. It follows that $\hat{\theta}$ is more efficient than $\tilde{\theta}$.

2.3.3 Consistency

Definition 2.7

An estimator $\hat{\theta}_n$ of θ, based on *n* observations, is said to be consistent if the probability of making errors of any given size ε, tends to zero as n tends to infinity; that is, if

$$\lim_{n\to\infty} P\left|\hat{\theta}_n - \theta\right| \geq \varepsilon = 0, \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (2.3.15)$$

for any positive ε.

If $\hat{\theta}_n$ is a consistent estimator of θ, then the sampling distribution of $\hat{\theta}_n$ becomes more and more concentrated at the true value of θ (see Fig. 2.3). A consistent estimator of θ is therefore close to θ for large sample sizes. Consistency is a very important property for an estimator to have. It is a poor estimator that does not approach its target as the sample size increases.

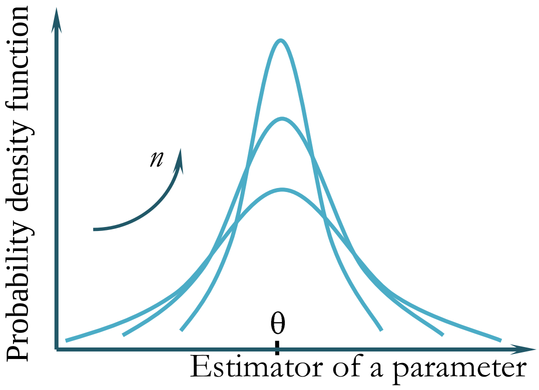

Fig. 2.3: ***Illustration of consistency***

Example 2.16

Suppose X is $N(\mu,\ \sigma^2)$ and consider a random sample of size *n of X*. Let $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} x_i$.

We know that $\bar{X}_n$ is $N\ \ \mu,\ \sigma^2/n$. Therefore for any fixed positive ε,

$$P\ \left|\bar{X}_n - \mu\right| \geq \varepsilon \quad = \quad 1 - P\ \left|\bar{X}_n - \mu\right| \ < \varepsilon$$

$$= \quad 1 - P\ -\varepsilon < \bar{X}_n - \mu < \varepsilon \ = 1 - \left[\Phi\left(\frac{\varepsilon\sqrt{n}}{\sigma}\right) - \Phi\left(-\frac{\varepsilon\sqrt{n}}{\sigma}\right)\right],$$

where Φ denotes the distribution function of the standard normal distribution. As $n \to \infty$, this tends to $1-(1-0)=0$. We have thus shown that $\bar{X}_n$ is consistent as an estimator of the mean of a normal population.

It is not always as easy, as in Example 2.16, to show that Equation (2.3.15) holds in order to establish consistency. The following Theorem gives an easier method for showing that an estimator $\hat{\theta}_n$ is a consistent estimator of θ.

Theorem 2.1

An estimator $\hat{\theta}_n$, based on a sample of size *n*, is consistent for θ if and only if

$$\lim_{n\to\infty} E\ \ \hat{\theta}_n \ \ = \theta \quad \text{and} \quad \lim_{n\to\infty} V\ \ \hat{\theta}_n \ \ = 0.$$

Proof

By Chebyshev's inequality, given ε > 0,

$$P\ \left|\hat{\theta}_n - \theta\right| \geq \varepsilon \ \ \leq \frac{1}{\varepsilon^2} E\left[\ \hat{\theta}_n - \theta\ ^2\right] = \frac{1}{\varepsilon^2}\left[V(\hat{\theta}_n) + \ E(\hat{\theta}_n) - \theta\ ^2\right]. \ \ldots\ldots. \quad (2.3.16)$$

(i) If $\lim_{n\to\infty} V\ \ \hat{\theta}_n \ \ = 0$ and $\lim_{n\to\infty} E\ \ \hat{\theta}_n \ \ = \theta$, then $\lim_{n\to\infty} P\ \left|\hat{\theta}_n - \theta\right| \geq \varepsilon \ \ \leq 0$.

But $\lim_{n\to\infty} P\ \left|\hat{\theta}_n - \theta\right| \geq \varepsilon \ \ \geq 0$. Hence, if $\lim_{n\to\infty} V\ \ \hat{\theta}_n \ \ = 0$ and $\lim_{n\to\infty} E\ \ \hat{\theta}_n \ \ = \theta$, then given ε > 0, $\lim_{n\to\infty} P\ \left|\hat{\theta}_n - \theta\right| \geq \varepsilon \ \ = 0$, proving that $\hat{\theta}_n$ is consistent for θ.

