Chapter 5

Statistical estimation

Chapter outline

Objective

In this chapter we study some statistical methods to find estimators of population parameters and study their properties. This will include methods of finding a point estimation as well as an interval estimation of the unknown population parameters.

Mathematical Statistics with Applications in R. https://doi.org/10.1016/B978-0-12-817815-7.00005-1

C. R. Rao
(Source: https://news.psu.edu/story/160566/2011/02/18/academics/cr-rao-receives-33rd-honorary-doctoral-degree).

Calyampudi Radhakrishna (C. R.) Rao (1920−) is a contemporary statistician whose work has influenced not just statistics, but such diverse fields as anthropology, biometry, demography, economics, genetics, geology, and medicine. Several statistical terms and equations are named after Rao. He has worked with many other famous statisticians such as Blackwell, Fisher, and Neyman and has had dozens of theorems named after him. Rao earned an MA in mathematics and another MA in statistics, both in India, and earned his PhD and ScD at Cambridge University. The following was stated in the preface to the 1991 special issue of the *Journal of Quantitative Economics* in Rao's honor: "Dr. Rao is a very distinguished scientist and a highly eminent statistician of our time. His contributions to statistical theory and applications are well known, and many of his results, which bear his name, are included in the curriculum of courses in statistics at bachelor's and master's level all over the world. He is an inspiring teacher and has guided the research work of numerous students in all areas of statistics. His early work had greatly influenced the course of statistical research during the last four decades. One of the purposes of this special issue is to recognize Dr. Rao's own contributions to econometrics and acknowledge his major role in the development of econometric research in India." The importance of statistics can be summarized in Rao's own words: "If there is a problem to be solved, seek statistical advice instead of appointing a committee of experts. Statistics can throw more light than the collective wisdom of the articulate few" http://www.finse.uio.no/events/international-workshops/introduction-to-estimation/.

5.1 Introduction

In statistical analysis, the estimation of a population's parameters plays a very significant role. In most applied problems, a certain numerical characteristic of the physical phenomenon may be of interest; however, its value may not be observable directly. Instead, suppose it is possible to observe one or more random variables, the distribution of which depends on the characteristic of interest. Our objective will be to develop methods that use the observed values of random variables (sample data) to gain information about the unknown and unobservable characteristic of the population.

In studying a real-world phenomenon, we begin with a random sample of size n taken from the totality of a population. In estimation theory, it is assumed the observations are random with a probability distribution dependent on some parameters of interest. The initial step in statistically analyzing these data is to be able to identify the probability distribution that characterizes this information. Since the parameters of a distribution are its defining characteristics, it becomes necessary to know the parameters. In the present chapter, we shall assume that the form of the population distribution is known (such as binomial, normal, etc.) but the parameters of the distribution (p for a binomial, μ and σ^2 for a normal, etc.) are unknown. We shall estimate these parameters using the data from our random sample. It is extremely important to have the best possible estimate of the population parameter(s). Having such estimates will lead to a better and more accurate statistical analysis.

For example, for phosphate mining in Florida, we may be interested in estimating the average radioactivity from both uranium and radium in a clay settling area of a mining site. Suppose that a random sample of 10 such sites resulted in a sample average of 40 pCi/g (picocuries/gram) of radioactivity. We may use this value as an estimate of the average radioactivity for all of the settling areas of mining sites in Florida. We may also want to know a range of values of radioactivity with certain confidence. Since many Florida crops are grown on clay settling areas, these types of estimates are important for assessing the risks associated with radioactivity ingested by eating food from the crops grown on these clay settling areas.

There are two types of estimators, namely, point estimator and interval estimator. First, we will introduce statistical point estimation methods, discuss their properties, and illustrate their usefulness with a number of applications. Point estimation gives a single "best guess" for the parameter(s) of interest. The importance of point estimates lies in the fact that many statistical formulas are based on them. For example, the point estimates of mean and standard deviation are needed in the calculation of confidence intervals (CIs) and in many formulas for hypothesis testing. These topics will be covered subsequently. In general, the point estimates will differ from the true parameter values by varying amounts depending on the sample values obtained. In addition, the point estimates do not convey any measure of reliability. To deal with these issues, we will also introduce so-called interval estimation or CIs.

5.2 The methods of finding point estimators

Let $X_1, \ldots, X_n$ be independent and identically distributed (iid) random variables (in statistical language, a random sample) with a probability density function (pdf) or probability mass function (pmf) $f(x, \theta_1, \ldots, \theta_l)$, where $\theta_1, \ldots, \theta_l$ are the unknown population parameters (characteristics of interest). For example, a normal pdf has parameters μ (the mean) and σ^2 (the variance). The actual values of these parameters are not known. The problem in point estimation is to determine statistics $g_i(X_1, \ldots, X_n)$, $i = 1, \ldots, l$, which can be used to estimate the value of each of the parameters—that is, to assign an appropriate value for the parameters $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_l)$ based on observed sample data from the population. These statistics are called estimators for the parameters, and the values calculated from these statistics using particular sample data values are called estimates of the parameters. Estimators of θ_i are denoted by $\widehat{\theta}_i$, where $\widehat{\theta}_i = g_i(X_1, \ldots, X_n), i = 1, \ldots l$. Observe that the estimators are random variables. As a result, an estimator has a distribution (which we called the sampling distribution in Chapter 4). When we actually run the experiment and observe the data, let the observed values of the random variables $X_1, \ldots, X_n$ be $x_1, \ldots, x_n$; then, $\widehat{\theta}(X_1, \ldots, X_n)$ is an estimator, and its value $\widehat{\theta}(x_1, \ldots, x_n)$ is an estimate. For example, in case of the normal distribution, the parameters of interest are $\theta_1 = \mu$, and $\theta_2 = \sigma^2$, that is, $\boldsymbol{\theta} = (\mu, \sigma^2)$. If the estimators of μ and σ^2 are $\overline{X} = (1/n)\sum_{i=1}^{n} X_i$ and $S^2 = (1/(n-1))\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2$, respectively, then the corresponding estimates are $\overline{x} = (1/n)\sum_{i=1}^{n} x_i$ and $s^2 = (1/n-1)\sum_{i=1}^{n}(x_i - \overline{x})^2$, the mean and variance corresponding to the particular observed sample values. In this book, we use capital letters such as $\overline{X}$ and S^2 to represent the estimators, and lowercase letters such as $\overline{x}$ and s^2 to represent the estimates.

There are many methods available for estimating the true value(s) of the parameter(s) of interest. Three of the more popular methods of estimation are the method of moments, the method of maximum likelihood, and Bayes' method. A very popular procedure among econometricians to find a point estimator is the generalized method of moments. In this chapter we study only the method of moments and the method of maximum likelihood for obtaining point estimators and some of their desirable properties. In Chapter 10, we shall discuss Bayes' method of estimation.

There are many criteria for choosing a desired point estimator. Heuristically, some of them can be explained as follows. An estimator, $\widehat{\theta}$, is unbiased if the mean of its sampling distribution is the parameter θ. The bias of $\widehat{\theta}$ is given by $B = E\left(\widehat{\theta}\right) - \theta$. The estimator has the sufficiency property if it fully uses all the sample information. Minimal sufficient statistics are those that are sufficient for the parameter and are functions of every other set of sufficient statistics for those same parameters. A method attributable to Lehmann and Scheffé can be used to find a minimal sufficient statistic. In addition, the estimator is said to satisfy the consistency property if the sample estimator has a high probability of being close to the population value θ for a large sample size. The concept of efficiency is based on comparing variances of the different unbiased estimators. If there are two unbiased estimators, it is desirable to have the one with the smaller variance. However, some of these properties will not be discussed in this book.

How do we find a good point estimator with desirable properties? To answer this question, we will study two methods of finding point estimators, namely, the method of moments and the method of maximum likelihood.

5.2.1 The method of moments

One of the oldest methods for finding point estimators is the method of moments. This is a very simple procedure for finding an estimator for one or more population parameters. Let $\mu_k' = E[X^k]$ be the kth moment about the origin of a random variable X, whenever it exists. Let $m'_k = (1/n)\Sigma_{i=1}^{n} X_i^k$ be the corresponding kth sample moment. Then, the estimator of μ'_k by the method of moments is m'_k. The method of moments is based on matching the sample moments with the corresponding population (distribution) moments and is founded on the assumption that sample moments should provide good estimates of the corresponding population moments. Because the population moments $\mu'_k = h_k(\theta_1, \theta_2, ..., \theta_l)$ are often functions of the population parameters, we can equate corresponding population and sample moments and solve for these parameters in terms of the moments.

Method of moments

Choose as estimates those values of the population parameters that are solutions of the equations $\mu'_k = m'_k, k = 1, 2, ..., l.$ Here μ'_k is a function of the population parameters.

For example, the first population moment is $\mu'_1 = E(X)$, and the first sample moment is $\overline{X} = \sum_{i=1}^{n} X_i/n$. Hence, the moment estimator of μ'_1 is $\overline{X}$. If $k = 2$, then the second population and sample moments are $\mu'_2 = E(X^2)$ and $m'_2 = (1/n)\sum_{i=1}^{n} X_i^2$, respectively. Basically, we can use the following procedure to find point estimators of the population parameters using the method of moments.

The method of moments procedure

Suppose there are l parameters to be estimated, say $\theta = (\theta_1, ..., \theta_l)$.

1. Find l population moments, $\mu'_k, k = 1, 2, ..., l$. μ'_k will contain one or more parameters $\theta_1, ..., \theta_l$.
2. Find the corresponding l sample moments, $m'_k, k = 1, 2, ..., l$. The number of sample moments should equal the number of parameters to be estimated.
3. From the system of equations, $\mu'_k = m'_k,\ k = 1, 2, ..., l$, solve for the parameter $\theta = (\theta_1, ..., \theta_l)$; this will be a moment estimator of $\widehat{\theta}$.

The following examples illustrate the method of moments for population parameter estimation.

EXAMPLE 5.2.1

Let $X_1, ..., X_n$ be a random sample from a Bernoulli population with parameter p.

(a) Find the moment estimator for p.

(b) Tossing a coin 10 times and equating heads to value 1 and tails to value 0, we obtained the following values:

0 1 1 0 1 0 1 1 1 0

Obtain a moment estimate for p, the probability of success (head).

Solution

(a) *For the Bernoulli random variable,* $\mu'_k = E[X] = p$, *so we can use* m'_1 *to estimate* p. *Thus,*

$$m'_1 = \widehat{p} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

Let

$$Y = \sum_{i=1}^{n} X_i.$$

Then, the method of moments estimator for p *is* $\widehat{p} = Y/n$. *That is, the ratio of the total number of heads to the total number of tosses will be an estimate of the probability of success.*

(b) *Note that this experiment results in Bernoulli random variables. Thus, using* (a) *with* $Y = 6$, *we get the moment estimate of* p *as* $\widehat{p} = \frac{6}{10} = 0.6$.

We would use this value $\hat{p} = 0.6$, *to answer any probabilistic questions for the given problem. For example, what is the probability of obtaining exactly 8 heads out of 10 tosses of this coin? This can be obtained by using the binomial formula (or R-command: pbinom(8,10, 0.6)-pbinom(7,10, 0.6)), with* $\hat{p} = 0.6$, *that is,*

$$P(X = 8) = \binom{10}{8}(0.6)^8(0.4)^{10-8} = 0.1209324.$$

In Example 5.2.1, we used the method of moments to find a single parameter. We demonstrate in Example 5.2.2 how this method is used for estimating more than one parameter.

EXAMPLE 5.2.2

Let $X_1, \ldots, X_n$ *be a random sample from a gamma probability distribution with parameters* α *and* β. *Find moment estimators for the unknown parameters* α *and* β.

Solution

For the gamma distribution (see Section 3.2.5),

$$E[X] = \alpha\beta \quad \text{and} \quad E\left[X^2\right] = \alpha\beta^2 + \alpha^2\beta^2.$$

Because there are two parameters, we need to find the first two moment estimators. Equating sample moments to distribution (theoretical) moments, we have:

$$\frac{1}{n}\sum_{i=1}^{n} X_i = \overline{X} = \alpha\beta, \quad \text{and} \quad \frac{1}{n}\sum_{i=1}^{n} X_i^2 = \alpha\beta^2 + \alpha^2\beta^2.$$

Solving for α *and* β *we obtain the estimates as* $\alpha = (\overline{x}/\beta)$ *and* $\beta = \left[\left\{(1/n)\Sigma_{i=1}^{n} x_i^2 - \overline{x}^2\right\}/\overline{x}\right]$
Therefore, the method of moments estimators for α *and* β *are:*

$$\hat{\alpha} = \frac{\overline{X}}{\hat{\beta}}$$

and

$$\hat{\beta} = \frac{\frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X}^2}{\overline{X}} = \frac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}{n\overline{X}},$$

which implies that:

$$\hat{\alpha} = \frac{\overline{X}}{\hat{\beta}} = \frac{\overline{X}^2}{\frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X}^2} = \frac{\overline{X}^2}{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}.$$

Thus, we can use these values in the gamma pdf to answer questions concerning the probabilistic behavior of the random variable X.

The following example shows that once we find the moments estimator theoretically, the estimate can be obtained by simply substituting a sample statistic into the formula.

EXAMPLE 5.2.3

Let the distribution of X be $N(\mu, \sigma^2)$.

(a) For a given sample of size n, use the method of moments to estimate μ and σ^2.

(b) The following data (rounded to the third decimal digit) were generated using Minitab from a normal distribution with mean 2 and standard deviation of 1.5:

3.163	1.883	3.252	3.716	−0.049	−0.653	0.057	2.987
4.098	1.670	1.396	2.332	1.838	3.024	2.706	0.231
3.830	3.349	−0.230	1.496				

Obtain the method of moments estimates of the true mean and the true variance.

Solution

(a) *For the normal distribution,* $E(X) = \mu$, *and because* $Var(X) = EX^2 - \mu^2$, *we have the second moment as* $E(X^2) = \sigma^2 + \mu^2$. *Equating sample moments to distribution moments we have:*

$$\frac{1}{n}\sum_{i=1}^{n} X_i = \mu_1' = \mu$$

and

$$\mu_2' = \frac{1}{n}\sum_{i=1}^{n} X_i^2 = \sigma^2 + \mu^2.$$

Solving for μ *and* σ^2, *we obtain the moment estimators as:*

$$\widehat{\mu} = \overline{X}$$

and

$$\widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X}^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2.$$

(b) *Because we know that the estimator of the mean is* $\widehat{\mu} = \overline{X}$ *and the estimator of the variance is* $\widehat{\sigma}^2 = (1/n)\Sigma_{i=1}^{n} X_i^2 - \overline{X}^2$, *from the data the estimates are* $\widehat{\mu} = 2.005$, *and* $\widehat{\sigma}^2 = 6.12 - (2.005)^2 = 2.1$. *Notice that the true mean is* 2 *and the true variance is* 2.25, *which we used to simulate the data.*

In general, using the population pdf we evaluate the lower order moments, finding expressions for the moments in terms of the corresponding parameters. Once we have population (theoretical) moments, we equate them to the corresponding sample moments to obtain the moment estimators.

EXAMPLE 5.2.4

Let $X_1, \ldots, X_n$ be a random sample from a uniform distribution on the interval $[a, b]$. Obtain method of moment estimators for a and b.

Solution

Here, a *and* b *are treated as parameters. That is, we know only that the sample comes from a uniform distribution on some interval, but we do not know from which interval. Our interest is to estimate this interval. The pdf of a uniform distribution is:*

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

Hence, the first two population moments are:

$$\mu_1 = E(X) = \int_a^b \frac{x}{b-a}\,dx = \frac{a+b}{2} \quad \text{and} \quad \mu_2 = E\left(X^2\right) = \int_a^b \frac{x^2}{b-a}\,dx = \frac{a^2+ab+b^2}{3}.$$

The corresponding sample moments are:

$$\widehat{\mu}_1 = \overline{X} \quad \text{and} \quad \widehat{\mu}_2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2.$$

Equating the first two sample moments to the corresponding population moments, we have:

$$\widehat{\mu}_1 = \frac{a+b}{2} \quad \text{and} \quad \widehat{\mu}_2 = \frac{a^2+ab+b^2}{3},$$

which, solving for a *and* b, *results in the moment estimators of* a *and* b,

$$\widehat{a} = \widehat{\mu}_1 - \sqrt{3\left(\widehat{\mu}_2 - \widehat{\mu}_1^2\right)} \quad \text{and} \quad \widehat{b} = \widehat{\mu}_1 + \sqrt{3\left(\widehat{\mu}_2 - \widehat{\mu}_1^2\right)}.$$

In Example 5.2.4, if $a = -b$, that is, $X_1, \ldots, X_n$ is a random sample from a uniform distribution on the interval $(-b, b)$, the problem reduces to a one-parameter estimation problem. However, in this case $E(X_i) = 0$, so the first moment cannot be used to estimate b. It becomes necessary to use the second moment. For the derivation, see Exercise 5.2.3.

It is important to observe that the method of moments estimators need not be unique. The following is an example of the nonuniqueness of moment estimators.

EXAMPLE 5.2.5

Let $X_1, \ldots, X_n$ be a random sample from a Poisson distribution with parameter $\lambda > 0$. Show that both $(1/n)\sum_{i=1}^{n} X_i$ and $(1/n)\sum_{i=1}^{n} X_i^2 - \left((1/n)\sum_{i=1}^{n} X_i\right)^2$ are moment estimators of λ.

Solution

We know that $E(X) = \lambda$, *from which we have a moment estimator of* λ *as* $(1/n)\sum_{i=1}^{n} X_i$. *Also, because we have* $\mathrm{Var}(X) = \lambda$, *equating the second moments, we can see that:*

$$\lambda = E\left(X^2\right) - (EX)^2,$$

so that:

$$\widehat{\lambda} = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)^2.$$

Thus,

$$\widehat{\lambda} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

and

$$\widehat{\lambda} = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)^2.$$

Both are moment estimators of λ. *Thus, the moment estimators may not be unique. We generally choose* $\overline{X}$ *as an estimator of* λ, *for its simplicity.*

It is important to note that, in general, we have as many moment conditions as parameters. In Example 5.2.5, we have more moment conditions than parameters, because both the mean and the variance of Poisson random variables are the same. Given a sample, this results in two different estimates of a single parameter. One of the questions could be, Can these two estimators be combined in some optimal way? This is done by the so-called generalized method of moments. We will not deal with this topic. The method of moments often provides estimators when other methods fail to do so or when estimators are harder to obtain, as in the case of a gamma distribution. Compared with other methods, method of moments estimators are easier to compute and have some desirable properties that we will discuss in the ensuing section.