(ii) If $\hat{\theta}_n$ is consistent for θ, then from Equation (2.3.16),

$$\lim_{n\to\infty} V\ \ \hat{\theta}_n \ \ + \ \lim_{n\to\infty}\ \ E(\hat{\theta}_n) - \theta\ ^2 = 0.$$

Since $V\ \hat{\theta}_n \ \geq 0$ and $E(\hat{\theta}_n)-\theta\ ^{2} \geq 0$, if $\hat{\theta}_n$ is consistent, then

$$\lim_{n\to\infty} V\ \hat{\theta}_n\ = 0 \text{ and } \lim_{n\to\infty} E(\hat{\theta}_n) = \theta.$$

Remark

To show that $\hat{\theta}_n$ is a consistent estimator of θ, it is sufficient to show that

$$\lim_{n\to\infty} V\ \hat{\theta}_n\ = 0 \quad \text{and} \quad \lim_{n\to\infty} E\ \hat{\theta}_n\ = \theta.$$

Example 2.17

In Example 2.14, show that the moment and the maximum likelihood estimators of θ are consistent.

Solution

We first consider $\tilde{\theta}$, the moment estimator of θ.

$E(\tilde{\theta}) = \theta$, (see Equation (2.3.1) on page 59).

$V(\tilde{\theta}) = \theta^2/3n$, (see Example 2.15 on page 62).

It follows that $\lim_{n\to\infty} E\ \tilde{\theta}\ = \theta$ and $\lim_{n\to\infty} V\ \tilde{\theta}\ = 0$, proving that $\tilde{\theta}$ is a consistent estimator of θ.

We now consider $\hat{\theta}$, the maximum likelihood estimator of θ.

$E(\hat{\theta}) = n\theta/(n+1)$, (see Equation (2.3.2) on page 60)

$V(\hat{\theta}) = \dfrac{n\theta^2}{(n+1)^2(n+2)}$, (see Equation (2.3.13) on page 63)

It follows that

$$\lim_{n\to\infty} E(\hat{\theta}) = \lim_{n\to\infty} \frac{\theta}{1+\frac{1}{n}} = \theta$$

and $$\lim_{n\to\infty} V(\hat{\theta}) = \lim_{n\to\infty} \frac{\theta^2}{(1+2/n)(n+1)^2} = 0,$$

proving that $\hat{\theta}$ is also a consistent estimator of θ.

Example 2.18

Consider again the Cauchy distribution defined in Example 2.12. The characteristic function of the distribution is

$$\phi_X(t) = \exp -|t| + it\theta \ .$$

If we take a random sample of size n of X, the characteristic function of the sample mean is

$$\phi_{\bar{X}}(t) = \phi_X \ \frac{t}{n} \ ^{n} = \exp -|t/n| + it\theta/n \ ^{n} = \exp -|t| + it\theta = \phi_X(t).$$

It can be seen that the distribution of the sample mean is the same as that of the population, and so the sample mean is not a consistent estimator of θ.

2.3.4 Sufficiency

Let $X_1, X_2, ..., X_n$ be a random sample of size *n* from a population whose distribution depends on an unknown parameter θ. We have learnt that a statistic condenses the *n* random variables into a single random variable. Such condensing is appealing since we would rather work with unidimensional quantities than *n*-dimensional quantities. We are obviously interested to find out if we lost any "information" by this condensing process. The information about the parameter θ is contained in the sample $X_1, X_2, ..., X_n$, so when we say that a statistic loses no information, we mean that it contains all the information about θ that is contained in the sample. Such a statistic is called a "sufficient statistic". We now formalize the definition of a sufficient statistic.

Definition 2.8 (Sufficient statistic)

Let $X_1, X_2, ..., X_n$ be a random sample from a population whose distribution depends on an unknown parameter θ. A statistic $T = T \ X_1, X_2, ..., X_n$ is said to be sufficient for θ if the conditional distribution of $X_1, X_2, ..., X_n$ given $T = t$ does not depend on θ for any value *t* of *T*.

The definition says that a statistic $T = T \ X_1, X_2, ..., X_n$ is sufficient for θ if the conditional distribution of the sample given the value of the statistic, does not depend on θ. The idea is that if you know the value of a sufficient statistic for θ, then the sample values themselves are not needed and can tell you nothing more about θ, and this is true since the distribution of the sample given the sufficient statistic does not depend on θ.

We now give two examples that are illustrative of Definition 2.8.

Example 2.19

Let $X_1, X_2, \ldots, X_n$ denote a random sample from the distribution with probability mass function

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

Show that $T = \sum_{i=1}^{n} X_i$ is a sufficient statistic for θ.

Solution

We are required to show that $P\;\; X_1 = x_1,\, X_2 = x_2,\, \ldots,\;\; X_n = x_n \big| T = t$ does not depend on θ. Now,

$$P\;\; X_1 = x_1,\, X_2 = x_2,\, \ldots,\;\; X_n = x_n \big| T = t \;\; = \;\; \frac{P\left(X_1{=}x_1,\, X_2{=}x_2,\, \ldots,\, X_n{=}x_n,\; \sum_{i=1}^{n} X_i{=}t \right)}{P\left(\sum_{i=1}^{n} X_i{=}t \right)}$$

$$= \frac{P\;\; X_1{=}x_1,\, X_2{=}x_2,\, \ldots,\, X_n{=}t-(x_1+x_2+\ldots+x_{n-1})}{P\left(\sum_{i=1}^{n} X_i{=}t \right)}$$

$$= \frac{P(X_1{=}x_1)\,P(X_2{=}x_2)\ldots P\left(X_n{=}t-\sum_{i=1}^{n-1} x_i \right)}{P\left(\sum_{i=1}^{n} X_i{=}t \right)}, \quad \text{by independence}$$

$$= \frac{\theta^{x_1}(1-\theta)^{1-x_1}\;\theta^{x_2}(1-\theta)^{1-x_2}\;\ldots\;\theta^{\,t-\sum_{i=1}^{n-1} x_i}\;\;(1-\theta)^{\;1-t+\sum_{i=1}^{n-1} x_i}}{P\left(\sum_{i=1}^{n} X_i{=}t \right)}$$

$$= \frac{\theta^t(1-\theta)^{n\,-\,\mathrm{t}}}{P\left(\sum_{i=1}^{n} X_i \;=\; t \right)}.$$

But

$$P\left(\sum_{i=1}^{n} X_i = t \right) \;=\; \begin{cases} \genfrac{}{}{0pt}{}{n}{t}\;\theta^t(1-\theta)^{n-t}, & t = 0,\ 1,\ \ldots,\ n, \\ 0, & \text{elsewhere.} \end{cases}$$

Therefore,

$$P\left(X_1 = x_1,\ X_2 = x_2,\ \ldots,\ X_n = x_n \middle| \sum_{i=1}^{n} X_i = t\right) = \frac{1}{\frac{n}{t}} = \frac{1}{\left(\frac{n}{\sum_{i=1}^{n} x_i}\right)}. \quad \cdots\cdots\cdots \qquad (2.3.17)$$

This conditional probability does not depend on θ. It follows that $T = \sum_{i=1}^{n} X_i$ is sufficient for θ.

Example 2.20

Let $X_1, X_2, \ldots, X_n$ denote a random sample from a distribution with p.d.f.

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

Show that $Y = \sum_{i=1}^{n} X_i$ is a sufficient statistic for θ.

Solution

The joint p.d.f. of $X_1, X_2, \ldots, X_n$ is

$$f_{\mathbf{X}}(x_1, x_2, \ldots, x_n) = \begin{cases} \theta^n e^{-\theta \sum_{i=1}^{n} x_i}, & x_i > 0,\ (i = 1, 2, \ldots n),\ \theta > 0, \\ 0, & \text{elsewhere}. \end{cases}$$

The p.d.f. of Y is

$$f_Y(y) = \begin{cases} \theta^n y^{n-1} e^{-\theta y} / (n-1)!, & y > 0,\ \theta > 0, \\ 0, & \text{elsewhere}. \end{cases}$$

The conditional p.d.f. of $X_1, X_2, \ldots, X_n$ given $Y = y$ is

$$f_{X|Y}(x_1, x_2, \ldots, x_n | y) = \frac{\theta^n e^{-\theta \sum_{i=1}^{n} x_i}}{\left\{\theta^n \left(\sum_{i=1}^{n} x_i\right)^{n-1} e^{-\theta \sum_{i=1}^{n} x_i}\right\} \Big/ (n-1)!}$$

$$= (n-1)! \Big/ \left(\sum_{i=1}^{n} x_i\right)^{n-1}, \qquad x_i > 0,\ (i = 1, 2, \ldots, n).$$

This conditional p.d.f. is independent of θ, and so Y is a sufficient statistic for θ.

Sufficient statistics do not always exist, but when they do, they are usually of lower dimension than $\boldsymbol{X} = (X_1, X_2, ..., X_n)$ and so an effective reduction in the data is achieved without loss of information about θ.

Remark

Let $X_1, X_2, ..., X_n$ denote a random sample of size *n* from a distribution that has p.d.f. or probability mass function $f(x;\, \theta),\ \ \theta \in \Omega$. Let $T = T(X_1, X_2, ..., X_n)$ be a statistic whose p.d.f. or probability mass function is $f_T(t;\, \theta)$. Then *T* is a sufficient statistic for θ if and only if

$$\frac{f(x_1;\, \theta)\, f(x_2;\, \theta)...\, f(x_n;\, \theta)}{f_T(t;\, \theta)} = W(x_1, x_2, ..., x_n), \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (2.3.18)$$

where *W* does not depend on $\theta \in \Omega$.

The factorization criterion

If we are to show by means of Definition 2.8 that a certain statistic *t* is or is not sufficient for a parameter θ, we must first find the p.d.f. or probability mass function of *t*. In many instances, it may be difficult to find the p.d.f. or probability mass function of *t*. This problem can be avoided by using the following factorization theorem which we state without proof.