5.2.2 The method of maximum likelihood

Now we will present an important method for finding estimators of parameters proposed by geneticist/statistician Sir Ronald A. Fisher around 1922 called the method of maximum likelihood. Even though the method of moments is intuitive and easy to apply, it usually does not yield "good" estimators. The method of maximum likelihood is intuitively appealing, because we attempt to find the values of the true parameters that would have most likely produced the data that we in fact observed. For most cases of practical interest, the performance of maximum likelihood estimators (MLEs) is optimal for large enough data. This is one of the most versatile methods for fitting parametric statistical models to data. First, we define the concept of a likelihood function.

Definition 5.2.1 *Let* $f(x_1, \ldots, x_n; \theta)$, $\theta \in \Theta \subseteq \mathbb{R}^k$, *be the joint probability (or density) function of* n *random variables* $X_1, \ldots, X_n$ *with sample values* $x_1, \ldots, x_n$. *The* **likelihood function** *of the sample is given by:*

$$L(\theta; x_1, \ldots, x_n) = f(x_1, \ldots, x_n; \theta), [= L(\theta), is\ a\ briefer\ notation].$$

We emphasize that L *is a function of* θ *for fixed sample values.*

The likelihood of a set of parameter values θ, given $x_1, \ldots, x_n$, is equal to the probability of those observed outcomes given the parameter values. If $X_1, \ldots, X_n$ are discrete iid random variables with probability function $p(x, \theta)$, then the likelihood function is given by:

$$L(\theta) = P(X_1 = x_1, \ldots, X_n = x_n)$$

$$= \prod_{i=1}^{n} P(X_i = x_i), \quad \text{(by multiplication rule for independent random variables)}$$

$$= \prod_{i=1}^{n} p(x_i, \theta)$$

and in the continuous case, if the density is $f(x, \theta)$, then the likelihood function is:

$$L(\theta) = \prod_{i=1}^{n} f(x_i, \theta).$$

It is important to note that the likelihood function, although it depends on the observed sample values $x = (x_1, \ldots, x_n)$, is to be regarded as a function of the parameter θ. In the discrete case, $L(\theta; x_1, \ldots, x_n)$ gives the probability of observing $x = (x_1, \ldots, x_n)$, for a given θ. Thus, the likelihood function is a statistic, depending on the observed sample $x = (x_1, \ldots, x_n)$.

EXAMPLE 5.2.6

Let $X_1, \ldots, X_n$ be iid $N(\mu, \sigma^2)$ random variables. Let $x_1, \ldots, x_n$ be the sample values. Find the likelihood function.

Solution

The density function for the normal variable is given by $f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$. *Hence, the likelihood:*

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

A statistical procedure should be consistent with the assumption that the best explanation of a set of data is provided by an estimator $\widehat{\theta}$, which will be the value of the parameter θ that maximizes the likelihood function. This value of θ will be called the MLE. The goal of maximum likelihood estimation is to find the parameter value(s) that makes the observed data most likely.

Definition 5.2.2 **Maximum likelihood estimators** *are those values of the parameters that maximize the likelihood function with respect to the parameter* θ. *That is,*

$$L\left(\widehat{\theta}; x_1 \ldots, x_n\right) = \max_{\theta \in \Theta} L(\theta; x_1, \ldots, x_n),$$

where Θ *is the set of possible values of the parameter* θ.

The method of maximum likelihood extends to the case of several parameters. Let $X_1, \ldots, X_n$ be a random sample with joint pmf (if discrete) or pdf (if continuous):

$$L(\theta_1, \ldots, \theta_m; x_1 \ldots, x_n) = f(x_1, x_2, \ldots, x_n; \theta_1, \theta_2, \ldots, \theta_m),$$

where the values of the parameters $\theta_1, \ldots, \theta_m$ are unknown and $x_1, \ldots, x_n$ are the observed sample values. Then, the maximum likelihood estimates $\widehat{\theta}_1, \ldots, \widehat{\theta}_m$ are those values of the $\theta_i' s$ that maximize the likelihood function, so that:

$$f\left(x_1, \ldots, x_n; \widehat{\theta}_1, \ldots, \widehat{\theta}_m\right) \geq f(x_1, \ldots, x_n; \theta_1, \ldots, \theta_m)$$

$$\text{for all allowable } \theta_1, \ldots, \theta_m.$$

Note that the likelihood function conveys to us how feasible the observed sample is as a function of the possible parameter values. Maximum likelihood estimates give the parameter values for which the observed sample is most likely to have been generated. In general, the maximum likelihood method results in the problem of maximizing a function of a single or several variables. Hence, in most situations, the methods of calculus can be used. In deriving the MLEs, however, there are situations in which the techniques developed are more problem specific. Sometimes we need to use numerical methods, such as Newton's method.

To find an MLE, we need only compute the likelihood function and then maximize that function with respect to the parameter of interest. In many cases, it is easier to work with the natural logarithm (ln) of the likelihood function, called the *log-likelihood function*. Because the natural logarithm function is increasing, the maximum value of the likelihood function, if it exists, will occur at the same point as the maximum value of the log-likelihood function. We now summarize the calculus-based procedure to find MLEs.

Procedure to find the maximum likelihood estimator

1. Define the likelihood function, $L(\theta)$.
2. Often it is easier to take the natural logarithm (ln) of $L(\theta)$.
3. When applicable, differentiate $\ln L(\theta)$ with respect to θ, and then equate the derivative to zero.
4. Solve for the parameter θ, and we will obtain $\widehat{\theta}$.
5. Check whether it is a maximizer or a global maximizer.

EXAMPLE 5.2.7

Suppose $X_1, \ldots, X_n$ is a random sample from a geometric distribution with parameter p, $0 \leq p \leq 1$. Find the MLE $\widehat{p}$.

Solution

For the geometric distribution, the pmf is given by:

$$f(x, p) = p(1-p)^{x-1}, \quad 0 \leq p \leq 1, \quad x = 1, 2, 3, \ldots.$$

Hence, the likelihood function is:

$$L(p) = \prod_{i=1}^{n} \left[p(1-p)^{x-1}\right] = p^n (1-p)^{-n+\sum_{i=1}^{n} x_i}.$$

Taking the natural logarithm of L(p),

$$\ln L = n \ln p + \left(-n + \sum_{i=1}^{n} x_i\right) \ln (1-p).$$

Taking the derivative with respect to p, *we have:*

$$\frac{d \ln L}{dp} = \frac{n}{p} - \frac{\left(-n + \sum_{i=1}^{n} x_i\right)}{(1-p)}.$$

Equating $\frac{d \ln L(p)}{dp}$ *to zero, we have:*

$$\frac{n}{p} - \frac{\left(-n + \sum_{i=1}^{n} x_i\right)}{(1-p)} = 0.$$

Solving for p,

$$p = \frac{n}{\sum_{i=1}^{n} x_i} = \frac{1}{\bar{x}}.$$

Thus, we obtain an MLE of p *as:*

$$\hat{p} = \frac{n}{\sum_{i=1}^{n} X_i} = \frac{1}{\bar{X}}.$$

We remark that $(1/\bar{X})$ *is the maximum likelihood estimate of* p. *It can be shown that* $\hat{p}$ *is a global maximum.*

EXAMPLE 5.2.8

(a) Suppose $X_1, \ldots, X_n$ is a random sample from a Poisson distribution with parameter λ. Find MLE $\hat{\lambda}$.

(b) Traffic engineers use the Poisson distribution to model light traffic. This is based on the rationale that when the rate is approximately constant in light traffic, the distribution of counts of cars in a given time interval should be Poisson. The following data show the number of vehicles turning left in 15 randomly chosen 5-minute intervals at a specific intersection. Calculate the maximum likelihood estimate.

10	17	12	6	12	11	9	6
10	8	8	16	7	10	6	

Solution

(a) *We have the pmf:*

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

Hence, the likelihood function is:

$$L(\lambda) = \prod_{i=1}^{n} \frac{\lambda^{x_i} e^{-\lambda}}{x_i!} = \frac{\lambda^{\sum_{i=1}^{n} x_i} e^{-n\lambda}}{\prod_{i=1}^{n} x_i!}.$$

Then, taking the natural logarithm, we have:

$$\ln L(\lambda) = \sum_{i=1}^{n} x_i \ln\lambda - n\lambda - \sum_{i=1}^{n} \ln(x_i!)$$

and differentiating with respect to λ *results in:*

$$\frac{d \ln L(\lambda)}{d\lambda} = \frac{\sum_{i=1}^{n} x_i}{\lambda} - n$$

and

$$\frac{d \ln L(\lambda)}{d\lambda} = 0, \ \ \textit{implies} \ \ \frac{\sum_{i=1}^{n} x_i}{\lambda} - n = 0.$$

That is,

$$\lambda = \frac{\sum_{i=1}^{n} x_i}{n} = \overline{x}.$$

Hence, the MLE of λ is:

$$\widehat{\lambda} = \overline{X}.$$

(b) *From (a) we have the estimate as:*

$$\widehat{\lambda} = \overline{x} = 9.8,$$

or approximately 10 vehicles per 5 minutes turn left at this intersection.

It can be verified that the second derivative is negative and, hence, we really have a maximum.

Sometimes the method of derivatives cannot be used for finding the MLE. For example, the likelihood is not differentiable in the range space. In this case, we need to make use of the special structures available in the specific situation to solve the problem. The following is one such case.

EXAMPLE 5.2.9

Let $X_1, \ldots, X_n$ be a random sample from $U(0, \theta)$, $\theta > 0$. Find the MLE of θ.

Solution

Note that the pdf of the uniform distribution is:

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

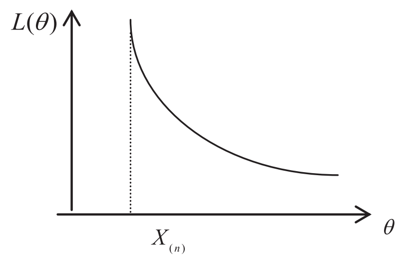


FIGURE 5.1 Likelihood function for uniform probability distribution.

Hence, the likelihood function is given by:

$$L(\theta,\ x_1, x_2,\ \ldots, x_n) = \begin{cases} \dfrac{1}{\theta^n}, & 0 \le x_1, x_2, \ldots, x_n \le \theta \\ 0, & \text{otherwise.} \end{cases}$$

When $\theta \ge \max(x_i)$, *the likelihood is* $(1/\theta^n)$, *which is positive and decreasing as a function of* θ *(for fixed* n*). However, for* $\theta < \max(x_i)$ *the likelihood drops to 0, creating a discontinuity at the point* $\max(x_i)$ *(this is the minimum value of* θ *that can be chosen that still satisfies the condition* $0 \le x_i \le \theta$*), and Fig. 5.1 shows that the maximum occurs at this point. Hence, we will not be able to find the derivative. Thus, the MLE is the largest order statistic,*

$$\widehat{\theta} = \max\,(X_i) = X_{(n)}.$$

In the previous example, because $E(X) = (\theta/2)$, we can see that $\theta = 2E(X)$. Hence, the method of moments estimator for θ is $\widehat{\theta} = 2\overline{X}$. Sometimes the method of moments estimator can give meaningless results. To see this, suppose we observe values 3, 5, 6, and 18 from a $U(0, \theta)$ distribution. Clearly, the maximum likelihood estimate of θ is 18, whereas the method of moments estimate is 16, which is not quite acceptable, because we have already observed a value of 18.

As mentioned earlier, if the unknown parameter θ represents a vector of parameters, say, $\theta = (\theta_1, \ldots, \theta_l)$, then the MLEs can be obtained from solutions of the system of equations:

$$\frac{\partial}{\partial \theta} \ln L(\theta_1, \ldots, \theta_n) = 0, \text{for} \quad i = 1, \ldots, l.$$

These are called the maximum likelihood equations and the solutions are denoted by $\left(\widehat{\theta}_1, \ldots, \widehat{\theta}_l\right)$.

EXAMPLE 5.2.10

Let $X_1, \ldots, X_n$ be $N(\mu, \sigma^2)$.

(a) If μ is unknown and $\sigma^2 = \sigma_0^2$ is known, find the MLE for μ.
(b) If $\mu = \mu_0$ is known and σ^2 is unknown, find the MLE for σ^2.
(c) If μ and σ^2 are both unknown, find the MLE for $\theta = (\mu, \sigma^2)$.

Solution

To avoid notational confusion when taking the derivative, let $\theta = \sigma^2$. *Then, the likelihood function is:*

$$L(\mu, \theta) = (2\pi\theta)^{-n/2} \exp\left(-\frac{\sum_{i=1}^{n} (x_i - \mu)^2}{2\theta} \right)$$

or

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

(a) *When* $\theta = \theta_0 = \sigma_0^2$ *is known, the problem reduces to estimating only one parameter,* μ. *Differentiating the log-likelihood function with respect to* μ,

$$\frac{\partial}{\partial \mu} (\ln L(\mu, \theta_0)) = \frac{2 \sum_{i=1}^{n} (x_i - \mu)}{2\theta_0}.$$

Setting the derivative equal to zero and solving for μ,

$$\sum_{i=1}^{n} (x_i - \mu) = 0.$$

From this,

$$\sum_{i=1}^{n} x_i = n\mu \quad or \quad \mu = \overline{x}.$$

Thus, we get $\widehat{\mu} = \overline{X}$.

(b) *When* $\mu = \mu_0$ *is known, the problem reduces to estimating only one parameter,* $\sigma^2 = \theta$. *Differentiating the log-likelihood function with respect to* θ,

$$\frac{\partial \ln L(\mu, \theta)}{\partial \theta} = \frac{-n}{2\theta} + \frac{\sum_{i=1}^{n} (x_i - \mu)^2}{2\theta^2}.$$

Setting the derivative equal to zero and solving for θ, *we get:*

$$\widehat{\theta} = \widehat{\sigma}^2 = \frac{\sum_{i=1}^{n} (X_i - \mu_0)^2}{n}.$$

(c) *When both* μ *and* θ *are unknown, we need to differentiate with respect to both* μ *and* θ *individually:*

$$\frac{\partial \ln L(\mu, \theta)}{\partial \mu} = \frac{2 \sum_{i=1}^{n} (x_i - \mu)}{2\theta}$$

and

$$\frac{\partial \ln L(\mu, \theta)}{\partial \theta} = \frac{-n}{2\theta} + \frac{\sum_{i=1}^{n}(x_i - \mu)^2}{2\theta^2}.$$

Setting the derivatives equal to zero and solving simultaneously, we obtain:

$$\widehat{\mu} = \overline{X},$$

$$\widehat{\sigma}^2 = \widehat{\theta} = \frac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}{n} = S'^2.$$

Note that in (a) and (c), the estimates for μ are the same; however, in (b) and (c), the estimates for σ^2 are different.

At times, the MLEs may be hard to calculate. It may be necessary to use numerical methods to approximate values of the estimate. The following example gives one such case.

EXAMPLE 5.2.11

Let $X_1, \ldots, X_n$ be a random sample from a population with gamma distribution and parameters α and β. Find MLEs for the unknown parameters α and β.

Solution

The pdf for the gamma distribution is given by:

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

The likelihood function is given by:

$$L = L(\alpha, \beta) = \frac{1}{(\Gamma(\alpha)\beta^{\alpha})^n}\prod_{i=1}^{n} x_i^{\alpha-1} e^{-\sum_{i=1}^{n} x_i/\beta}.$$

Taking the logarithms gives:

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

Now taking the partial derivatives with respect to α and β and setting both equal to zero, we have:

$$\frac{\partial}{\partial \alpha}\ln L = -n\frac{\Gamma'(\alpha)}{\Gamma(\alpha)} - n \ln \beta + \sum_{i=1}^{n} \ln x_i = 0$$

$$\frac{\partial}{\partial \beta}\ln L = -n\frac{\alpha}{\beta} + \sum_{i=1}^{n} \frac{x_i}{\beta^2} = 0.$$

Solving the second one to get β in terms of α, we have:

$$\beta = \frac{\overline{x}}{\alpha}.$$

Substituting this β in the first equation, we have to solve:

$$-n\frac{\Gamma'(\alpha)}{\Gamma(\alpha)} - n \ln \frac{\overline{x}}{\alpha} + \sum_{i=1}^{n} \ln x_i = 0$$

for $\alpha > 0$. There is no closed-form solution for α and β. In this case, one can use numerical methods such as the Newton–Raphson method to solve for α, and then use this value to find β.

There are many references available on the web (such as http://www.mn.uio.no/math/tjenester/kunnskap/kompendier/num_opti_likelihoods.pdf) explaining the Newton–Raphson method for the gamma distribution.

In only a few cases are we able to obtain a simple form for the maximum likelihood equation that can be solved by setting the first derivative to zero. Often, we cannot write an equation that can be differentiated to find the MLE parameter estimates. This is especially true in the situation in which the model is complex and involves many parameters. Evaluating the likelihood exhaustively for all values of the parameters becomes almost impossible, even with modern computers. This is why so-called optimization algorithms have become indispensable to statisticians. The purpose of an optimization algorithm is to find as fast as possible the set of parameter values that make the observed data most likely. There are many such algorithms available. We describe the Newton–Raphson method in Project 5F, and another powerful algorithm, known as the expectation maximization algorithm, is given in Section 13.4.

We have been introduced to several classical discrete and continuous *pdfs*, such as the binomial, Poisson, Gaussian (normal), gamma, and exponential *pdfs*, among others. Note that when we use one of these *pdfs* to study a given set of data we refer to it as parametric analysis, because each of the classical *pdfs* contains at least one parameter that plays a major role in the shape of the probability distribution that characterizes the behavior of the phenomenon of interest.

5.2.2.1 Some additional probability distributions

Now, we will introduce some additional probability distributions that play major roles in analyzing data, or information, in health science, environmental science, engineering, business, and economics, among many other important areas in our society. We shall study the **three-parameter gamma *pdf*** and the **Weibull *pdf***. The **Rayleigh *pdf*** and the **power exponential *pdf*** are other examples, which will be given in this chapter. Each of these *pdfs* will be applied to real data: cancer data, hurricane data, global warming data, and environmental (rainfall) data in Chapter 14.