Theorem 2.2 (The Fisher-Neyman factorization theorem)

A necessary and sufficient condition for a single statistic *t* to be sufficient for a single parameter θ is that the likelihood function factorizes in the form

$$l(\theta) = g(t;\, \theta) h(x_1, x_2, ..., x_n),$$

where $h(x_1, x_2, ..., x_n)$ is nonnegative and does not involve θ and *g* is nonnegative and depends on the observations only through *t*.

A general proof of this theorem is given by Cox and Hinkley (1974), Kendall and Stuart (Vol 2, 1973), and Lindgren (1993).

Theorem 2.2 characterises sufficiency and, as the following examples show, is usually much easier to work with than the definition of sufficiency.

Example 2.21

Let $X_1, X_2, ..., X_n$ denote a random sample of size *n* from the Poisson distribution with an unknown mean μ. Show that $\sum_{i=1}^{n} X_i$ is a sufficient statistic for μ.

Solution

The likelihood function is

$$l(\mu) = \prod_{i=1}^{n} \mu^{x_i} e^{-\mu} \bigg/ x_i\ !$$

$$= \frac{\mu^{\sum_{i=1}^{n} x_i} e^{-n\mu}}{\prod_{i=1}^{n} x_i\ !} = \left(\mu^{\sum_{i=1}^{n} x_i} e^{-n\mu}\right)\left(\frac{1}{\prod_{i=1}^{n} x_i\ !}\right)$$

$$= \mu^{t} e^{-n\mu} \left(1 \bigg/ \prod_{i=1}^{n} x_i\ !\right) = g(t;\ \mu) h(x_1,\ x_2,\ ...,\ x_n),$$

where $t = \sum_{i=1}^{n} x_i$, $g(t;\ \mu) = \mu^t e^{-n\mu}$ and $h(x_1,\ x_2,\ ...,\ x_n) = 1 \big/ \prod_{i=1}^{n} x_i$!. Since $g(t;\ \mu)$ depends on the sample only through t and $h(x_1,\ x_2,\ ...,\ x_n)$ does not depend on μ, by the factorization theorem, $\sum_{i=1}^{n} X_i$ is a sufficient statistic for μ.

Example 2.22

Let $X_1,\ X_2,\ ...,\ X_n$ denote a random sample of size n from a distribution with p.d.f.

$$f(x;\theta) = \begin{cases} 3\theta x^2 e^{-\theta x^3}, & x \geq 0,\ \theta > 0, \\ 0, & \text{elsewhere.} \end{cases}$$

Find a sufficient statistic for θ.

Solution

The likelihood function is

$$l(\theta) = \prod_{i=1}^{n} 3\theta x_i^2 e^{-\theta x_i^3} = 3\theta^{\ n} \left(\prod_{i=1}^{n} x_i^2\right) e^{-\theta \sum_{i=1}^{n} x_i^3}$$

$$= 3\theta^{\ n} e^{-\theta \sum_{i=1}^{n} x_i^3} \left(\prod_{i=1}^{n} x_i^2\right) = 3\theta^{\ n} e^{-\theta t} \left(\prod_{i=1}^{n} x_i^2\right) = g(t;\ \theta) h(x_1,\ x_2,\ ...,\ x_n),$$

where $t = \sum_{i=1}^{n} x_i^3$, $g(t;\ \theta) = 3\theta^{\ n} e^{-\theta t}$ and $h(x_1,\ x_2,\ ...,\ x_n) = \prod_{i=1}^{n} x_i^2$. Since g (t; θ) depends on the sample only through t and $h(x_1, x_2, ..., x_n)$ does not depend on θ, $\sum_{i=1}^{n} X_i^3$ is a sufficient statistic for θ (see Theorem 2.2).

Example 2.23

Let $X_1, X_2, ..., X_n$ denote a random sample from a distribution that is $N(\theta, \sigma^2)$, $-\infty < \theta < \infty$, where σ is known. The likelihood function is

$$l(\theta) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \exp\left[-\sum_{i=1}^{n} x_i - \theta\ {}^2 \Big/ 2\sigma^2\right].$$

But $\sum_{i=1}^{n} x_i - \theta\ {}^2 = \sum_{i=1}^{n}\left[\ x_i - \bar{x}\ +\ \bar{x} - \theta\ \right]^2$

$$= \sum_{i=1}^{n} x_i - \bar{x}\ {}^2 + n\ \bar{x} - \theta\ {}^2,$$

because $2\sum_{i=1}^{n} x_i - \bar{x}\ \ \bar{x} - \theta\ = 2\ \bar{x} - \theta\ \sum_{i=1}^{n} x_i - \bar{x}\ = 0.$

Thus,

$$l(\theta) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \exp\left[-\left\{\sum_{i=1}^{n} x_i - \bar{x}\ {}^2 + n(\bar{x} - \theta)^2\right\} \Big/ 2\sigma^2\right]$$

$$= \left\{\exp\left[\frac{-n\ \bar{x} - \theta\ {}^2}{2\sigma^2}\right]\right\}\left\{\frac{\exp\left[-\sum_{i=1}^{n} x_i - \bar{x}\ {}^2 \Big/ 2\sigma^2\right]}{\sigma\sqrt{2\pi}\ {}^n}\right\}. \quad \text{.............} \quad (2.3.19)$$

Because the first factor of the right-hand side of this equation depends on $x_1, x_2, ..., x_n$, only through $\bar{x}$, and the second factor does not depend on θ, the factorization theorem implies that $\bar{X}$ is a sufficient statistic for θ.