In Example 5.2.11, we have studied the two-parameter gamma probability distribution (pdf); here we shall introduce the three-parameter version, which is useful when we analyze data that exhibit positive skewness. The *three-parameter gamma pdf* is given by:

$$f(x) = \frac{1}{\beta^{\alpha}\boldsymbol{\Gamma}(\alpha)}(x-\gamma)^{\alpha-1} \quad \exp -\frac{(x-\gamma)}{\beta},$$

where $x > \gamma,\quad \beta > 0$ and $\boldsymbol{\Gamma}_{(\alpha)=}\ \int_0^{\infty} x^{\alpha-1}e^{-x}dx.$

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

$$\begin{aligned} F(x) &= P(X \le x) \\ &= \int_{\gamma}^{x} \frac{1}{\beta^{\alpha}\boldsymbol{\Gamma}_{(\alpha)}}(y-\gamma)^{\alpha-1} \exp -\frac{(y-\gamma)}{\beta}dy \\ &= \boldsymbol{\Gamma}_{\frac{x-\gamma}{\beta}}(\alpha).\frac{1}{\boldsymbol{\Gamma}(\alpha)}. \end{aligned}$$

The expected value is given by:

$$E(X) = \int_0^{\infty} xf(x)dx = \gamma + \alpha\beta.$$

Note that when the location parameter $\gamma = 0$ we obtain the two-parameter gamma (pdf).

EXAMPLE 5.2.12

Given a random sample, $X_1, \ldots, X_n$ from a three-parameter gamma distribution, obtain the MLEs of the parameters.

Solution

The likelihood function is given by:

$$L(\alpha,\ \beta,\ \gamma) = \pi_{i=1}^{n}\ f(x_i)$$

$$= \left(\frac{1}{\beta^{\alpha}\boldsymbol{\Gamma}_{(\alpha)}}\right)^{n}\sum_{i=1}^{n}(x_i-\gamma)^{\alpha-1}\pi_{i=1}^{n}\exp-\left(\frac{x_i-\gamma}{\beta}\right),$$

and the log-likelihood function $\ell(\alpha,\ \beta,\ \gamma)$ *of* $\mathrm{L}(\alpha,\beta,\gamma)$ *is given by:*

$$\ell(\alpha,\ \beta,\ \gamma) = -n\,\alpha\ln\beta - n\ln\boldsymbol{\Gamma}_{(\alpha)} + (\alpha-1)\sum_{i=1}^{n}\ln(x_i-\gamma) - \sum_{i=1}^{n}\frac{x_i-\gamma}{\beta}.$$

$$(\alpha-1)\sum_{i=1}^{n}\ln(x_1-\gamma) - \sum_{i=1}^{n}\frac{x_1-\gamma}{\beta}.$$

The maximum likelihood estimator (*MLE*) can be obtained by setting $\frac{\partial\ell}{\partial\alpha} = 0,\quad \frac{\partial\ell}{\partial\beta} = 0$ and $\frac{\partial\ell}{\partial\gamma} = 0.$ *That is,*

$$\frac{\partial\ell}{\partial\beta} = -\frac{n\alpha}{\beta} + \frac{\sum_{i=1}^{n}(x_i-\gamma)}{\beta^2} = 0,$$

which results in the MLE of β *being:*

$$\widehat{\beta} = \frac{\sum_{i=1}^{n}(x_i-\widehat{\gamma})}{n\widehat{\alpha}}, \tag{5.1}$$

$$\frac{\partial\ell}{\partial\alpha} = -n\ln\beta - n\frac{\boldsymbol{\Gamma}'_{(\alpha)}}{\boldsymbol{\Gamma}_{(\alpha)}} + \sum_{i=1}^{n}\ln(x_i-\gamma) = 0.$$

Substituting $\widehat{\beta}$ *in the above expression we have:*

$$\ln\widehat{\alpha} - \frac{\boldsymbol{\Gamma}'_{(\hat{\alpha})}}{\boldsymbol{\Gamma}_{(\hat{\alpha})}} = \ln\left[\frac{1}{n}\sum_{i=1}^{n}(x_1-\widehat{\gamma})\right] - \frac{1}{n}\sum_{i=1}^{n}\ln(x_i-\widehat{\gamma}), \tag{5.2}$$

where $\frac{\boldsymbol{\Gamma}'_{(\alpha)}}{\boldsymbol{\Gamma}_{(\alpha)}}$ *is called the* ***digamma function****, which is defined as the logarithmic derivative of the gamma function. Now,*

$$\frac{\partial\ell}{\partial\alpha} = -(\alpha-1)\sum_{i=1}^{n}\frac{1}{(x_i-\gamma)} + \sum_{i=1}^{n}\frac{1}{\beta} = 0,$$

which reduces to:

$$\sum_{i=1}^{n}\frac{1}{x_i-\widehat{\gamma}} = \frac{n}{\widehat{\beta}(\widehat{\alpha}-1)}. \tag{5.3}$$

Thus, we can proceed to numerically solve (5.1)–(5.3) to obtain (numerically) an approximate MLE $\widehat{\alpha}$, $\widehat{\beta}$, and $\widehat{\gamma}$ *so that we can apply the subject pdf to real data.*

We can also use the cumulative probability distribution of the three-parameter gamma pdf to obtain the quantile, x_p, for which $F(x_p) = 1-p$, that is,

$$F(x_p) = \frac{\boldsymbol{\Gamma}_{x_p}-\gamma^{(\alpha)}}{\beta}\cdot\frac{1}{\boldsymbol{\Gamma}_{(\alpha)}} = 1-p.$$

Substituting the *MLE* for α, β, and γ, that is, $\widehat{\alpha}$, $\widehat{\beta}$, and $\widehat{\gamma}$, we can proceed to obtain approximate estimates of x_p.

The Weibull probability distribution is very important in characterizing the behavior of health, engineering, and environmental data, among others. The ***Weibull pdf*** is given by:

$$f(x) = \frac{\alpha}{\beta}\left(\frac{x-\gamma}{\beta}\right)^{\alpha-1}\exp\left[-\left(\frac{x-\gamma}{\beta}\right)^{\alpha}\right],$$

where $x > 0$, and the shape parameter α, is greater than zero; the scale parameter β is $\beta > 0$; and the location parameter γ is $x > \gamma$. The cumulative probability distribution of the Weibull pdf is given by:

$$F(x) = P(X \le x) = \int_{\gamma}^{x} \frac{\alpha}{\beta}\left(\frac{t-\gamma}{\beta}\right)^{\alpha-1} \exp\left[-\left(\frac{t-\gamma}{\beta}\right)^{\alpha}\right] dt$$

$$= 1 - \exp\left[-\left(\frac{x-\gamma}{\beta}\right)^{\alpha}\right].$$

When $\gamma = 0$, the subject pdf is reduced to a two-parameter Weibull and it is commonly used because of the difficulty in estimating the three-parameter Weibull pdf.

EXAMPLE 5.2.13

For a random sample $X_1, \ldots, X_n$ drawn from the three-parameter Weibull *pdf*, obtain MLEs for the parameters.

Solution

The likelihood function, $L(\alpha, \beta, \gamma)$*, is given by:*

$$L(\alpha, \beta, \gamma) = \alpha^n \beta^{-n\alpha} \left\{ \underset{i=1}{\overset{n}{\pi}} (x_i - \gamma) \right\}^{\alpha-1} \exp\left\{ -\beta^{-\alpha} \sum_{i=1}^{n} (x_i - \gamma)^{\alpha} \right\}$$

and the log-likelihood function $\ell(\alpha, \beta, \gamma)$ *of* $L(\alpha, \beta, \gamma)$ *is given by:*

$$\ell(\alpha, \beta, \gamma) = n \ln \alpha - n\alpha \ln \beta + (\alpha - 1).$$

$$.\sum_{i=1}^{n} \ln(x_i - \gamma) - \beta^{-\alpha} \sum_{i=1}^{n} (x_i - \gamma)^{\alpha}.$$

Setting $\frac{\partial \ell}{\partial \alpha} = 0$, $\frac{\partial \ell}{\partial \beta} = 0$ and $\frac{\partial \ell}{\partial \gamma} = 0$ *and taking the partial derivatives and substituting* $\alpha = \widehat{\alpha}$, $\beta = \widehat{\beta}$, *and* $\gamma = \widehat{\gamma}$ *and simplifying the resulting expression, we have:*

$$\widehat{\alpha} + \sum_{i=1}^{n} \ln(x_i - \widehat{\gamma}) = \frac{n \sum_{i=1}^{n} (x_i - \widehat{\gamma})^{\widehat{\alpha}} \ln(x_i - \widehat{\gamma})}{\sum_{i=1}^{n} (x_i - \widehat{\gamma})^{\widehat{\alpha}}},$$

$$\frac{n\widehat{\alpha} \sum_{i=1}^{n} (x_i - \widehat{\gamma})^{\widehat{\alpha}-1}}{\sum_{i=1}^{n} (x_i - \widehat{\gamma})^{\widehat{\alpha}}} = (\widehat{\alpha} - 1) \sum_{i=1}^{n} \frac{1}{x_i - \widehat{\gamma}}$$

and

$$\widehat{\beta} = \left\{ \frac{1}{n} \sum_{i=1}^{n} (x_i - \widehat{\gamma})^{\widehat{\alpha}} \right\}^{\frac{1}{\widehat{\alpha}}}.$$

The above equation cannot be analytically solved without further restrictions, so we cannot obtain exact values for $\widehat{\alpha}$, $\widehat{\beta}$, and $\widehat{\gamma}$*; however, there are software packages that we can use to obtain approximate estimates of the subject parameters.*

One of the solutions for Example 5.2.13 is given in http://math.ut.ee/acta/12/Bartkute-Sakalauskas.pdf. Thus, we can see from the previous examples that even though MLEs are elegant estimators, sometimes it is not easy or possible to obtain explicit forms. For these estimates to perform parametric analysis on a given set of data that represent a real-world phenomenon of interest, we will need numerical approximations.

We can use the cumulative probability distribution function $F(x)$ to the **quantile** x_p for which $F(x_p) = 1 - p$, which reduces to:

$$x_p = \gamma + \beta(-\ln p)^{\frac{1}{\alpha}}.$$

Thus, using the MLE of the parameters, we have:

$$\widehat{x}_p = \widehat{\gamma} + \widehat{\beta}(-\ln p)^{\frac{1}{\widehat{\alpha}}}.$$

The graphs in Fig. 5.2 illustrate how the Weibull pdf varies with the shape parameter α (Fig. 5.2A) and with the scale parameter β (Fig. 5.2B).

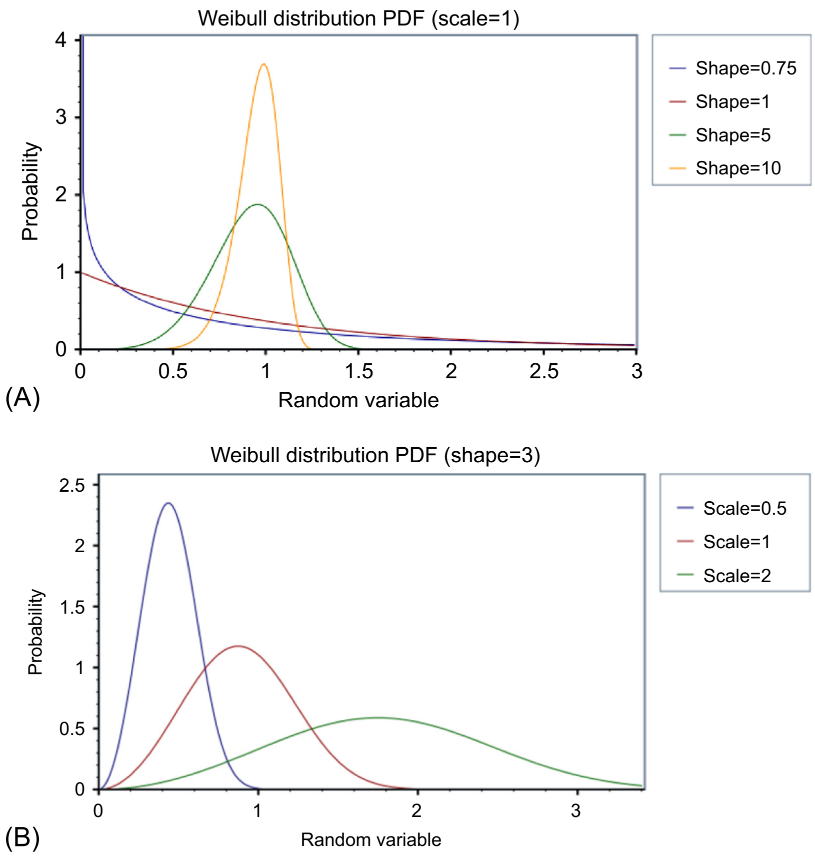


FIGURE 5.2 (A) Weibull distribution with different shape parameters. (B) Weibull distribution with different scale parameters. *PDF*, probability density function.

The ***exponential power*** *or* ***error probability distribution*** is usually applicable in characterizing continuous data that are very nonsymmetric with respect to their mean. It has been shown to be useful in analyzing environmental, engineering, and health data, among others. It is characterized by three parameters that offer the flexibility of addressing different skewness behaviors. Let X be a continuous random variable that characterizes the behavior of a certain problem of interest; the power exponential or error pdf is given by:

$$f(x) = \lambda\left(e^{1-e^{\lambda x^k}}\right)e^{\lambda x^k}x^{k-1}, \quad x > 0,\ \lambda > 0,\ k > 0$$

where λ and k are location and shape parameters, respectively.

The cumulative probability distribution function of the random variable X that follows the exponential power pdf is given by:

$$F(x) = 1 - e^{1-e^{\lambda x^k}}, \quad x > 0,\ \lambda > 0,\ k > 0.$$

The population mean and variance of X are mathematically intractable. Obtaining an MLE analytically is difficult.

The ***Rayleigh distribution*** characterizes the behavior of a continuous random variable that represents many real-world problems. This pdf arises when there is a two-dimensional vector, for example, wind velocity data as measured by an anemometer and wind range that consists of speed value and direction, and both components are normally distributed, are not correlated, and have equal variance. Let X be a continuous random variable that assumes such data; the Rayleigh pdf of the random variable X is given by:

$$f(x;\sigma) = \frac{x}{\sigma^2}e^{\left(\frac{-x^2}{2\sigma^2}\right)}, \quad x > 0,$$

where the scale parameter $\sigma > 0$. The pdf of various values of parameters is given in Fig. 5.3.

The cdf is given by:

$$P(X \le x) = \frac{1}{\sigma^2}\int_0^x \frac{t}{\sigma^2} e^{\left(\frac{-t^2}{2\sigma^2}\right)} dt = 1 - e^{\frac{-x^2}{2\sigma^2}}, \ \ x > 0, \ \ \sigma > 0.$$

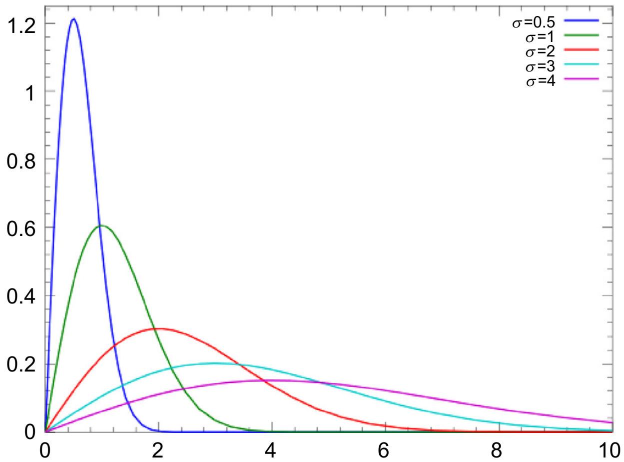


FIGURE 5.3 Rayleigh probability density function for various values of σ.

The expected value and the variance are given by:

$$E(X) = \sigma\sqrt{\frac{\pi}{2}} = 1.25\sigma$$

and

$$Var(X) = \frac{4-\pi}{2}\sigma^2 = 0.429\sigma^2.$$

For a random sample $X_1, \ldots, X_n$ from the Rayleigh pdf, we can verify that the MLE of σ is given by:

$$\widehat{\sigma} = \left[\frac{1}{2n}\sum_{i=1}^{n} X_i^2\right].$$

Sometimes, it may be necessary to estimate a function of a parameter. The following invariance property of MLEs is very useful in those cases.

Theorem 5.2.1 *Let* $\mathrm{h}(\theta)$ *be a one-to-one function of* θ. *If* $\widehat{\theta} = \left(\widehat{\theta}_1, \ldots, \widehat{\theta}_l\right)$ *is the MLE of* $\theta = (\theta_1, \ldots, \theta_l)$, *then the MLE of a function* $\mathrm{h}(\theta) = (\mathrm{h}_l(\theta), \ldots, \mathrm{h}_k(\theta))$ *of these parameters is* $\mathrm{h}\left(\widehat{\theta}\right) = \left(\mathrm{h}_1\left(\widehat{\theta}\right), \ldots, \mathrm{h}_k\left(\widehat{\theta}\right)\right)$ *for* $1 \le \mathrm{k} \le \mathrm{l}$.

As a consequence of the invariance property, in Example 5.2.10, we can obtain the estimator of the true standard deviation as $\widehat{\sigma} = \sqrt{\widehat{\sigma}^2} = \sqrt{(1/n)\Sigma_{i=1}^{n}\left(Xi - \overline{X}\right)^2}$.

It is also known that, under very general conditions on the joint distribution of the sample and for a large sample size n, the MLE $\widehat{\theta}$ is approximately the minimum variance unbiased estimator (MVUE; this concept is introduced in the next section) of θ.

EXERCISES 5.2

5.2.1. Let $X_1, \ldots, X_n$ be a random sample of size n from the geometric distribution for which p is the probability of success.

(a) Use the method of moments to find a point estimator for p.

(b) Use the following data (simulated from geometric distribution) to find the moment estimate for p:

2	5	7	43	18	19	16	11	22
4	34	19	21	23	6	21	7	12

How will you use this information? (The pmf of a geometric distribution is $f(x) = p(1-p)^{x-1}$, for $x = 1, 2, \ldots$. Also, $\mu = 1/p$.)

5.2.2. Let $X_1, \ldots, X_n$ be a random sample of size n from the exponential distribution whose pdf (by taking $\theta = 1/\beta$ in Definition 3.2.7) is:

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

(a) Use the method of moments to find a point estimator for θ.

(b) Find the MLE of θ.

(c) Using the invariance property, obtain an MLE of the variance.

(d) The following data represent the time intervals between the emissions of beta particles:

0.9	0.1	0.1	0.8	0.9	0.1	0.1	0.7	1.0	0.2
0.1	0.1	0.1	2.3	0.8	0.3	0.2	0.1	1.0	0.9
0.1	0.5	0.4	0.6	0.2	0.4	0.2	0.1	0.8	0.2
0.5	3.0	1.0	0.5	0.2	2.0	1.7	0.1	0.3	0.1
0.4	0.5	0.8	0.1	0.1	1.7	0.1	0.2	0.3	0.1

Assuming the data follow an exponential distribution, obtain a moment estimate for the parameter θ. Interpret.

5.2.3. Let $X_1, \ldots, X_n$ be a random sample from a uniform distribution on the interval $(\theta - 1, \theta + 1)$.

(a) Find a moment estimator for θ.

(b) Use the following data to obtain a moment estimate for θ:

11.72	12.81	12.09	13.47	12.37

5.2.4. The probability density of a one-parameter Weibull distribution is given by:

$$f(x) = \begin{cases} 2\alpha x e^{-\alpha x^2}, & x > 0,\ \alpha > 0 \\ 0, & \text{otherwise.} \end{cases}$$

(a) Using a random sample of size n, obtain a moment estimator for α.

(b) Assuming that the following data are from a one-parameter Weibull population,

1.87	1.60	2.36	1.12	0.15
1.83	0.64	1.53	0.73	2.26

obtain a moment estimate of α.

5.2.5. Let $X_1, \ldots, X_n$ be a random sample from the truncated exponential distribution with pdf:

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

(a) Find the method of moments estimate of θ.
(b) Show that the MLE of θ is $\min(X_i, i=1, \ldots, n)$.

5.2.6. Let $X_1, \ldots, X_n$ be a random sample from a distribution with pdf:

$$f(x) = \begin{cases} \dfrac{1+\alpha x}{2}, & -1 \le x \le 1, \quad \text{and} \quad -1 \le \alpha \le 1 \\ 0, & \text{otherwise.} \end{cases}$$

Find the moment estimators for α.

5.2.7. Let $X_1, \ldots, X_n$ be a random sample from a population with pdf:

$$f(x) = \begin{cases} \dfrac{2\alpha^2}{x^3}, & x \ge \alpha \\ 0, & \text{otherwise.} \end{cases}$$

Find a method of moments estimator for α.

5.2.8. Let $X_1, \ldots, X_n$ be a random sample from a negative binomial distribution with pmf:

$$p(x, r, p) = \binom{x+r-1}{r-1} p^x (1-p)^x, \quad 0 \le p \le 1, x = 0, 1, 2, \ldots.$$

Find method of moments estimators for r and p. (Here $E[X] = r(1-p)/p$ and $E[X^2] = r(1-p)(r - rp + 1)/p^2$.)

5.2.9. Let $X_1, \ldots, X_n$ be a random sample from a distribution with pdf:

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

Use the method of moments to obtain an estimator of θ.

5.2.10. Let $X_1, \ldots, X_n$ be a random sample from a distribution with pdf:

$$f(x) = \begin{cases} \dfrac{2\beta - 2x}{\beta^2}, & 0 < x < \beta \\ 0, & \text{otherwise.} \end{cases}$$

Use the method of moments to obtain an estimator of β.

5.2.11. Let $X_1, \ldots, X_n$ be a random sample with common mean μ and variance σ^2. Obtain a method of moments estimator for σ.

5.2.12. Let $X_1, \ldots, X_n$ be a random sample from the beta distribution with parameters α and β. Find the method of moments estimator for α and β.

5.2.13. Let $X_1, X_2, \ldots, X_n$ be a random sample from a distribution with unknown mean μ and variance σ^2. Show that the method of moments estimators for μ and σ^2 are, respectively, the sample mean $\overline{X}$ and $S'^2 = (1/n)\sum_{i=1}^{n}\left(X - \overline{X}\right)^2$. Note that $S'^2 = [(n-1)/n]S^2$ where S^2 is the sample variance.

5.2.14. Let $X_1, \ldots, X_n$ be a random sample recorded as heads or tails resulting from tossing a coin n times with unknown probability p of heads. Find the MLE $\widehat{p}$ of p. Also, using the invariance property, obtain an MLE for $q = 1 - p$. How would you use the results you have obtained?

5.2.15. Let X be a random variable representing the time between successive arrivals at a checkout counter in a supermarket. The values of X in minutes (rounded to the nearest minute) are:

1	2	3	7	11	4	13
12	7	3	2	11	7	2

Assume that the pdf of X is $f(x) = (1/\theta)e^{-(x/\theta)}$. Use these data to find MLE $\widehat{\theta}$. How can you use this estimate you have just derived? Also find the method of moment estimate.

5.2.16. The pdf of a random variable X is given by:

$$f(x) = \begin{cases} \dfrac{2x}{\alpha^2} e^{-x^2/\alpha^2}, & x > 0,\ \alpha > 0 \\ 0, & \text{otherwise.} \end{cases}$$

Using a random sample of size n, obtain MLE $\widehat{\alpha}$ for α.

5.2.17. The pdf of a random variable X is given by:

$$P(X = n) = \frac{1}{n!} \exp\left(\alpha n - e^{\alpha}\right), n = 0, 1, 2, \ldots.$$

Using a random sample of size n, obtain MLE $\widehat{\alpha}$ for α.

5.2.18. Let $X_1, \ldots, X_n$ be a random sample from a two-parameter Weibull distribution with pdf:

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

Obtain maximum likelihood equations and indicate how to obtain the MLEs of α and β.

5.2.19. Let $X_1, \ldots, X_n$ be a random sample from a Rayleigh distribution with pdf:

$$f(x) = \begin{cases} \dfrac{x}{\alpha} e^{-x^2/2\alpha}, & x > 0 \\ 0, & \text{otherwise.} \end{cases}$$

Find the MLEs of α.

5.2.20. Let $X_1, \ldots, X_n$ be a random sample from a two-parameter exponential population with density:

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

Find MLEs for θ and υ when both are unknown.

5.2.21. Let $X_1, \ldots, X_n$ be a random sample from the shifted exponential distribution with:

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

Obtain the MLEs of θ and λ.

5.2.22. Let $X_1, \ldots, X_n$ be a random sample on [0, 1] with pdf:

$$f(x) = \frac{\Gamma(2\theta)}{\Gamma(\theta)^2} [x(1-x)]^{\theta-1}, \theta > 0.$$

What equation does the maximum likelihood estimate of θ satisfy?

5.2.23. Let $X_1, \ldots, X_n$ be a random sample with pdf:

$$f(x) = \begin{cases} (\alpha+1)x^{\alpha}, & 0 \leq x \leq 1 \\ 0, & \text{otherwise.} \end{cases}$$

Find the MLE of α.

5.2.24. Let $X_1, \ldots, X_n$ be a random sample from a uniform distribution with pdf:

$$f(x) = \begin{cases} \dfrac{1}{3\theta+2}, & 0 \leq x \leq 3\theta+2 \\ 0, & \text{otherwise.} \end{cases}$$

Obtain the MLE of θ.

5.2.25. Let $X_1, \ldots, X_n$ be a random sample from a Cauchy distribution with pdf:

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

Obtain maximum likelihood equations and indicate how to obtain the MLE for β.

5.2.26. The following data represent the amounts of leakage of a fluorescent dye from the bloodstream into the eye in patients with abnormal retinas:

1.6	1.4	1.2	2.2	1.8	1.7
1.8	6.3	2.4	2.3	18.9	22.8

Assuming that these data come from a normal distribution, find the maximum likelihood estimate of (μ, σ).

5.2.27. Let $X_1, \ldots, X_n$ be a random sample from a population with gamma distribution and parameters α and β. Show that the MLE of $\mu = \alpha\beta$ is the sample mean $\widehat{\mu} = \overline{X}$.

5.2.28. The lifetime X of a certain brand of component used in a machine can be modeled as a random variable with pdf $f(x) = (1/\theta)e^{-(x/\theta)}$. The reliability $R(x)$ of the component is defined as $R(x) = 1 - F(x)$. Suppose $X_1, X_2, \ldots, X_n$ are the lifetimes of n components randomly selected and tested. Find the MLE of $R(x)$.

5.2.29. Using the method explained in Project 4A, generate 20 observations of a random variable having an exponential distribution with mean and standard deviation both equal to 2. What is the maximum likelihood estimate of the population mean? How much is the observed error?

5.2.30. Let $X_1, \ldots, X_n$ be a random sample from a Pareto distribution (named after the economist Vilfredo Pareto) with shape parameter a. The density function is given by:

$$f(x) = \begin{cases} \dfrac{a}{x^{a+1}}, & x \geq 1 \\ 0, & \text{otherwise.} \end{cases}$$

(The Pareto distribution is a skewed, heavy-tailed distribution. Sometimes it is used to model the distribution of incomes.) Show that the MLE of a is:

$$\widehat{a} = \frac{n}{\sum_{i=1}^{n} \ln(X_i)}.$$

5.2.31. Let $X_1, \ldots, X_n$ be a random sample from $N(\theta, \theta)$, $0 < \theta < \infty$. Find the maximum likelihood estimate of θ.

5.3 Some desirable properties of point estimators

Two different methods of finding estimators for population parameters have been introduced in the preceding section. We have seen that it is possible to have several estimators for the same parameter. For a practitioner of statistics, an important question is going to be which of many available sample statistics, such as mean, median, smallest observation, or largest observation, should be chosen to represent all of the sample? Should we use the method of moments estimator, the MLE, or an estimator obtained through some other method such as the least squares (we will see this method in Chapter 7)? Now we introduce some common ways to distinguish between them by looking at some desirable properties of these estimators.

5.3.1 Unbiased estimators

It is desirable to have the property that the expected value of an estimator of a parameter is equal to the true value of the parameter. Such estimators are called unbiased estimators.

Definition 5.3.1 *A point estimator* $\widehat{\theta}$ *is called an* **unbiased estimator** *of the parameter* θ *if* $E\left(\widehat{\theta}\right) = \theta$ *for all possible values of* θ. *Otherwise* $\widehat{\theta}$ *is said to be* **biased**. Furthermore, the ***bias*** of $\widehat{\theta}$ *is given by:*

$$B = E\left(\widehat{\theta}\right) - \theta.$$

Note that the bias is nothing but the expected value of the (random) error, $E\left(\widehat{\theta}-\theta\right)$. Thus, the estimator is unbiased if the bias is 0 for all values of θ. The bias occurs when a sample does not accurately represent the population from which the sample is taken. It is important to observe that to check whether $\widehat{\theta}$ is unbiased, it is not necessary to know the value of the true parameter. Instead, one can use the sampling distribution of $\widehat{\theta}$. We demonstrate the basic procedure through the following example.

EXAMPLE 5.3.1

Let $X_1, \ldots, X_n$ be a random sample from a Bernoulli population with parameter p. Show that the method of moments estimator is also an unbiased estimator.

Solution

We can verify that the moment estimator of p *is:*

$$\widehat{p} = \frac{\sum_{i=1}^{n} X_i}{n} = \frac{Y}{n}.$$

Because for binomial random variables, E(Y) = np*, it follows that:*

$$E(\widehat{p}) = E\left(\frac{Y}{n}\right) = \frac{1}{n}E(Y) = \frac{1}{n}\bullet np = p.$$

Hence, $\widehat{p} = Y/n$ *is an unbiased estimator for* p.

In fact, we have the following result, which states that the sample mean is always an unbiased estimator of the population mean.

Theorem 5.3.1 *The mean of a random sample* $\overline{X}$ *is an unbiased estimator of the population mean* μ.

Proof. Let $X_1, \ldots, X_n$ be random variables with mean μ. Then, the sample mean is $\overline{X} = (1/n)\Sigma_{i=1}^{n} X_i$:

$$E\overline{X} = \frac{1}{n}\sum_{i=1}^{n} EX_i = \frac{1}{n}\cdot n\mu = \mu.$$

Hence, $\overline{X}$ is an unbiased estimator of μ.

How is this interpreted in practice? Suppose that a data set is collected with n numerical observations $x_1, \ldots, x_n$. The resulting sample mean may be either less than or greater than the true population mean, μ (remember, we do not know this value). If the sampling experiment was repeated many times, then the average of the estimates calculated over these repetitions of the sampling experiment will equal the true population mean.

If we have to choose among several different estimators of a parameter θ, it is desirable to select one that is unbiased. The following result states that the sample variance $S^2 = (1/n-1)\Sigma_{i=1}^{n}\left(X_i - \overline{X}\right)^2$ is an unbiased estimator of the population variance σ^2. This is one of the reasons that in the definition of the sample variance, instead of dividing by n, we divide by $(n-1)$.

Theorem 5.3.2 *If* S^2 *is the variance of a random sample from an infinite population with finite variance* σ^2*, then* S^2 *is an unbiased estimator for* σ^2.

Proof. Let $X_1, \ldots, X_n$ be a random sample with variance $\sigma^2 < \infty$. We have:

$$E\left(S^2\right) = \frac{1}{n-1}E\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2 = \frac{1}{n-1}E\left[\sum_{i=1}^{n}\left\{(X_i-\mu) - \left(\overline{X}-\mu\right)\right\}^2\right]$$

$$= \frac{1}{n-1}\left[\sum_{i=1}^{n} E\{X_i-\mu\}^2 - nE\left\{\overline{X}-\mu\right\}^2\right].$$

Because $E\left\{(X_i-\mu)^2\right\} = \sigma^2$ and $E\left\{\left(\overline{X}-\mu\right)^2\right\} = \sigma^2\big/n$, it follows that:

$$E(S^2) = \frac{1}{n-1}\left[\sum_{i=1}^{n} \sigma^2 - n\frac{\sigma^2}{n}\right] = \sigma^2.$$

Hence, S^2 is an unbiased estimator of σ^2.

It is important to observe the following:

1. S^2 is not an unbiased estimator of the variance of a finite population.
2. Unbiasedness may not be retained under functional transformations, that is, if $\widehat{\theta}$ is an unbiased estimator of θ, it does not follow that $f\left(\widehat{\theta}\right)$ is an unbiased estimator of $f(\theta)$.
3. MLEs or moment estimators are not, in general, unbiased.
4. In many cases it is possible to alter a biased estimator by multiplying by an appropriate constant to obtain an unbiased estimator.

The following example will show that unbiased estimators need not be unique.

EXAMPLE 5.3.2

Let $X_1, \ldots, X_n$ be a random sample from a population with finite mean μ. Show that the sample mean $\overline{X}$ and $\frac{1}{3}\overline{X} + \frac{2}{3}X_1$ are both unbiased estimators of μ.

Solution

By Theorem 5.3.1, $\overline{X}$ is unbiased. Now:

$$E\left[\frac{1}{3}\overline{X} + \frac{2}{3}X_1\right] = \frac{1}{3}\mu + \frac{2}{3}\mu = \mu.$$

Hence, $\frac{1}{3}\overline{X} + \frac{2}{3}X_1$ is also an unbiased estimator of μ.

How many unbiased estimators can we find? In fact, the following example shows that if we have two unbiased estimators, there are infinitely many unbiased estimators.

EXAMPLE 5.3.3

Let $\widehat{\theta}_1$ and $\widehat{\theta}_2$ be two unbiased estimators of θ. Show that:

$$\widehat{\theta}_3 = a\widehat{\theta}_1 + (1-a)\widehat{\theta}_2, 0 \le a \le 1$$

is an unbiased estimator of θ. Note that $\widehat{\theta}_3$ is a convex combination of $\widehat{\theta}_1$ and $\widehat{\theta}_2$. In addition, assume that $\widehat{\theta}_1$ and $\widehat{\theta}_2$ are independent, $Var\left(\widehat{\theta}_1\right) = \sigma_1^2$ and $Var\left(\widehat{\theta}_2\right) = \sigma_2^2$. How should the constant a be chosen to minimize the variance of $\widehat{\theta}_3$?

Solution

We are given that $E\left(\widehat{\theta}_1\right) = \theta$ and $E\left(\widehat{\theta}_2\right) = \theta$. Therefore,

$$\begin{aligned} E\left(\widehat{\theta}_3\right) &= E\left[a\widehat{\theta}_1 + (1-a)\widehat{\theta}_2\right] = aE\widehat{\theta}_1 + (1-a)E\widehat{\theta}_2 \\ &= a\theta + (1-a)\theta = \theta. \end{aligned}$$

Hence, $\widehat{\theta}_3$ is unbiased. By independence,

$$\begin{aligned} Var\left(\widehat{\theta}_3\right) &= Var\left[a\widehat{\theta}_1 + (1-a)\widehat{\theta}_2\right] \\ &= a^2 Var\left(\widehat{\theta}_1\right) + (1-a)^2 Var\left(\widehat{\theta}_2\right) \\ &= a^2\sigma_1^2 + (1-a)^2\sigma_2^2. \end{aligned}$$

To find the minimum,

$$\frac{d}{da} Var\left(\widehat{\theta}_3\right) = 2a\sigma_1^2 - 2(1-a)\sigma_2^2 = 0,$$

gives us:

$$a = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2}.$$

Because $\frac{d^2}{da^2} V\left(\widehat{\theta}_3\right) = 2\sigma_1^2 + 2\sigma_2^2 > 0$, $V\left(\widehat{\theta}_3\right)$ *has a minimum at this value of* 'a'. *Thus, if* $\sigma_1^2 = \sigma_2^2$, *then* $a = 1/2$.

EXAMPLE 5.3.4

Let $X_1, \ldots, X_n$ be a random sample from a population with pdf:

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

Show that the method of moments estimator for the population parameter β is unbiased.

Solution

From Section 5.2, we have seen that the method of moments estimator for β *is the sample mean* $\overline{X}$, *and the population mean is* β. *Because* $E\left(\overline{X}\right) = \mu = \beta$, *the method of moments estimator for the population parameter* β *is unbiased.*

As we have seen, there can be many unbiased estimators of a parameter θ. Which one of these estimators can we choose? If we have to choose an unbiased estimator, it will be desirable to choose the one with the least variance. If an estimator is biased, then we should prefer the one with low bias as well as low variance. Generally, it is better to have an estimator that has low bias as well as low variance. This leads us to the following definition.