We have learnt that a sufficient statistic for θ contains all the information that is in the sample about θ. We might want to use a function $W\ T(\boldsymbol{X})$ of a sufficient statistic $T(\boldsymbol{X})$ as an estimator of θ. It is therefore of interest to know the conditions under which $W\ T(\boldsymbol{X})$ is sufficient for θ. The result is given in the following theorem.

Theorem 2.3

If t is a sufficient statistic for θ, then any one-to-one function of t is also sufficient for θ.

Proof

If t is a sufficient statistic for θ, then by the factorization theorem,

$$l(\theta) = g(t;\, \theta)h(x_1, x_2, ..., x_n),$$

where $h(x_1, x_2, ..., x_n)$ does not depend on θ. Let $y = w(t)$ be a one-to-one function of t. Then, $t = w^{-1}(y)$. It follows that

$$l(\theta) = g\;\; w^{-1}(y);\, \theta\;\; h\;\; x_1, x_2, ..., x_n\;\; .$$

Hence, by the factorization theorem, Y is also sufficient for θ.

Example 2.24

Let $X_1, X_2, ..., X_n$ denote a random sample of size n from the beta distribution with p.d.f.

$$f(x;\, \lambda) = \begin{cases} \lambda x^{\lambda-1}, & 0 < x < 1,\ \lambda > 0, \\ 0, & \text{elsewhere.} \end{cases}$$

Show that $\sum_{i=1}^{n} \ln X_i$ is a sufficient statistic for λ.

Solution

The likelihood function is given by

$$\begin{aligned} l(\lambda) &= \prod_{i=1}^{n} f(x_i;\, \lambda) = \lambda^n \left(\prod_{i=1}^{n} x_i \right)^{\lambda-1} \\ &= \lambda^n \left(\prod_{i=1}^{n} x_i \right)^{\lambda} \left(\frac{1}{\prod_{i=1}^{n} x_i} \right) = g(T;\, \lambda)h(x_1, x_2, ..., x_n), \end{aligned}$$

where $T = \prod_{i=1}^{n} x_i$, $g(T;\, \lambda) = \lambda^n T^{\lambda}$ and $h(x_1, x_2,..., x_n) = 1 \Big/ \prod_{i=1}^{n} x_i$. Since $g(T;\, \lambda)$ depends on the sample only through T and $h(x_1, x_2,..., x_n)$ does not depend on λ, $T = \prod_{i=1}^{n} X_i$ is sufficient for λ. $\sum_{i=1}^{n} \ln X_i$ is a one-to-one function of $\prod_{i=1}^{n} X_i$, and so $\sum_{i=1}^{n} \ln X_i$ is also sufficient for λ.

Sufficiency when the range of X depends on the unknown parameter θ

In the following examples, we illustrate how to find a sufficient statistic for θ when the range of X depends on the unknown parameter θ.

Example 2.25

In Example 2.5, find a sufficient statistic for θ.

Solution

The likelihood function can be expressed in the form (see Equation (2.2.7) on page 50)

$$l(\theta) = \frac{1}{\theta^n} w(\theta - t), \quad \text{where } t = \max_{1 \le i \le n}(x_i) \text{ and } w(z) = \begin{cases} 1, & z \ge 0, \\ 0, & z < 0. \end{cases}$$

Thus, $\qquad l(\theta) = g(t;\, \theta) h(x_1, x_2, \ldots, x_n)$,

where $g(t;\, \theta) = \frac{1}{\theta^n} w(\theta - t)$ and $h(x_1, x_2, \ldots, x_n) \equiv 1$. Since g(t; θ) depends on the sample only through t, and $h(x_1, x_2, \ldots, x_n)$ does not depend on θ, $t = X_{(n)}$ is sufficient for θ.