Definition 5.3.2 *The* **mean square error** *of the estimator* $\widehat{\theta}$, *denoted by* $MSE\left(\widehat{\theta}\right)$, *is defined as:*

$$MSE\left(\widehat{\theta}\right) = E\left(\widehat{\theta} - \theta\right)^2.$$

Through the following calculations, we will now show that the MSE is a measure that combines both bias and variance:

$$\begin{aligned} MSE\left(\widehat{\theta}\right) &= E\left(\widehat{\theta} - \theta\right)^2 = E\left[\left(\widehat{\theta} - E\left(\widehat{\theta}\right)\right) + \left(E\left(\widehat{\theta}\right) - \theta\right)\right]^2 \\ &= E\left[\left(\widehat{\theta} - E\left(\widehat{\theta}\right)\right)^2 + \left(E\left(\widehat{\theta}\right) - \theta\right)^2 + 2\left(\widehat{\theta} - E\left(\widehat{\theta}\right)\right)\left(E\left(\widehat{\theta}\right) - \theta\right)\right] \\ &= E\left(\widehat{\theta} - E\left(\widehat{\theta}\right)\right)^2 + E\left(E\left(\widehat{\theta}\right) - \theta\right)^2 + 2E\left(\widehat{\theta} - E\left(\widehat{\theta}\right)\right)\left(E\left(\widehat{\theta}\right) - \theta\right) \\ &= Var\left(\widehat{\theta}\right) + \left[E\left(\widehat{\theta}\right) - \theta\right]^2. \end{aligned}$$

Letting $B = E\left(\widehat{\theta}\right) - \theta$, we get:

$$MSE\left(\widehat{\theta}\right) = Var\left(\widehat{\theta}\right) + B^2.$$

B is called the ***bias*** of the estimator. Also, $E\left(\widehat{\theta} - E\left(\widehat{\theta}\right)\right)\left(E\left(\widehat{\theta}\right) - \theta\right) = 0$.

Because the bias is zero for unbiased estimators, it is clear that $MSE\left(\widehat{\theta}\right) = Var\left(\widehat{\theta}\right)$. The MSE measures, on average, how close an estimator comes to the true value of the parameter. Hence, this could be used as a criterion for determining when one estimator is "better" than another. However, in general, it is difficult to find $\widehat{\theta}$ to minimize $MSE\left(\widehat{\theta}\right)$. For this reason, most of the time, we look only at unbiased estimators to minimize $Var\left(\widehat{\theta}\right)$. This leads to the following definition.

Definition 5.3.3 *The unbiased estimator* $\widehat{\theta}$ *that minimizes the MSE is called the MVUE of* θ.

EXAMPLE 5.3.5

Let X_1, X_2, X_3 be a sample of size $n = 3$ from a distribution with unknown mean μ, $-\infty < \mu < \infty$, where the variance σ^2 is a known positive number. Show that both $\widehat{\theta}_1 = \overline{X}$ and $\widehat{\theta}_2 = [(2X_1 + X_2 + 5X_3)/8]$ are unbiased estimators for μ. Compare the variances of $\widehat{\theta}_1$ and $\widehat{\theta}_2$.

Solution

We have:

$$E\left(\widehat{\theta}_1\right) = E\left(\overline{X}\right) = \frac{1}{3}\cdot 3\mu = \mu,$$

and

$$E\left(\widehat{\theta}_2\right) = \frac{1}{8}[2EX_1 + EX_2 + 5EX_3]$$

$$= \frac{1}{8}[2\mu + \mu + 5\mu] = \mu.$$

Hence, both $\widehat{\theta}_1$ *and* $\widehat{\theta}_2$ *are unbiased estimators.*
However,

$$Var\left(\widehat{\theta}_1\right) = \frac{\sigma^2}{3},$$

whereas

$$Var\left(\widehat{\theta}_2\right) = Var\left(\frac{2X_1 + X_2 + 5X_3}{8}\right)$$

$$= \frac{4}{64}\sigma^2 + \frac{1}{64}\sigma^2 + \frac{25}{64}\sigma^2 = \frac{30}{64}\sigma^2.$$

Because $\mathrm{Var}\left(\widehat{\theta}_1\right) < \mathrm{Var}\left(\widehat{\theta}_2\right)$, *we see that* $\overline{X}$ *is a better unbiased estimator in the sense that the variance of* $\overline{X}$ *is smaller.*

It is important to observe that the MLEs are not always unbiased, but it can be shown that for such estimators the bias goes to zero as the sample size increases.

5.3.2 Sufficiency

In the statistical inference problems on a parameter, one of the major questions is: Can a specific statistic replace the entire data without losing pertinent information? Suppose $X_1, \ldots, X_n$ is a random sample from a probability distribution with unknown parameter θ. In general, statisticians look for ways of reducing a set of data so that these data can be more easily understood without losing the meaning associated with the entire collection of observations. Intuitively, a statistic U is a sufficient statistic for a parameter θ if U contains all the information available in the data about the value of θ. For example, the sample mean may contain all the relevant information about the parameter μ, and in that case $U = \overline{X}$ is called a sufficient statistic for μ. An estimator that is a function of a sufficient statistic can be deemed to be a "good" estimator, because it depends on fewer data values. When we have a sufficient statistic U for θ, we need to concentrate only on U because it exhausts all the information that the sample has about θ. That is, knowledge of the actual n observations does not contribute anything more to the inference about θ.

Definition 5.3.4 *Let* $X_1, \ldots, X_n$ *be a random sample from a probability distribution with unknown parameter* θ. *Then, the statistic* $U = g(X_1, \ldots, X_n)$ *is said to be* ***sufficient*** *for* θ *if the conditional pdf or pmf of* $X_1, \ldots, X_n$ *given* $U = u$ *does* *not depend on θ for any value of* u. *An estimator of* θ *that is a function of a sufficient statistic for* θ *is said to be a* ***sufficient estimator*** *of* θ.

EXAMPLE 5.3.6

Let $X_1, \ldots, X_n$ be iid Bernoulli random variables with parameter θ. Show that $U = \Sigma_{i=1}^{n} X_i$ is sufficient for θ.

Solution

The joint pmf of $X_1, \ldots, X_n$ *is:*

$$f(X_1, \ldots, X_n; \theta) = \theta^{\sum_{i=1}^{n} X_i}(1-\theta)^{n-\sum_{i=1}^{n} X_i}, \quad 0 \le \theta \le 1.$$

Because $U = \Sigma_{i=1}^{n} X_i$ *we have:*

$$f(X_1, \ldots, X_n; \theta) = \theta^{U}(1-\theta)^{n-U}, \quad 0 \le U \le n.$$

Also, because $U \sim B(n, \theta)$, *we have:*

$$f(u; \theta) = \binom{n}{u}\theta^{U}(1-\theta)^{n-U}.$$

Also,

$$f(x_1, \ldots, x_n | U = u) = \frac{f(x_1, \ldots, x_n, u)}{f_U(u)} = \begin{cases} \dfrac{f(x_1, \ldots, x_n)}{f_U(u)}, & u = \sum x_i \\ 0, & \text{otherwise.} \end{cases}$$

Therefore,

$$f(x_1, \ldots, x_n | U = u) = \begin{cases} \dfrac{\theta^{u}(1-\theta)^{n-u}}{\binom{n}{u}\theta^{u}(1-\theta)^{n-u}} = \dfrac{1}{\binom{n}{u}} & \text{if } u = \sum x_i \\ 0, & \text{otherwise.} \end{cases}$$

which is independent of θ. *Therefore,* U *is sufficient for* θ.

EXAMPLE 5.3.7

Let $X_1, \ldots, X_n$ be a random sample from $U(0, \theta)$. That is,

$$f(x) = \begin{cases} \dfrac{1}{\theta}, & \text{if } 0 < x < \theta \\ 0, & \text{otherwise.} \end{cases}$$

Show that $U = \max\limits_{1 \le i \le n} X_i$ is sufficient for θ.

Solution

The joint density or the likelihood function is given by:

$$f(x_1, \ldots, x_n; \theta) = \begin{cases} \dfrac{1}{\theta^n}, & \text{if } 0 < x_1, \ldots, x_n < \theta \\ 0, & \text{otherwise.} \end{cases}$$

The joint pdf $f(x_1, \ldots, x_n; \theta)$ *can be equivalently written as:*

$$f(x_1, \ldots, x_n; \theta) = \begin{cases} \dfrac{1}{\theta^n}, & \text{if } x_{\min} > 0, x_{\max} < \theta \\ 0, & \text{otherwise.} \end{cases}$$

Now, we can compute the pdf of U:

$$\begin{aligned} F(u) &= P(U \le u) = P(X_1, \ldots, X_n \le u) \\ &= \prod_{i=1}^{n} P(X_i \le u) \quad \text{(because of independence)} \\ &= \prod_{i=1}^{n} \left(\int_0^u \frac{1}{\theta} dx \right) = \frac{u^n}{\theta^n}, \quad 0 < u < \theta. \end{aligned}$$

The pdf of U *may now be obtained as:*

$$f(u) = \frac{d}{du} F(u) = \frac{n u^{n-1}}{\theta^n}, \quad 0 < u < \theta$$

Moreover,

$$f(x_1, \ldots, x_n | u) = \begin{cases} \dfrac{f(x_1, \ldots, x_n, u)}{f_U(u)} = \dfrac{f(x_1, \ldots x_n)}{f_U(u)}, & \text{if } u = x_{\max} \text{ and } x_{\min} > 0 \\ 0, & \text{otherwise.} \end{cases}$$

Using the expressions for $f(x_1, \ldots, x_n)$ *and* $f_U(u)$ *we obtain:*

$$f(x_1, \ldots, x_n | u = u) = \begin{cases} \dfrac{1/\theta^n}{n u^{n-1}/\theta^n} = \dfrac{1}{n u^{n-1}}, & \text{if } u = x_{\max} \text{ and } x_{\min} > 0 \\ 0, & \text{otherwise} \end{cases}$$

$f(X_1, \ldots, X_n|U)$ *is a function of* u *and* $x_{\min}$, *which is independent of* θ. *Hence,* $U = \max_{1 \le i \le n} X_i$ *is sufficient for* θ.

The outcome $X_1, \ldots, X_n$ is always sufficient, but we will exclude this trivial statistic from consideration. In the previous two examples, we were given a statistic and asked to check whether it was sufficient. It can often be tedious to check whether a statistic is sufficient for a given parameter based directly on the foregoing definition. If the form of the statistic is not given, how do we guess what is the sufficient statistic? Now think of working out the conditional probability by hand for each of our guesses! In general, this will be a tedious way to go about finding sufficient statistics. Fortunately, the Neyman−Fisher factorization theorem makes it easier to spot a sufficient statistic. The following result will give us a convenient way of verifying the sufficiency of a statistic through the likelihood function.

Neyman−Fisher factorization criteria

Theorem 5.3.3 *Let* U *be a statistic based on the random sample* $X_1, \ldots, X_n$. *Then,* U *is a sufficient statistic for* θ *if and only if the joint pdf (or pf)* $f(x_1, \ldots, x_n; \theta)$ *(which depends on the parameter* θ*) can be factored into two nonnegative functions:*

$$f(x_1, \ldots, x_n; \theta) = g(u, \theta)\, h(x_1, \ldots, x_n), \quad \text{for all } x_1, \ldots, x_n,$$

where $g(u, \theta)$ *is a function only of* u *and* θ *and* $h(x_1, \ldots, x_n)$ *is a function of only* $x_1, \ldots, x_n$ *and not of* θ.

Proof (discrete case). We will give the proof only in the discrete case, even though the result is also true for the continuous case. First suppose that $U(X_1, \ldots, X_n)$ is sufficient for θ. Then, $X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n$ if and only if $X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n$ and $U(X_1, \ldots, X_n) = U(x_1, \ldots, x_n) = u$ (*say*). Therefore,

$$f(x_1, \ldots, x_n; \theta) = P_\theta(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n \text{ and } U = u)$$
$$= P_\theta(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n | U = u) P_\theta(U = u).$$

Because U is assumed to be sufficient for θ, the conditional probability $P_\theta(X_1 = x_1, ..., X_n = x_n | U = u)$ does not depend on θ. Let us denote this conditional probability by $h(x_1, \ldots, x_n)$. Clearly $P_\theta(U = u)$ is a function of u and θ. Let us denote this by $g(u, \theta)$.

It now follows from the equation above that:

$$f(x_1, \ldots, x_n; \theta) = g(u, \theta) h(x_1, \ldots, x_n),$$

as was to be shown.

To prove the converse, assume that:

$$f(x_1, \ldots, x_n; \theta) = g(u, \theta) h(x_1, \ldots, x_n).$$

Define the set A_u as:

$$A_u = \{(x_1, \ldots, x_n) : U(x_1, \ldots, x_n) = u\}.$$

That is, A_u is the set of all $(x_1, \ldots, x_n)$ such that U maps it into u. We note that A_u does not depend on θ. Now:

$$P_\theta(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n | U = u)$$

$$= \frac{P_\theta(X_1 = x_1, X_2 = x_2, \ldots X_n = x_n \quad \text{and} \quad U = u)}{P_\theta(U = u)}$$

$$= \begin{cases} \dfrac{P_\theta(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n \quad \text{and} \quad U = u)}{P_\theta(U = u)}, & \text{if } (x_1, \ldots, x_n) \in A_u \\ \\ 0, & \text{if } (x_1, \ldots, x_n) \notin A_u. \end{cases}$$

If $(x_1, \ldots, x_n) \notin A_u$, then, clearly,

$$f(x_1, \ldots, x_n; \theta) = P_\theta(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n | U = u),$$

which is independent of θ.

If $(x_1, \ldots, x_n) \in A_u$, then, using the factorization criterion, we obtain:

$$P_\theta(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n | U = u)$$

$$= \frac{P_\theta(X_1 = x_1, X_2 = x_2, \ldots X_n = x_n)}{P_\theta(U = u)}$$

$$= \frac{f(x_1, \ldots, x_n; \theta)}{P_\theta(U = u)} = \frac{g(u, \theta)\, h\, (x_1, \ldots, x_n)}{\sum\limits_{(x_1, \ldots, x_n) \in A_u} g(u, \theta) h(x_1, \ldots, x_n)}$$

$$= \frac{g(u, \theta) h(x_1, \ldots, x_n)}{g(u, \theta) \sum\limits_{(x_1, \ldots, x_n) \in A_u} h(x_1, \ldots, x_n)} = \frac{h(x_1, \ldots, x_n)}{\sum\limits_{(x_1, \ldots, x_n) \in A_u} h(x_1, \ldots, x_n)}$$

Therefore, the conditional distribution of $X_1, \ldots, X_n$ given U does not depend on θ, proving that U is sufficient.

One can use the following procedure to verify that a given statistic is sufficient. This procedure is based on factorization criteria rather than using the definition of sufficiency directly.

Procedure to verify sufficiency

1. Obtain the joint pdf or pf $f_\theta(x_1, \ldots, x_n)$.
2. If necessary, rewrite the joint pdf or pf in terms of the given statistic and parameter so that one can use the factorization theorem.
3. Define the functions g and h in such a way that g is a function of the statistic and parameter only and h is a function of the observations only.
4. If step 3 is possible, then the statistic is sufficient. Otherwise, it is not sufficient.

In general, it is not easy to use the factorization criterion to show that a statistic U is *not* sufficient. We now give some examples using the factorization theorem.

EXAMPLE 5.3.8

Let $X_1, \ldots, X_n$ denote a random sample from a geometric population with parameter p. Show that $\overline{X}$ is sufficient for p.

Solution

For the geometric distribution, the pf is given by:

$$f(x,p) = \begin{cases} p(1-p)^{x-1}, & x \geq 1 \\ 0, & \text{otherwise.} \end{cases}$$

Hence, the joint pf is:

$$f(x_1, \ldots, x_n; p) = p^n (1-p)^{-n+\sum_{i=1}^{n} x_i}$$

$$= \begin{cases} p^n (1-p)^{n\overline{x}-n}, & \text{if } x_1, \ldots, x_n \geq 1 \\ 0, & \text{otherwise.} \end{cases}$$

Take,

$$g(\overline{x}, p) = p^n (1-p)^{n\overline{x}-n} \quad \text{and} \quad h(x_1, \ldots, x_n) = \begin{cases} 1, & \text{if } x_i \geq 1 \\ 0, & \text{otherwise.} \end{cases}$$

Thus, $\overline{\mathrm{X}}$ *is sufficient for* p.

EXAMPLE 5.3.9

Let $X_1, \ldots, X_n$ denote a random sample from a $U(0, \theta)$ with pdf:

$$f_\theta(x) = \begin{cases} \dfrac{1}{\theta}, & 0 < x < \theta, \quad \theta > 0 \\ 0, & \text{otherwise.} \end{cases}$$

Show that $X_{(n)} = \max\limits_{1 \leq i \leq n} X_i$ is sufficient for θ, using the factorization theorem.

Solution

The likelihood function of the sample is:

$$f_\theta(x_1, \ldots, x_n) = \begin{cases} \dfrac{1}{\theta^n}, & \text{if } 0 < x_1, \ldots, x_n < \theta, \\ 0, & \text{otherwise.} \end{cases}$$

We can now write $f_\theta(x_1, \ldots, x_n)$ *as:*

$$f_\theta(x_1, \ldots, x_n) = h\,(x_1, \ldots, x_n) g\big(\theta, x_{(n)}\big), \text{ for all } x_1, \ldots, x_n$$

where

$$h\,(x_1, \ldots, x_n) = \begin{cases} 1, & \text{if } x_1, \ldots, x_n > 0 \\ 0, & \text{otherwise} \end{cases}$$

and

$$g\big(\theta; x_{(n)}\big) = \begin{cases} \dfrac{1}{\theta^n}, & \text{if } 0 < x_{(n)} < \theta, \\ 0, & \text{otherwise.} \end{cases}$$

From the factorization theorem, we now conclude that $X_{(n)}$ is sufficient for θ. In the next definition, we introduce the concept of joint sufficiency.

Definition 5.3.5 *Two statistics U_1 and U_2 are said to be* **jointly sufficient** *for the parameters θ_1 and θ_2 if the conditional distribution of $X_1, \ldots, X_n$ given U_1 and U_2 does not depend on θ_1 or θ_2. In general, the statistic $\mathbf{U} = (U_1, \ldots, U_n)$ is jointly sufficient for $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_n)$ if the conditional distribution of $X_1, \ldots, X_n$ given $\mathbf{U}$ is free of $\boldsymbol{\theta}$.*

Now we state the factorization criteria for joint sufficiency analogous to the single population parameter case.

The factorization criteria for joint sufficiency

Theorem 5.3.4 *The two statistics U_1 and U_2 are jointly sufficient for θ_1 and θ_2 if and only if the likelihood function can be factored into two nonnegative functions,*

$$f(x_1, \ldots, x_n; \theta_1, \theta_2) = g(u_1, u_2; \theta_1, \theta_2)\ h(x_1, \ldots, x_n)$$

where $g(u_1, u_2; \theta_1, \theta_2)$ is only a function of u_1, u_2; θ_1 and θ_2, and $h(x_1, \ldots, x_n)$ is free of θ_1 or θ_2.