Example 2.26

Let $X_1, X_2, \ldots, X_n$ denote a random sample from a distribution with p.d.f.

$$f(x;\, \theta) = \begin{cases} \dfrac{g(x)}{h(\theta)}, & 0 < x \le b(\theta), \\ 0, & \text{elsewhere,} \end{cases}$$

where $\int_0^{b(\theta)} g(x)dx = h(\theta)$ and where θ is an unknown parameter and $b(\theta)$ is monotone in θ. The likelihood function is

$$l(\theta) = \begin{cases} \left\{ \prod_{i=1}^{n} g(x_i) \right\} \Big/ \ h(\theta)^{\ n}, & 0 < x_i \le b(\theta),\ i = 1, 2, \ldots, n, \\ 0, & \text{elsewhere.} \end{cases}$$

$$= \begin{cases} \left\{ \prod_{i=1}^{n} g(x_i) \right\} \Big/ \ h(\theta) \ ^{n}, & x_{(n)} \le b(\theta), \\ 0, & x_{(n)} > b(\theta), \end{cases}$$

where $x_{(n)} = \max\limits_{1 \le i \le n} (x_i)$. Let $t = x_{(n)}$. Then,

$$l(\theta) = \frac{w \ b(\theta) - t}{h(\theta) \ ^{n}} \prod_{i=1}^{n} g(x_i), \quad \text{where } w(z) = \begin{matrix} 1, & z \ge 0, \\ 0, & z < 0. \end{matrix}$$

Thus, $l(\theta) = g(t;\ \theta) h(x_1, x_2, ..., x_n)$,

where $g(t;\ \theta) = \dfrac{w \ b(\theta) - t}{h(\theta) \ ^{n}}$ and $h(x_1, x_2, ..., x_n) = \prod\limits_{i=1}^{n} g(x_i)$. By the factorization theorem, $X_{(n)}$ is sufficient for θ.

Example 2.27

In Example 2.26, let

$$f(x;\ \theta) = \begin{cases} \dfrac{g(x)}{h(\theta)}, & a(\theta) \le x < \infty, \\ 0, & \text{elsewhere,} \end{cases}$$

where θ is an unknown parameter and $a(\theta)$ is monotone in θ and where

$$\int_{a(\theta)}^{\infty} g(x) dx = h(\theta).$$

The likelihood function is given by

$$l(\theta) = \begin{cases} \left\{ \prod_{i=1}^{n} g(x_i) \right\} \Big/ \ h(\theta) \ ^{n}, & x_i \ge a(\theta),\ i = 1,\ 2,\ ...,\ n, \\ 0, & \text{elsewhere.} \end{cases}$$

$$= \begin{cases} \left\{ \prod_{i=1}^{n} g(x_i) \right\} \Big/ \ h(\theta) \ ^{n}, & x_{(1)} \ge a(\theta), \\ 0, & x_{(1)} < a(\theta), \end{cases}$$

where $x_{(1)} = \min\limits_{1 \le i \le n} (x_i)$. If we let $t = x_{(1)}$, then

$$l(\theta) = \frac{w \ t - a(\theta)}{h(\theta) \ ^{n}} \prod_{i=1}^{n} g(x_i), \quad \text{where } w(z) = \begin{matrix} 1, & z \ge 0, \\ 0, & z < 0. \end{matrix}$$

Thus, $l(\theta) = g(t;\ \theta) h(x_1, x_2, ..., x_n)$, where $g(t;\,\theta) = \frac{w\ t\ -\ a(\theta)}{h(\theta)^{\ n}}$, and $h(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} g(x_i)$. By the factorization theorem, $X_{(1)}$ is sufficient for θ.

Jointly sufficient statistics

The concept of sufficiency generalizes, and if we have many unknown parameters, we may have a set of statistics which are jointly sufficient. An obvious extension of the factorization theorem then applies.

A much fuller discussion of sufficiency is given, for example, by Cox and Hinkley (1974), Cramér (1952), and Wilks (1962).

Exercise 2(b)

1. Let $X_1, X_2, \ldots, X_n$ be a random sample of size n from a population with variance σ^2.

   (a) Show that $\frac{1}{n}\sum_{i=1}^{n} X_i - \bar{X}^{\ 2}$ is a biased estimator of σ^2.

   (b) Find the amount of bias in the estimator.

   (c) What happens to the bias as *n* increases?

2. Let $Y_1, Y_2, \ldots, Y_n$ be a random sample from a population with mean μ and variance σ^2.

   (a) Under what conditions is $W = \sum_{i=1}^{n} a_i Y_i$ an unbiased estimator of μ?

   (b) The following $(n+1)$ estimators are proposed for μ. $W_1 = \ Y_1 + Y_2 + \ldots + Y_n \ /n$, $W_{i+1} = 2Y_i - Y_1 \quad i = 1,\ 2,\ \ldots,\ n$. Show that all $(n+1)$ estimators are unbiased and determine the best amongst these $(n + 1)$ estimators.

3. Let $X_1, X_2, \ldots, X_n$ be a random sample of size *n* from a population with p.d.f.

   $$f(x;\lambda) = ce^{-x/\lambda}, \quad x > 0,\ \ \lambda > 0.$$

   (a) Find the value of the constant *c*.

   (b) Find the maximum likelihood estimator of λ and show that it is unbiased and consistent.

4. Let $X_1, X_2, ..., X_n$ be a random sample of size n of X, where X has the p.d.f.

$$f(x;\ \theta) = 2x/\theta^2, \quad 0 \le x \le \theta.$$

(a) Find the maximum likelihood estimator of θ and show that it is sufficient, biased and consistent.

(b) Find the method of moments estimator of θ.

5. Let $X_1, X_2, ..., X_n$ be a random sample of size n of X, where X has the p.d.f.

$$f(x;\ \theta) = \frac{x}{\theta} e^{-x^2/2\theta}, \quad x \ge 0, \ \ \theta > 0,$$

where θ is an unknown parameter.

(a) Show that X^2 has the exponential distribution.

(b) Find a sufficient statistic for θ.

(c) Find the maximum likelihood estimator of θ and show that it is unbiased and consistent.

6. Let $X_1, X_2, ..., X_n$ be a random sample of size n of the random variable X with p.d.f.

$$f(x;\ \mu) = 2x\mu e^{-\mu x^2}, \quad x > 0, \quad \mu > 0,$$

where μ is an unknown parameter. Let $W = \frac{1}{n}\sum_{i=1}^{n} X_i^2$. Show that

(a) W is sufficient for μ, (b) W is unbiased and consistent for $1/\mu$.

7. Refer to Question 4. Let a_n be the unbiased estimator of θ, based on the maximum likelihood estimator of θ, and let b_n denote the moment estimator of θ. Show that the efficiency of b_n relative to a_n is $2/(n+1)$.

8. Let X be a Bernoulli random variable with probability mass function

$$P(X = x) = \begin{cases} p^x(1-p)^{1-x}, & x = 0,\ 1, \ \ 0 < p < 1, \\ 0, & \text{otherwise,} \end{cases}$$

where p is an unknown parameter.

(a) Find the maximum likelihood estimator of p and show that it is unbiased and consistent.

(b) Find the moment estimator of p.

9. A random variable Y has the geometric distribution with an unknown parameter p. That is, $P(Y = y) = pq^y, \quad y = 0, 1, 2, ...,$ where $p + q = 1$.

(a) Find the moment estimator of p.

(b) Find the maximum likelihood estimator of p.

10. Suppose that $\hat{\theta}_1, \hat{\theta}_2$ and $\hat{\theta}_3$ are estimators of θ such that, $E(\hat{\theta}_1) = E(\hat{\theta}_2) = \theta$, $E(\hat{\theta}_3) \neq \theta, V(\hat{\theta}_1) = 9, V(\hat{\theta}_2) = 7$ and $E(\hat{\theta}_3 - \theta)^2 = 6$. Compare these three estimators. Which do you prefer? Why?

11. Let $X_1, X_2, ..., X_7$ denote a random sample from a population with mean μ and variance σ^2. Consider the following estimators of μ.

$$\hat{\theta}_1 = \ X_1 \ + X_2 \ + ... + X_7 \ /7, \quad \hat{\theta}_2 = \ 2X_1 \ - X_6 \ + X_4 \ /2.$$

(a) Is either estimator unbiased? (b) Which is the better estimator of μ? Why?

12. Let $X_1, X_2, ..., X_n$ denote a random sample of size n from a distribution which is $N(0, \theta)$, $0 < \theta < \infty$. Show that $\sum_{i=1}^{n} X_i^2$ is a sufficient statistic for θ.

13. Let $X_1, X_2, ..., X_n$ denote a random sample of size n from the Poisson distribution with an unknown mean μ, and let $T = \sum_{i=1}^{n} X_i$.

(a) Show that $P \ X_1 = x_1, X_2 = x_2, ..., \ X_n = x_n \middle| T = t \ = \dfrac{t!}{x_1!\, x_2! ... x_n!} \ \dfrac{1}{n}^{\ t}$, which is a multinomial distribution, and independent of μ. By definition 2.8, T is sufficient for μ.

(b) Use the factorization theorem to show that T is sufficient for μ.

14. Let $X_1, X_2, ..., X_n$ be a random sample of size n from a geometric distribution with probability mass function

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

Show that $\sum_{i=1}^{n} X_i$ is sufficient for θ.

15. Let $X_1, X_2, ..., X_n$ be a random sample of size n from the beta distribution with p.d.f.

$$f(x) \ = \ \frac{1}{\beta(\theta,\ 2)} x^{\theta - 1}(1-x), \qquad 0 < x < 1,\ \theta > 0.$$

Show that $\sum_{i=1}^{n} \ln X_i$ is a sufficient statistic for θ.

16. Let $X_1, X_2, ..., X_n$ denote a random sample of size *n* from a population with p.d.f.

$$f(x;\ \theta) = \begin{cases} 3x^2/\theta^3, & 0 < x < \theta, \quad \theta > 0, \\ 0, & \text{elsewhere,} \end{cases}$$

where θ is an unknown parameter.

(a) Find the maximum likelihood estimator of θ and show that it is sufficient, biased and consistent.

(b) Find the moment estimator of θ.