EXAMPLE 5.3.10

Let $X_1, \ldots, X_n$ be a random sample from $N(\mu, \sigma^2)$.

(a) If μ is unknown and $\sigma^2 = \sigma_0^2$ is known, show that $\overline{X}$ is a sufficient statistic for μ.

(b) If $\mu = \mu_0$ is known and σ^2 is unknown, show that $\Sigma_{i=1}^n (X_i - \mu_0)^2$ is sufficient for σ^2.

(c) If μ and σ^2 are both unknown, show that $\Sigma_{i=1}^n X_i$ and $\Sigma_{i=1}^n X_i^2$ are jointly sufficient for μ and σ^2.

Solution

The likelihood function of the sample is:

$$L = \frac{1}{(2\pi)^{n/2}\sigma^n} \exp\left[-\frac{\sum_{i=1}^n (X_i - \mu)^2}{2\sigma^2} \right]$$

$$= \frac{1}{(2\pi)^{n/2}\sigma^n} \exp\left[\frac{1}{2\sigma^2} \left(\sum_{i=1}^n x_i^2 - 2\mu \sum_{i=1}^n x_i + n\mu^2 \right) \right]$$

$$= (2\pi)^{-n/2}\sigma^{-n} \exp\left(-\frac{\sum_{i=1}^n x_i^2}{2\sigma^2} \right) \exp\left(\frac{2\mu n\overline{x}}{2\sigma^2} \right) \exp\left(-\frac{n\mu^2}{2\sigma^2} \right).$$

(a) *When $\sigma^2 = \sigma_0^2$ is known, use the factorization criteria, with:*

$$g(\overline{x}, \mu) = \exp\left(\frac{2n\mu\overline{x} - n\mu^2}{2\sigma_0^2} \right)$$

and

$$h(x_1, \ldots, x_n) = (2\pi)^{-n/2}\sigma^{-n} \exp\left(-\frac{\sum_{i=1}^n x_i^2}{2\sigma^2} \right).$$

Therefore, $\overline{X}$ is sufficient for. μ

(b) *When $\mu = \mu_0$ is known, let*

$$g\left(\sum_{i=1}^n (X_i - \mu)^2, \sigma^2 \right) = \sigma^{-n} \exp\left| -\frac{\sum_{i=1}^n (x_i - \mu)^2}{2\sigma^2} \right|$$

and

$$h(x_1, \ldots, x_n) = \frac{1}{(2\pi)^{n/2}}.$$

Thus $\Sigma_{i=1}^{n}(X_i - \mu)^2$ *is sufficient for* σ^2.

(c) *When both* μ *and* σ^2 *are unknown, use:*

$$g\left(\sum_{i=1}^{n} x_i \sum_{i=1}^{n} x_i^2, \mu, \sigma^2\right) = \sigma^{-n} \exp\left| -\frac{\sum_{i=1}^{n} x_i^2 - 2\mu \sum_{i=1}^{n} x_i + n\mu^2}{2\sigma^2} \right|$$

and

$$h(x_1, \ldots, x_n) = \frac{1}{(2\pi)^{n/2}}.$$

Hence, $\Sigma_{i=1}^{n} X_i$ and $\Sigma_{i=1}^{n} X_i^2$ *are jointly sufficient for* μ *and* σ^2.

EXAMPLE 5.3.11

Suppose that we have a random sample $X_1, \ldots, X_n$ from a discrete distribution given by:

$$f_\theta(x) = C(\theta)2^{-x/\theta}, \quad x = \theta, \theta+1, \theta+2, \ldots; \quad \theta > 0,$$

where $C(\theta) > 0$ is a normalizing constant. Using the factorization theorem, find a sufficient statistic for θ.

Solution

The joint density function $f(x_1, \ldots, x_n; \theta)$ *of the sample* $X_1, \ldots, X_n$ *is:*

$$f(x_1, \ldots, x_n; \theta) = \begin{cases} C(\theta)2^{-\sum_{i=1}^{n}(x_i/\theta)}, & x_1, x_2, \ldots, x_n \text{ are integers } \geq \theta \\ 0, & \text{otherwise.} \end{cases}$$

The function $f(x_1, \ldots, x_n; \theta)$ *can be written as:*

$$f(x_1, \ldots, x_n; \theta) = h(x_1, \ldots, x_n)C(\theta)2^{-\sum_{i=1}^{n}(x_i/\theta)} g_1(\theta, x_{(1)}),$$

where $x_{(1)} = \min_i(x_1, \ldots, x_n)$ *and*

$$h(x_1, x_2, \ldots, x_n) = \begin{cases} 1, & \text{if } x_j - x_{(1)} \geq 0 \text{ is an integer for } j = 1, 2, \ldots, n \\ 0, & \text{otherwise} \end{cases}$$

and

$$g_1(\theta, x_{(1)}) = \begin{cases} 1, & \text{if } x_{(1)} \geq \theta \\ 0, & \text{otherwise.} \end{cases}$$

Thus,

$$f(x_1, \ldots, x_n; \theta) = h(x_1, \ldots, x_n) g\left(\theta, \sum x_i, x_{(1)}\right),$$

where $g(\theta, \Sigma x_i, x_{(1)}) = C(\theta)2{-}\Sigma_{i=1}^{n}(x_i/\theta)\ g_1(\theta, x_{(1)})$. *Using the factorization theorem, we conclude that* $(\Sigma x_i, x_{(1)})$ *is jointly sufficient for* θ. *This result shows that even for a single parameter, we may need more than one statistic for sufficiency.*

When using the factorization criterion, one has to be careful in cases where the range space depends on the parameter.

Using the factorization criterion, we can prove the following result, which says that if we have a unique MLE, then that estimator will be a function of the sufficient statistic.

Theorem 5.3.5 *If* U *is a sufficient statistic for* θ, *the MLE of* θ, *if unique, is a function of* U.

Proof. Because U is sufficient, by Theorem 5.3.3, the joint pdf can be factored as:

$$f(x_1, \ldots, x_n; \theta) = g(u, \theta)h(x_1, \ldots, x_n).$$

This depends on θ only through the statistic U. To maximize L we need to maximize $g(U, \theta)$.

Many common distributions such as Poisson, normal, gamma, and Bernoulli are members of the exponential family of probability distributions. The exponential family of distributions has density functions of the form:

$$f(x;\theta) = \begin{cases} \exp[k(x)c(\theta) + S(x) + d(\theta)], & \text{if } x \in B \\ 0, & x \notin B \end{cases}$$

where B does not depend on the parameter θ.

EXAMPLE 5.3.12

Write the following in exponential form:

(a) $\dfrac{e^{-\lambda}\lambda^x}{x!}$

(b) $p^x(1-p)^{1-x}$

(c) $\frac{1}{\sqrt{2\pi}}e^{-(x-\mu)^2/2}$

Solution

(a) *We have:*

$$\frac{e^{-\lambda}\lambda^x}{x!} = \exp[x \ln\lambda - \ln x! - \lambda].$$

Here $k(x) = x$, $c(\lambda) = \ln \lambda$, $S(x) = -\ln(x!)$ *and* $d(\lambda) = -\lambda$.

(b) *Similarly,*

$$p^x(1-p)^{1-x} = \exp\left[x \ln\left(\frac{p}{1-p}\right) + \ln(1-p)\right], \quad x = 0 \text{ or } 1.$$

(c) *This is the standard normal density:*

$$\frac{1}{\sqrt{2\pi}}e^{-(x-\mu)^2/2} = \exp\left[x\mu - \frac{x^2}{2} - \frac{\mu^2}{2} - \frac{1}{2}\ln(2\pi)\right], \quad -\infty < x < \infty.$$

Note that in the previous example, for each of the cases, $\Sigma_{i=1}^n X_i$ is a sufficient statistic for the parameter. In the next result, we give a generalization of this fact.

Theorem 5.3.6 *Let* $X_1, \ldots, X_n$ *be a random sample from a population with pdf or pmf of the exponential form:*

$$f(x;\theta) = \begin{cases} \exp[k(x)c(\theta) + S(x) + d(\theta)], & \text{if } x \in B \\ 0, & x \notin B \end{cases}$$

where B *does not depend on the parameter* θ. *The statistic* $\Sigma_{i=1}^n k(X_i)$ *is sufficient for* θ.

Proof. The joint density:

$$\begin{aligned} f(x_1, \ldots, x_n; \theta) &= \exp\left[c(\theta)\sum_{i=1}^n k(x_i) + \sum_{i=1}^n S(x_i) + nd(\theta)\right] \\ &= \left\{\exp\left[c(\theta)\sum_{i=1}^n k(x_i) + nd(\theta)\right]\right\}\left\{\exp\left[\sum_{i=1}^n S(x_i)\right]\right\}. \end{aligned}$$

Using the factorization theorem, the statistic $\Sigma_{i=1}^n k(X_i)$ is sufficient.

It does not follow that every function of a sufficient statistic is sufficient. However, any one-to-one function of a sufficient statistic is also sufficient. Every statistic need not be sufficient. When they do exist, sufficient estimators are very important, because if one can find a sufficient estimator it is ordinarily possible to find an unbiased estimator based on the sufficient statistic. Actually, the following theorem shows that if one is searching for an unbiased estimator with minimal variance, it has to be restricted to functions of a sufficient statistic.

Rao–Blackwell theorem

Theorem 5.4.7 *Let* $X_1, \ldots, X_n$ *be a random sample with joint pf or pdf* $f(x_1, \ldots, x_n; \theta)$ *and let* $U = (U_1, \ldots, U_n)$ *be jointly sufficient for* $\theta = (\theta_1, \ldots, \theta_n)$. *If* T *is any unbiased estimator of* $k(\theta)$, *and if* $T^* = E(T|U)$, *then:*

(a) T^* *is an unbiased estimator of* $k(\theta)$.
(b) T^* *is a function of* U *and does not depend on* θ.
(c) $\mathrm{Var}(T^*) \le \mathrm{Var}(T)$ *for every* θ, *and* $\mathrm{Var}(T^*) < \mathrm{Var}(T)$ *for some* θ *unless* $T^* = T$ *with probability 1.*

Proof

(a) By the property of conditional expectation and by the fact that T is an unbiased estimator of $k(\theta)$,

$$E(T^*) = E(E(T|U)) = E(T) = k(\theta).$$

Hence, T^* is an unbiased estimator of $k(\theta)$.
(b) Because U is sufficient for θ, the conditional distribution of any statistic (hence, for T), given U, does not depend on θ. Thus, $T^* = E(T|U)$ is a function of U.
(c) From the property of conditional probability, we have the following:

$$\begin{aligned} Var(T) &= E(Var(T|U)) + Var(E(T|U)) \\ &= E(Var(T|U)) + Var(T^*). \end{aligned}$$

Because $Var\,(T|U) \ge 0$ for all u, it follows that $E(Var\,(T|U)) \ge 0$. Hence, $Var\,(T^*) \le Var\,(T)$. We note that $Var\,(T^*) = Var\,(T)$ if and only if $Var\,(T|U) = 0$ or T is a function of U, in which case $T^* = T$ (from the definition of $T^* = E(T|U) = T$).

In particular, if $k(\theta) = \theta$, and T is an unbiased estimator of θ, then $T^* = E(T|U)$ will typically give the minimum variance unbiased estimator (MVUE) of θ. If T is the sufficient statistic that best summarizes the data from a given distribution with parameter θ, and we can find some function g of T such that $E(g(T)) = \theta$, it follows from the Rao–Blackwell theorem that $g(T)$ is the uniformly minimum variance unbiased estimator (UMVUE) for θ.

EXERCISES 5.3

5.3.1. Let $X_1, \ldots, X_n$ be a random sample from a population with density:

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

(a) Show that $\overline{X}$ is a biased estimator of θ.
(b) Show that $\overline{X}$ is an unbiased estimator of $\mu = 1 + \theta$.

5.3.2. The mean and variance of a finite population $\{a_1, \ldots, a_N\}$ are defined by:

$$\mu = \frac{1}{N}\sum_{i=1}^{N} a_i \quad \text{and} \quad \sigma^2 = \frac{1}{N}\sum_{i=1}^{N} (a_i - \mu)^2.$$

For a finite population, show that the sample variance S^2 is a biased estimator of σ^2.

5.3.3. For an infinite population with finite variance σ^2, show that the sample standard deviation S is a biased estimator for σ. Find an unbiased estimator of σ. (We have seen that S^2 is an unbiased estimator of σ^2. From this exercise, we see that a function of an unbiased estimator need not be an unbiased estimator.)

5.3.4. Let $X_1, \ldots, X_n$ be a random sample from an infinite population with finite variance σ^2. Define:

$$S'^2 = \frac{1}{n}\sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2.$$

Show that S'^2 is a biased estimator for σ^2, and that the bias of S'^2 is $-\frac{\sigma^2}{n}$. Thus, S'^2 is negatively biased, and so on average underestimates the variance. Note that S'^2 is the MLE of σ^2.

5.3.5. Let $X_1, \ldots, X_n$ be a random sample from a population with the mean μ. What condition must be imposed on the constants $c_1, c_2, \ldots, c_n$ so that:

$$c_1X_1 + c_2X_2 + \cdots + c_nX_n$$

is an unbiased estimator of μ?

5.3.6. Let $X_1, \ldots, X_n$ be a random sample from a geometric distribution with parameter θ. Find an unbiased estimate of θ.

5.3.7. Let $X_1, \ldots, X_n$ be a random sample from a $U(0, \theta)$ distribution. Let $Y_n = \max\{X_1, \ldots, X_n\}$. We know (from Example 5.3.4) that $\widehat{\theta}_1 = Y_n$ is an MLE of θ.

(a) Show that $\widehat{\theta}_2 = 2\overline{X}$ is a method of moments estimator.

(b) Show that $\widehat{\theta}_1$ is a biased estimator and $\widehat{\theta}_2$ is an unbiased estimator of θ.

(c) Show that $\widehat{\theta}_3 = \frac{n+1}{n}\widehat{\theta}_1$ is an unbiased estimator of θ.

5.3.8. Let $X_1, \ldots, X_n$ be a random sample from a population with mean μ and variance 1. Show that $\widehat{\mu}^2 = \overline{X}^2$ is a biased estimator of μ^2, and compute the bias.

5.3.9. Let $X_1, \ldots, X_n$ be a random sample from an $N(\mu, \sigma^2)$ distribution. Show that the estimator $\widehat{\mu} = \overline{X}$ is the MVUE for μ.

5.3.10. Let $X_1, \ldots, X_{n_1}$ be a random sample from an $N(\mu_1, \sigma^2)$ distribution and let $Y_1, \ldots, Y_{n2}$ be a random sample from an $N(\mu_2, \sigma^2)$ distribution. Show that the pooled estimator:

$$\widehat{\sigma}^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

is unbiased for σ^2, where S_1^2 and S_2^2 are the respective sample variances.

5.3.11. Let $X_1, \ldots, X_n$ be a random sample from an $N(\mu, \sigma^2)$ distribution. Show that the sample median, M, is an unbiased estimator of the population mean μ. Compare the variances of $\overline{X}$ and M. (Note: For the normal distribution, the mean, median, and mode all occur at the same location. Even though both $\overline{X}$ and M are unbiased, the reason we usually use the mean instead of the median as the estimator of μ is that $\overline{X}$ has a smaller variance than M.)

5.3.12. Let $X_1, \ldots, X_n$ be a random sample from a Poisson distribution with parameter λ. Show that the sample mean $\overline{X}$ is sufficient for λ.

5.3.13. Let $X_1, \ldots, X_n$ be a random sample from a population with density function:

$$f_\sigma(x) = \frac{1}{2\sigma}\exp\left(-\frac{|x|}{\sigma}\right), \quad -\infty < X < \infty, \quad \sigma > 0.$$

Find a sufficient statistic for the parameter σ.

5.3.14. Show that if $\widehat{\theta}$ is a sufficient statistic for the parameter θ and if the MLE of θ is unique, then the MLE is a function of this sufficient statistic $\widehat{\theta}$.

5.3.15. Let $X_1, \ldots, X_n$ be a random sample from an exponential population with parameter θ. Show that $\Sigma_{i=1}^n X_i$ is sufficient for θ. Also show that $\overline{X}$ is sufficient for θ.

5.3.16. The following is a random sample from an exponential distribution:

1.5	3.0	2.6	6.8	0.7	2.2	1.3	1.6	1.1	6.5
0.3	2.0	1.8	1.0	0.7	0.7	1.6	3.0	2.0	2.5
5.7	0.1	0.2	0.5	0.4					

(a) What is an unbiased estimate of the mean?

(b) Using (a) and these data, find two sufficient statistics for the parameter θ.

5.3.17. Let $X_1, \ldots, X_n$ be a random sample from a one-parameter Weibull distribution with pdf:

$$f(x) = \begin{cases} 2\alpha x e^{-\alpha x^2}, & x > 0 \\ 0, & \text{otherwise.} \end{cases}$$

(a) Find a sufficient statistic for α.

(b) Using (a), find a UMVUE for α.

5.3.18. Let $X_1, \ldots, X_n$ be a random sample from a population with density function:

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

Show that $\left(\min_{1\le i\le n} X_i, \max_{1\le i\le n} X_i\right)$ is sufficient for θ.

5.3.19. Let $X_1, \ldots, X_n$ be a random sample from a $G(1, \beta)$ distribution.

(a) Show that $U = \sum_{i=1}^{n} X_i$ is a sufficient statistic for β.

(b) The following is a random sample from a $G(1, \beta)$ distribution:

0.3	3.4	0.4	1.8	0.7	1.0	0.1	2.3	3.7	2.0
0.3	3.7	0.1	1.3	1.2	3.3	0.2	1.3	0.6	0.4

Find a sufficient statistic for β.

5.3.20. Show that X_1 is not sufficient for μ, if $X_1, \ldots, X_n$ is a sample from $N(\mu, 1)$.

5.3.21. Let $X_1, \ldots, X_n$ be a random sample from the truncated exponential distribution with pdf:

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

Show that $X_{(1)} = \min(X_i)$ is sufficient for θ.

5.3.22. Let $X_1, \ldots, X_n$ be a random sample from a distribution with pdf:

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

Show that $U = X_1, \ldots, X_n$ is a sufficient statistic for θ.

5.3.23. Let $X_1, \ldots, X_n$ be a random sample from a Rayleigh distribution with pdf:

$$f(x) = \begin{cases} \frac{2x}{\alpha} e^{-x^2/\alpha}, & x > 0 \\ 0, & \text{otherwise.} \end{cases}$$

Show that $\sum_{i=1}^{n} X_i^2$ is sufficient for the parameter α.