(c) Let a_n be the unbiased estimator of θ, based on the maximum likelihood estimator of θ, and let b_n denote the moment estimator of θ. Show that the efficiency of b_n relative to a_n is $5/(3n + 2)$.

17. Let $X_1, X_2, ..., X_n$ denote a random sample from the distribution with probability mass function

$$f(x;\ \theta) = \begin{cases} \theta^x(1-\theta)^{1-x}, & x = 0,\ 1, \quad 0 < \theta < 1, \\ 0, & \text{otherwise.} \end{cases}$$

If $Y = \sum_{i=1}^{n} X_i$, show that $P\ \ X_1 = x_1,\ X_2 = x_2,\ ...,\ \ X_n = x_n | Y = y\ \ = \frac{1}{\genfrac{}{}{0pt}{}{n}{y}}$.

18. Let $X_1, X_2, ..., X_n$ denote a random sample from a distribution with p.d.f.

$$f(x;\ \theta) = \begin{cases} e^{-(x-\theta)}, & x \geq \theta, \\ 0, & \text{elsewhere.} \end{cases}$$

Show that $\min_{1 \leq i \leq n} (X_i)$ is a sufficient statistic for θ.

19. Let $X_1, X_2, ..., X_n$ denote a random sample from the gamma distribution with p.d.f.

$$\alpha(\alpha x)^{\beta-1} \exp(-\alpha x)/\Gamma(\beta), \quad x > 0, \quad \alpha > 0, \quad \beta > 0.$$

Find:

(a) a sufficient statistic for α if β is known,

(b) a sufficient statistic for β if α is known.

20. Find the relative efficiency of $T_1 = X_1 + 2X_2 /3$ with respect to the mean, $T_2 = X_1 + X_2 /2$, when X_1 and X_2 are i.i.d., each with finite variance σ^2.

21. Referring to Example 2.5, estimating the parameter θ in a uniform distribution over the interval (0, θ):
    (a) find the relative efficiency of $2\bar{X}$, which is unbiased, with respect to the unbiased estimator $\frac{n+1}{n} X_{(n)}$.
    (b) find the relative efficiency of $2\bar{X}$ with respect to $2\tilde{X}$, where $\tilde{X}$ is the sample median. (Assume that n is odd).

22. In Example 2.15, estimating the parameter θ in a uniform distribution over the interval (0, θ), we found the mean and variance of $X_{(n)}$ to be

$$E\left[X_{(n)}\right] = \frac{n\theta}{n+1}, \quad V\left[X_{(n)}\right] = \frac{n\theta^2}{(n+1)^2(n+2)}.$$

    (a) Find a multiple of $\bar{X}$ that is an unbiased estimator of θ, and find is mean square error.
    (b) Find a multiple of $X_{(n)}$ that is an unbiased estimator of θ and find its mean square error.

23. If X has the exponential distribution with mean 1/λ, then a natural candidate for estimating λ from a random sample of size n is $\hat{\lambda} = 1/\bar{X}$.
    (a) Calculate $E(\hat{\lambda})$ when $n > 1$.
    (b) Find an unbiased estimator of λ and calculate its mean square error.
    (c) Show that a multiple of $1/\bar{X}$ with the smallest mean square error in estimating λ is $(n-2)/\sum X_i$.

24. Refer to Exercise 2(a), Question 5. Determine whether the maximum likelihood and the method of moments estimators of θ are biased.

References

Barnard, G. A., Jenkins, G. M., and Winsten, C. B. (1962). Likelihood inference and time series (with discussions), *J. R. Stat. Soc. A*, **125(3),** 321 – 372.

Cox, D. R. and Hinkley, D. V. (1974). Theoretical Statistics. *Chapman and Hall, London.*

Cramér, H. (1952). Mathematical Methods of Statistics. *Princeton U. P.*

Kendall, M. G. and Stuart, A. (1973). The Advanced Theory of Statistics. Vol. 2, Inference and Relationship, 3rd edition. *Griffin, London.*

Lindgren, B. W. (1993). Statistical Theory, 4th edition. *Chapman and Hall, London.*

Mood, A. M., Graybill, F. A. and Boes, D. (1974). Introduction to the theory of Statistics 3rd edition. *McGraw-Hill Book Company.*

Norden, R. H. (1972). A survey of maximum likelihood estimation, Part 1, *Int. Stat. Rev.* **40(3)**, 329-354.

Norden, R. H. (1973). A survey of maximum likelihood estimation, Part 2, *Int. Stat. Rev.* **41(1)**, 329-354.

Ofosu, J. B. and Hesse, C. A. (2011). Introduction to Probability and Probability Distributions. *E P P Books Services, Accra.*

Sprott, D. A. and Kalbefleisch, J. D. (1969). Examples of likelihoods and comparison with point estimates and large sample approximations. *J. Amer. Stat. Ass.* **64**, 486-484.

Wilks, S. S. (1962). Mathematical Statistics, *John Wiley and Sons Ltd., New York.*