5.4 A method of finding the confidence interval: pivotal method

In the previous sections, we studied methods for finding point estimators for the population parameters. In general, the estimates will differ from the true parameter values by varying amounts depending on the sample values obtained. In addition, the point estimates do not convey any measure of reliability.

Now, we discuss another type of estimation, called an *interval estimation.* Although point estimators are useful, interval estimators convey more information about the data that are used to obtain the point estimate. The purpose of using an interval estimator is to have some degree of confidence of securing the true parameter. For an interval estimator of a single parameter θ, we will use the random sample to find two quantities L and U such that $L < \theta < U$ with some probability. Because L and U depend on the sample values, they will be random. This interval (L, U) should have two properties: (1) $P(L < \theta < U)$ is high, that is, the true parameter θ is in (L, U) with high probability, and (2) the length of the interval (L, U) should be relatively narrow on average.

In summary, interval estimation goes a step beyond point estimation by providing, in addition to the estimating interval (L, U), a measure of one's confidence in the accuracy of the estimate. Interval estimators are called *confidence intervals* and the limits are called U and L, the *upper* and *lower confidence limits,* respectively. The associated levels of confidence are determined by specified probabilities. The width of the CI reflects the amount of variability inherent in the point estimate. Thus, our objective is to find a narrow interval with high probability of enclosing the true parameter, θ. We will restrict our attention to single-parameter estimation.

The probability that a CI will contain the true parameter θ is called the *confidence coefficient.* The confidence coefficient gives the fraction of the time that the constructed interval will contain the true parameter, under repeated sampling.

Let L and U be the lower and upper confidence limits for a parameter θ based on a random sample $X_1, \ldots, X_n$. Both L and U are functions of the sample. We can write the interval estimate of θ as:

$$P(L \leq \theta \leq U) = 1 - \alpha,$$

and we read it as we are $(1 - \alpha)100\%$ confident that the true parameter θ is located in the interval (L, U). The number $1 - \alpha$ is the confidence coefficient, and the interval (L, U) is referred to as a $((1 - \alpha)100\%$ CI) for θ. Thus, if we want a 95% CI for, say, population mean μ, then $\alpha = 0.05$. Note that for the discrete random variables, we may not be able to find a lower bound L and an upper bound U such that the probability, $P(L \leq \theta \leq U)$, is exactly $(1 - \alpha)$. In such a case we can choose L and U such that $P(L \leq \theta \leq U) \geq 1 - \alpha$.

How do we find the CI? For this, we use the error structure of the point estimator to obtain this interval. For instance, we know that the sample mean, $\overline{X}$, is a point estimate (MLE or unbiased estimator) of the population mean μ. In this case, we also know that the standard error of $\overline{X}$ is $\sigma/\sqrt{n}$. If the sample came from a normal population, then for a 95% CI for the mean, multiply the standard error by 1.96 and then add and subtract this product from the sample mean. From this we can also observe that, if everything else remains the same, the size of the CI reduces as the sample size increases.

EXAMPLE 5.4.1

As part of a promotion, the management of a large health club wants to estimate average weight loss for its members within the first 3 months after joining the club. They took a random sample of 45 members of this health club and found that they lost an average of 13.8 lb within the first 3 months of membership with a sample standard deviation of 4.2 lb. Find a 95% CI for the true mean. What if a random sample of 200 members of this health club also resulted in the same sample mean and sample standard deviation?

Solution

Here a point estimate of the true mean μ *is the sample mean* $\overline{x} = 13.8$ *lb. Because* $n = 45$ *is large enough, we can use the central limit theorem (CLT) and use approximate normality for the distribution of* $\overline{X}$ *with mean* μ *and the approximate standard error* $\left(4.2 / \sqrt{45}\right) = 0.626$. *Thus a 95% CI is* $13.8 \pm (1.96)(0.626)$, *resulting in the interval* $(12.57, 15.03)$. *Thus, on average, with 95% confidence, one can expect the true mean to lie in this interval.*

For $n = 200$, *the standard error is* $\left(4.2 / \sqrt{200}\right) \approx 0.297$. *Thus a 95% CI is* $13.8 \pm (1.96)(0.297)$ *resulting in the interval* $(13.22, 14.38)$. *Thus, the more sample values (that is, the more information) we have, the tighter (smaller width) the interval.*

The previous example was built on our knowledge of the sampling distribution of the sample mean. What if the sampling distribution of the statistic we are interested in is not readily available? More generally, our success in building CIs for an estimate of a parameter depends on identifying a quantity known as the pivot. We now describe this method.

The *pivotal method* is a general method of constructing a CI using a pivotal quantity. This relies on our knowledge of sampling distributions. Here we have to find a pivotal quantity with the following two characteristics:

(i) It is a function of the random sample (a statistic or an estimator $\widehat{\theta}$) and the unknown parameter θ, where θ is the only unknown quantity, and
(ii) It has a probability distribution that does not depend on the parameter θ.

Suppose that $\widehat{\theta} = \widehat{\theta}(X)$ is a point estimate of θ, and let $p\left(\widehat{\theta}, \theta\right)$ be the pivotal quantity. Now, for a given value of α $(0 < \alpha < 1)$, and constants a and b, with $(a < b)$, let

$$P\left(a \leq p\left(\widehat{\theta}, \theta\right) \leq b\right) = 1 - \alpha.$$

Hence, given $\widehat{\theta}$, the inequality is solved for θ to obtain a region of θ values, usually an interval corresponding to the observed $\widehat{\theta}$ value. This will be a desired CI.

From (i) and (ii), it is important to note that the pivotal quantity depends on the parameter, but its distribution is independent of the parameter. Let $X_1, \ldots, X_n$ be a random sample and let $\widehat{\theta}$ be a reasonable point estimate of θ. For instance, $\widehat{\theta}$ could be the maximum likelihood (or some other) estimator of θ. In general, finding a pivotal quantity may not be easy.

However, if $\widehat{\theta}$ is the sample mean $\overline{X}$ or sample variance S^2, we could find a pivotal quantity with known sampling distributions. Suppose $p\left(\widehat{\theta}, \theta\right)$ is a pivotal quantity with known probability distribution that is independent of θ. (Usually, the probability distribution of the pivotal quantity will be standard normal, t, χ^2, or F distribution.) The following are some of the standard pivotal quantities. If the sample $X_1, \ldots, X_n$ is from $N(\mu, \sigma^2)$:

With μ unknown and σ known, let $\overline{X}$ be the sample mean. Then the pivot is $\left(\overline{X} - \mu\right)/\left(\sigma/\sqrt{n}\right)$, which has an $N(0, 1)$ distribution (see comments after Corollary 4.2.2).

With μ unknown and σ unknown, then the pivot is $\left(\overline{X} - \mu\right)/\left(S/\sqrt{n}\right)$, which has a t distribution with $(n-1)$ degrees of freedom (see Theorem 4.2.9). If n is large, using CLT, the distribution of the pivot is approximately $N(0, 1)$.

If σ^2 is unknown, then the pivot is $(n-1)S^2/\sigma^2$, which has a χ^2 distribution with $(n-1)$ degrees of freedom (see Theorem 4.2.8).

The following examples illustrate the pivotal method.

EXAMPLE 5.4.2

Suppose we have a random sample $X_1, \ldots, X_n$ from $N(\mu, 1)$. Construct a 95% CI for μ.

Solution

Here the confidence coefficient is 0.95. We know that the MLE of μ *is* $\overline{X}$, *which has an* $N(\mu, 1/n)$ *distribution. Note that this distribution depends on the unknown value of* μ, *and hence,* $\overline{X}$ *cannot be a pivot. However, taking the z-transform of* $\overline{X}$ *we obtain the pivotal quantity as:*

$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} = \frac{\overline{X} - \mu}{1/\sqrt{n}},$$

which has an $N(0, 1)$ *distribution that is a function of the sample measurements and does not depend on* μ. *Hence, this* Z *can be taken as a pivot* $p\left(\widehat{\theta}, \theta\right)$. *Now to find* a *and* b *such that* $P\left(a \le Z\right) = p\left(\widehat{\theta}, \theta\right) \le b\Big) = 0.95$. *One such choice is to find the value of* a *such that* $P(-a \le Z \le a) = 0.95$. *From the normal table,*

$$P\left(-z_{\alpha/2} \le Z \le z_{\alpha/2}\right) = 0.95,$$

where $z_{\alpha/2}$ *represents the value of* z *with tail area* $\alpha/2$. *This implies* $a = z_{\alpha/2} = 1.96$. Hence,

$$P(-1.96 \le Z \le 1.96) = 0.95,$$

or, using the definition of Z *and solving for* μ, *we obtain:*

$$P\left(\overline{X} - \frac{1.96}{\sqrt{n}} \le \mu \le \overline{X} + \frac{1.96}{\sqrt{n}}\right) = 0.95.$$

Hence, a 95% CI for μ *is* $\left(\overline{X} - (1.96/\sqrt{n}),\ \overline{X} + (1.96/\sqrt{n})\right)$ *Thus, the lower confidence limit* L *is* $\overline{X} - (1.96/\sqrt{n})$ *and the upper confidence limit* U *is* $\overline{X} + (1.96/\sqrt{n})$.

From the derivation of Example 5.4.1, it follows that:

$$P\left(\left|\overline{X} - \mu\right| < z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha.$$

Thus, for a normal population with known variance σ^2, if $\overline{X}$ is used as an estimator of the true mean μ, the probability that the error will be less than $z_{\alpha/2}\sigma/\sqrt{n}$ is $1-\alpha$. It is important to note that there is some arbitrariness in choosing a CI for a given problem. There may be several pivotals for $\widehat{\theta}$ that could be used. Also, it is not necessary to allocate equal probability to the two tails of the distribution; however, doing so may result in the shortest length CI for a given confidence coefficient.

When we make the statement of the form:

$$P\left(\overline{X} - \frac{1.96}{\sqrt{n}} \le \mu \le \overline{X} + \frac{1.96}{\sqrt{n}}\right) = 0.95,$$

we mean that, in an infinite series of trials in which repeated samples of size n are drawn from the same population and 95% CIs for μ are calculated by the same method for each of the samples, the proportion of intervals that actually include μ will be 0.95. Fig. 5.4 illustrates this idea, where the vertical line represents the position of true mean μ and each of the horizontal lines represents a 95% CI of the sample, and 20 samples of size n are taken.

A statement of the type $P(\bar{x}-(1.96/\sqrt{n}) \leq \mu \leq \bar{x}+(1.96/\sqrt{n})) = 0.95$, where $\bar{x}$ is the observed sample mean, is misleading. Once we calculate this interval using a particular sample, then either this interval contains the true mean μ or not, and hence, the probability will be either 0 or 1. Thus, the correct interpretation of CI for the population mean is that if samples of the same size, n, are drawn repeatedly from a population, and a CI is calculated from each sample, then 95% of these intervals should contain the population mean. This is often stated as "We are 95% confident that the true mean is in the interval $\left(\overline{X} - z_{\alpha/2}\left(\sigma/\sqrt{n}\right),\ \overline{X} - z_{\alpha/2}\left(\sigma/\sqrt{n}\right)\right)$." This concept of CI is attributed to Neyman.

We can follow the accompanying procedure to find a CI for the parameter θ.

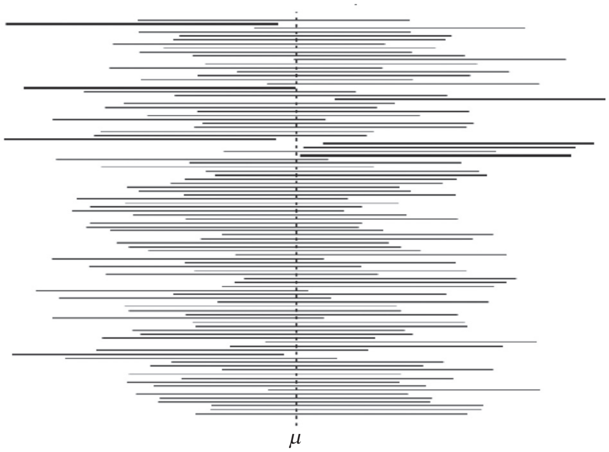


FIGURE 5.4 The 95% confidence intervals for μ.

Procedure to find a confidence interval for θ using the pivot

1. Find an estimator $\widehat{\theta}$ of θ: usually the MLE of θ works.
2. Find a function of θ and $\widehat{\theta}, p\left(\theta, \widehat{\theta}\right)$ (pivot), such that the probability distribution of $p(.,.)$ does not depend on θ.
3. Find a and b such that $P\left(a \leq p\left(\theta, \widehat{\theta}\right) \leq b\right) = 1 - \alpha$. Choose a and b such that $P\left(p\left(\theta, \widehat{\theta}\right) \leq a\right) = \alpha/2$ and $P\left(p\left(\theta, \widehat{\theta}\right) \geq b\right) = \alpha/2$ (see Fig. 5.5 where the shaded area in each side is $\alpha/2$).
4. Now, transform the pivot CI to a CI for the parameter θ. That is, work with the inequality in step 3 and rewrite it as $P(L \leq \theta \leq U) = 1 - \alpha$, where L is the lower confidence limit and U is the upper confidence limit.

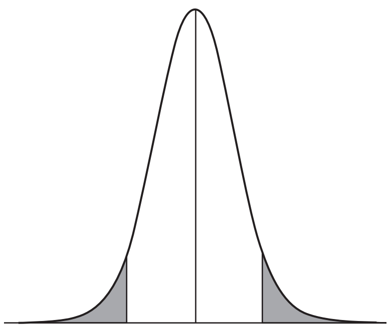


FIGURE 5.5 Probability density of the pivot.

The following example is given to show that the success of finding a pivotal quantity depends on our ability to find the right transformation of the statistic and its distribution so that the transformed variable is a pivot.

EXAMPLE 5.4.3

Suppose the random sample $X_1, \ldots, X_n$ has $U(0, \theta)$ distribution. Construct a 90% CI for θ and interpret. Identify the upper and lower confidence limits.

Solution

From Example 5.3.4, we know that:

$$U = \max_{1\le i\le n} X_i$$

is the MLE of θ. *The random variable* U *has the pdf:*

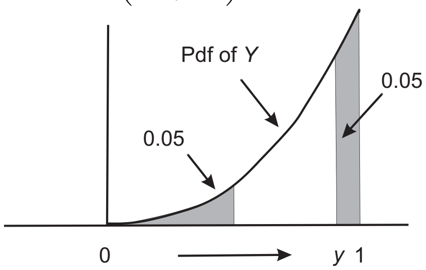
$$f_U(u) = nu^{n-1}/\theta^n, \quad 0 \le u \le \theta.$$

This is not independent of the parameter θ. *Let* $Y = U/\theta$, *then (using the Jacobians described in Chapter 3) the pdf of* Y *is given by:*

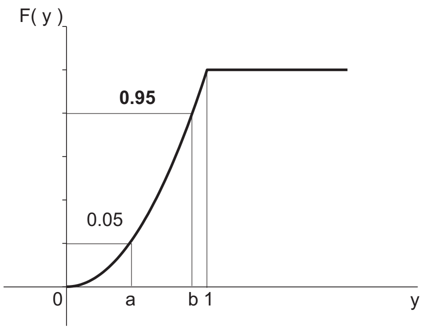
$$f_Y(y) = ny^{n-1}, \quad 0 \le y \le 1.$$

Hence, Y *satisfies the two characteristics of the pivotal quantity. Thus,* $Y = U/\theta$ *is a pivot. Now, we have to find* a *and* b *such that:*

$$p\left(a \le \frac{U}{\theta} \le b\right) = 0.90.$$



To find a *and* b *we use the cdf of* Y, $F_Y(y) = y^n$, $0 \le y \le 1$, *as follows:*



$$F_Y(a) = 0.05 \quad \text{and} \quad F_Y(b) = 0.95,$$

which implies that:

$$a^n = 0.05 \quad \text{and} \quad b^n = 0.95$$

resulting in:

$$a = \sqrt[n]{0.05} \quad \text{and} \quad b = \sqrt[n]{0.95}.$$

Write:

$$P\left(\sqrt[n]{0.05} < \frac{U}{\theta} < \sqrt[n]{0.95}\right) = 0.90.$$

Solving, the 90% CI for θ *is:*

$$\left(\frac{U}{\sqrt[n]{0.95}}, \frac{U}{\sqrt[n]{0.05}}\right)$$

or

$$P\left(\frac{U}{\sqrt[n]{0.95}} \le \theta \le \frac{U}{\sqrt[n]{0.05}}\right) = 0.90.$$

Thus, the lower confidence limit is $U/\sqrt[n]{0.95}$ *and the upper confidence limit is* $U/\sqrt[n]{0.05}$, *and the 90% CI is* $\left(U/\sqrt[n]{0.95},\ U/\sqrt[n]{0.05}\right)$

We can interpret this in the following manner. In a large number of trials in which repeated samples are taken from a population with uniform pdf with parameter θ, approximately 90% of the intervals will contain θ. For instance, if we observed $n = 20$ values from a uniform distribution with the maximum observed value being 15, then a 90% CI for θ is (15.04, 17.42). Thus, we are 90% confident that these data came from a uniform distribution upper limit falling somewhere in this interval.

It is important to note that the pivotal method may not be applicable in all situations. For example, in the binomial case, to find a CI for p, there is no quantity that satisfies the two conditions of a pivot. However, if the sample size is large, then the z-score of sample proportion can be used as a pivot with approximate standard normal distribution. For the pivotal method to work, there is the practical necessity that the distribution of the pivotal quantity make it easy to compute the probabilities. In cases where the pivotal method does not work, we may need to use other techniques such as the method based on sampling distributions (see Project 4A). A proper discussion of these methods is beyond the level of this book.

EXERCISES 5.4

5.4.1. **(a)** Suppose we construct a 99% CI. What are we 99% confident about?
(b) Which of these CIs is wider, 90% or 99%?
(c) In computing a CI, when do you use the t distribution and when do you use z, with normal approximation?
(d) How does the sample size affect the width of a CI?

5.4.2. Suppose X is a random sample of size $n = 1$ from a uniform distribution defined on the interval $(0, \theta)$. Construct a 98% CI for θ and interpret.

5.4.3. Consider the probability statement:

$$P\left(-2.81 \le Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \le 2.75\right) = \kappa,$$

where $\overline{X}$ is the mean of a random sample of size n from an $N(\mu, \sigma^2)$ distribution with known σ^2.
(a) Find κ.
(b) Use this statement to find a CI for μ.
(c) What is the confidence level of this CI?
(d) Find a symmetric CI for μ.

5.4.4. A random sample of size 50 from a particular brand of 16-oz tea packets produced a mean weight of 15.65 oz. Assume that the weights of these brands of tea packets are normally distributed with standard deviation of 0.59 oz. Find a 95% CI for the true mean μ.

5.4.5. Let $X_1, \ldots, X_n$ be a random sample from an $N(\mu, \sigma^2)$, where the value of σ^2 is unknown.
(a) Construct a $(1 - \alpha)100\%$ CI for σ^2, choosing an appropriate pivot. Interpret its meaning.
(b) Suppose a random sample from a normal distribution gives the following summary statistics: $n = 21$, $\overline{x} = 44.3$, and $s = 3.96$. Using (a), find a 90% CI for σ^2. Interpret its meaning.

5.4.6. Let $X_1, \ldots, X_n$ be a random sample from a gamma distribution with $\alpha = 2$ and unknown β. Construct a 95% CI for β.

5.4.7. Let $X_1, \ldots, X_n$ be a random sample from an exponential distribution with pdf $f(x) = (1/\theta)e^{-x/\theta}$, $\theta > 0$, $x > 0$. Construct a 95% CI for θ and interpret. (Hint: Recall that $\sum_{i=1}^{n} X_i$ has a gamma distribution with $\alpha = n$, $\beta = \theta$.)

5.4.8. Let $X_1, \ldots, X_n$ be a random sample from a Poisson distribution with parameter λ.
(a) Construct a 90% CI for λ.

(b) Suppose that the number of raisins in a bowl of a particular brand of cereal is observed to be 25. Assuming that the number of raisins in a bowl is Poisson distributed, estimate the expected number of raisins per bowl with a 90% CI.

(c) How many bowls of cereal need to be sampled to estimate the expected number of raisins per bowl with a standard error of less than 0.2?

5.4.9. Let $X_1, \ldots, X_n$ be a random sample from an $N(\mu, \sigma^2)$.

(a) Construct a $(1-\alpha)100\%$ CI for μ when the value of σ^2 is known.

(b) Construct a $(1-\alpha)100\%$ CI for μ when the value of σ^2 is unknown.

5.4.10. Let $X_1, \ldots, X_n$ be a random sample from an $N(\mu_1, \sigma^2)$ population and $Y_1, \ldots, Y_n$ be an independent random sample from an $N(\mu_2, \sigma^2)$ distribution where σ^2 is assumed to be known. Construct a $(1-\alpha)100\%$ interval for $(\mu_1 - \mu_2)$. Interpret its meaning.

5.4.11. Let $X_1, \ldots, X_n$ be a random sample from a uniform distribution on $[\theta, \theta+1]$. Find a 99% CI for θ, using an appropriate pivot.

5.5 One-sample confidence intervals

In this section, we will find CIs for the one-sample case for both large- and small-sample situations.

5.5.1 Large-sample confidence intervals

If the sample size is large, then by the CLT, certain sampling distributions can be assumed to be approximately normal. That is, if θ is an unknown parameter (such as μ, p, $(\mu_1 - \mu_2)$, $(p_1 - p_2)$), then for large samples, by the CLT, the z-transform:

$$z = \frac{\widehat{\theta} - \theta}{\sigma_{\hat{\theta}}},$$

possesses an approximately standard normal distribution, where $\widehat{\theta}$ is the MLE of θ and $\sigma_{\hat{\theta}}$ is its standard deviation. Then as in Example 5.4.1, the pivotal method can be used to obtain the CI for the parameter θ. For $\theta = \mu$, $n \geq 30$ will be considered large; for the binomial parameter p, n is considered large if np and $n(1-p)$ are both greater than 5. Note that these numbers are only a rule of thumb.

Procedure to calculate large-sample confidence interval for θ

1. Find an estimator (such as the MLE) of θ, say $\widehat{\theta}$.
2. Obtain the standard error, $\sigma_{\hat{\theta}}$ *of* $\widehat{\theta}$.
3. Find the z-transform $z = \left(\widehat{\theta} - \theta\right) / \sigma_{\hat{\theta}}$. Then z has an approximately standard normal distribution.
4. Using the normal table, find two tail values $-z_{\alpha/2}$ and $z_{\alpha/2}$.
5. An approximate $(1-\alpha)100\%$ CI for θ is $\left(\widehat{\theta} - z_{\alpha/2}\sigma_{\hat{\theta}}, \widehat{\theta} + z_{\alpha/2}\sigma_{\hat{\theta}}\right)$, that is,

$$P\left(\widehat{\theta} - z_{\alpha/2}\sigma_{\hat{\theta}} \leq \theta \leq \widehat{\theta} + z_{\alpha/2}\sigma_{\hat{\theta}}\right) = 1 - \alpha.$$

6. *Conclusion:* We are $(1-\alpha)100\%$ confident that the true parameter θ lies in the interval $\left(\widehat{\theta} - z_{\alpha/2}\sigma_{\hat{\theta}}, \widehat{\theta} + z_{\alpha/2}\sigma_{\hat{\theta}}\right)$

EXAMPLE 5.5.1

Let $\widehat{\theta}$ be a statistic that is normally distributed with mean θ and standard deviation $\sigma_{\hat{\theta}}$, where σ is assumed to be known. Find a CI for θ that possesses a confidence coefficient equal to $1 - \alpha$.

Solution

The z-transform of $\widehat{\theta}$ *is:*

$$Z = \frac{\widehat{\theta} - \theta}{\sigma_{\hat{\theta}}}$$

and has a standard normal distribution. Select two tail values $-z_{\alpha/2}$ *and* $z_{\alpha/2}$ *such that:*

$$P\left(-z_{\alpha/2} \le Z \le z_{\alpha/2}\right) = 1 - \alpha.$$

Because of symmetry, this is the shortest interval that contains the area $1 - \alpha$*. Then,*

$$P\left(\widehat{\theta} - z_{\alpha/2}\sigma_{\hat{\theta}} \le \theta \le \widehat{\theta} + z_{\alpha/2}\sigma_{\hat{\theta}}\right) = 1 - \alpha.$$

Therefore, the confidence limits of θ *are* $\widehat{\theta} - z_{\alpha/2}\sigma_{\hat{\theta}}$ *and* $\widehat{\theta} + z_{\alpha/2}\sigma_{\hat{\theta}}$*. Hence, the* $(1-\alpha)100\%$ *CI for* θ *is given by* $\widehat{\theta} \pm z_{\alpha/2}\sigma_{\hat{\theta}}$.

In particular, for a large sample of size n, let $\widehat{\theta} = \overline{X}$ be the sample mean. Then the large-sample $(1-\alpha)100\%$ CI for the population mean μ is:

$$\overline{X} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \simeq \overline{X} \pm z_{\alpha/2}\frac{S}{\sqrt{n}}$$

where S is a point estimate of σ. That is,

$$P\left(\overline{X} - z_{\alpha/2}\frac{S}{\sqrt{n}} \le \mu \le \overline{X} \pm z_{\alpha/2}\frac{S}{\sqrt{n}}\right) = 1 - \alpha.$$

As we have seen in Section 5.4, the correct interpretation of this CI is that in a repeated sampling, approximately $(1-\alpha)$ 100% of all intervals of the form $\overline{X} \pm z_{\alpha/2}(S/\sqrt{n})$ include μ, the true mean. Suppose $\bar{x}$ and s are the sample mean and the sample standard deviation, respectively, for a particular set of n observed sample values $x_1, \ldots, x_n$. Then we do not know whether the particular interval $\left(\bar{x} - z_{\alpha/2}(s/\sqrt{n}),\ \bar{x} - z_{\alpha/2}(s/\sqrt{n})\right)$ contains μ. However, the procedure that produced this interval does capture the true mean in approximately $(1-\alpha)100\%$ of cases. This interpretation will be assumed hereafter, when we make a statement such as, "We are 95% confident that the true mean will lie in the interval (74.1, 79.8)."

EXAMPLE 5.5.2

Two statistics professors want to estimate average scores for an elementary statistics course that has two sections. Each professor teaches one section and each section has a large number of students. A random sample of 50 scores from each section produced the following results:

(a) Section I: $\bar{x}_1 = 77.01$, $s_1 = 10.32$
(b) Section II: $\bar{x}_2 = 72.22$, $s_2 = 11.02$

Calculate 95% CIs for each of these two samples.

Solution

Because $n = 50$ *is large, we could use normal approximation. For* $\alpha = 0.05$*, from the normal table:* $z_{\alpha/2} = z_{0.025} = 1.96$. *The CIs are as follows:*

(a) *We have:*

$$\bar{x}_1 \pm z_{\alpha/2}\frac{s_1}{\sqrt{n}} = 77.01 \pm 1.96\left(\frac{10.32}{\sqrt{50}}\right),$$

which gives a 95% CI (74.149, 79.871).

(b) *We can compute:*

$$\bar{x}_2 \pm z_{\alpha/2}\frac{s_2}{\sqrt{n}} = 72.22 \pm 1.96\left(\frac{11.02}{\sqrt{50}}\right),$$

which gives the interval (69.165, 75.275).

It may be noted that if the population is normal with a known variance σ^2, we can use $\overline{X} \pm z_{\alpha/2}(\sigma/\sqrt{n})$ as the CI for the population mean μ, irrespective of the sample size. However, if σ^2 is unknown, to use $\overline{X} \pm z_{\alpha/2}(s/\sqrt{n})$ as an approximate CI for μ, the sample size has to be large for the CLT to hold. However, to use this approximate procedure, we do not need the condition that samples arise from a normal distribution. We will consider sample size to be large if $n \ge 30$ (applicable to estimators of the mean). If not, we shall use the small-sample procedure discussed in the next section.

EXAMPLE 5.5.3

Fifteen vehicles were observed at random for their speeds (in mph) on a highway with speed limit posted as 70 mph, and it was found that their average speed was 73.3 mph. Suppose that from past experience we can assume that vehicle speeds are normally distributed with $\sigma = 3.2$. Construct a 90% CI for the true mean speed μ, of the vehicles on this highway. Interpret the result.

Solution

Because the population is given to be normal with standard deviation $\sigma = 3.2$*, sample size need not be large, given* $\bar{x} = 73.3$ *and* $\sigma = 3.2$*. Here,* $n = 15$*, and* $\alpha = 0.10$*. Thus,* $z_{\alpha/2} = z_{0.05} = 1.645$*. Hence, a 90% CI for* μ *is given by:*

$$73.3 - 1.645\frac{3.2}{\sqrt{15}} < \mu < 73.3 + 1.645\frac{3.2}{\sqrt{15}}$$

or

$$71.681 < \mu < 74.919.$$

Interpretation: We are 90% confident that the true mean speed μ *of the vehicles on this highway is between 71.681 and 74.919 mph.*

5.5.2 Confidence interval for proportion, *p*

Consider a binomial distribution with parameter p. Let X be the number of successes in n trials. Then the MLE $\widehat{p}$ of p is $\widehat{p} = X/n$. It can be shown, using the procedure outlined at the beginning of this section, that an approximate large-sample $(1-\alpha)100\%$ CI for p is:

$$\left(\widehat{p} - z_{\alpha/2}\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}, \widehat{p} + z_{\alpha/2}\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}\right).$$

That is,

$$P\left(\widehat{p} - z_{\alpha/2}\sqrt{\frac{\widehat{p}1-\widehat{p})}{n}} < p < \widehat{p} + z_{\alpha/2}\sqrt{\frac{\widehat{p}1-\widehat{p})}{n}}\right) = 1-\alpha.$$

A natural question is, "How do we determine that the sample size we have is sufficient for the normal approximation that is used in the foregoing formula?" There are various rules of thumb that are used to determine the adequacy of the sample size for normal approximation. Some of the popular rules are that np and $n(1-p)$ should be greater than 10, or that $\widehat{p} \pm 2\sqrt{\widehat{p}(1-\widehat{p})/n}$ should be contained in the interval (0,1), or $np(1-p) \geq 10$, etc. All of these rules perform poorly when p is nearer to 0 or 1. There have been many works on coverage analysis for CIs. We refer to a survey article by Lee et al., for more details on this topic. For simplicity of calculations, we will use the rule that np and $n(1-p)$ are both greater than 5.

EXAMPLE 5.5.4

An auto manufacturer gives a bumper-to-bumper warranty for 3 years or 36,000 miles for its new vehicles. In a random sample of 60 of its vehicles, 20 of them needed five or more major warranty repairs within the warranty period. Estimate the true proportion of vehicles from this manufacturer that need five or more major repairs during the warranty period, with confidence coefficient 0.95. Interpret.

Solution

Here we need to find a 95% CI for the true proportion, p*. Here,* $\widehat{p} = 20/60 = 1/3$*. For* $\alpha = 0.05$*,* $z_{\alpha/2} = z_{0.025} = 1.96$*. Hence, a 95% CI for* p *is:*

$$\widehat{p} \pm z_{\alpha/2}\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}} = \frac{1}{3} \pm 1.96\sqrt{\frac{\left(\frac{1}{3}\right)\left(\frac{2}{3}\right)}{60}},$$

which gives the CI as (0.21405, 0.45262)*. That is, we are 95% confident that the true proportion of vehicles from this manufacturer that need five or more major repairs during the warranty period will lie in the interval* (0.21405, 0.45262).

5.5.2.1 Margin of error and sample size

In real-world problems, the estimates of the proportion p are usually accompanied by a margin of error, rather than a CI. For example, in the news media, especially leading up to election time, we hear statements such as "The CNN/USA Today/Gallup poll of 818 registered voters taken on June 27–30 showed that if the election were held now, the president would beat his challenger 52% to 40%, with 8% undecided. The poll had a margin of error of plus or minus 4 percentage points." What is this "margin of error"? According to the American Statistical Association, the margin of error is a common summary of sampling error that quantifies uncertainty about a survey result. Thus, the margin of error is nothing but a CI. The number quoted in the foregoing statement is half the maximum width of a 95% CI, expressed as a percentage.

Let b be the width of a 95% CI for the true proportion, p. Let $\widehat{p} = x/n$ be an estimate for p where x is the number of successes in n trials. Then,

$$b = \frac{x}{n} + 1.96\sqrt{\frac{(x/n)(1-(x/n))}{n}} - \left(\frac{x}{n} - 1.96\sqrt{\frac{(x/n)(1-(x/n))}{n}}\right)$$

$$= 3.92\sqrt{\frac{(x/n)(1-(x/n))}{n}} \le 3.92\sqrt{\frac{1}{4n}},$$

because $(x/n)(1-(x/n)) = \widehat{p}(1-\widehat{p}) \le \frac{1}{4}$.

Thus, the margin of error associated with $\widehat{p} = (x/n)$ is $100d\%$, where:

$$d = \frac{\max b}{2} = \frac{3.92\sqrt{\frac{1}{4n}}}{2} = \frac{1.96}{2\sqrt{n}}.$$

From the foregoing derivation, it is clear that we can compute the margin of error for other values of α by replacing 1.96 with the corresponding value of $z_{\alpha/2}$.

A quick look at the formula for the CI for proportions reveals that a larger sample would yield a shorter interval (assuming other things being equal) and hence, a more precise estimate of p. The larger sample is costlier in terms of time, resources, and money, whereas samples that are too small may result in inaccurate inferences. Then, it becomes beneficial for finding out the minimum sample size required (thus less costly) to achieve a prescribed degree of precision (usually, the minimum degree of precision acceptable). We have seen that the large-sample $(1-\alpha)100\%$ CI for p is:

$$\widehat{p} - z_{\alpha/2}\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}} < p < \widehat{p} + z_{\alpha/2}\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}.$$

Rewriting it, we have:

$$|\widehat{p} - p| \le z_{\alpha/2}\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}} = \frac{z_{\alpha/2}}{\sqrt{n}}\sqrt{\widehat{p}(1-\widehat{p})},$$

which shows that, with probability $(1-\alpha)$, the estimate $\widehat{p}$ is within $z_{\alpha/2}\sqrt{\widehat{p}(1-\widehat{p})/n}$ units of p. Because $\widehat{p}(1-\widehat{p}) \le 1/4$, for all values of $\widehat{p}$, we can write the foregoing inequality as:

$$|\widehat{p} - p| \le \frac{z_{\alpha/2}}{\sqrt{n}}\sqrt{\frac{1}{4}} = \frac{z_{\alpha/2}}{2\sqrt{n}}.$$

If we wish to estimate p at level $(1-\alpha)$ to within d units of its true value, that is $|\widehat{p} - p| \le d$, the sample size must satisfy the condition $\left(z_{\alpha/2}\big/(2\sqrt{n})\right) \le d$, or

$$n \ge \frac{z^2_{\alpha/2}}{4d^2}.$$

Thus, to estimate p at level $(1-\alpha)$ to within d units of its true value, take the minimal sample size as $n = z^2_{\alpha/2}\big/(4d^2)$, and if this is not an integer, round up to the next integer.

Sometimes, we may have an initial estimate $\widetilde{p}$ of the parameter p from a similar process or from a pilot study or simulation. In this case, we can use the following formula to compute the minimum required size of the sample to estimate p, at level $(1-\alpha)$, to within d units:

$$n = \frac{z_{\alpha/2}^2 \tilde{p}(1-\tilde{p})}{d^2}$$

and, if this is not an integer, we round up to the next integer.

A similar derivation for calculation of sample size for estimation of the population mean μ at level $(1-\alpha)$ with margin of error E is given by:

$$n = \frac{z_{\alpha/2}^2 \sigma^2}{E^2}$$

and, if this is not an integer, rounding up to the next integer. This formula can be used only if we know the population standard deviation, σ. Although it is unlikely we will know σ when the population mean itself is not known, we may be able to determine σ from an earlier similar study or from a pilot study/simulation.

EXAMPLE 5.5.5

A dendritic tree is a branched formation that originates from a nerve cell. To study brain development, researchers want to examine the brain tissues from adult guinea pigs. How many cells must the researchers select (randomly) so as to be 95% sure that the sample mean is within 3.4 cells of the population mean? Assume that a previous study has shown $\sigma = 10$ cells.

Solution

A 95% confidence corresponds to $\alpha = 0.05$. *Thus, from the normal table,* $z_{\alpha/2} = z_{0.025} = 1.96$. *Given that* $E = 3.4$ *and* $\sigma = 10$, *and using the sample size formula, the required sample size* n *is:*

$$n = \frac{z_{\alpha/2}^2 \sigma^2}{E^2} = \frac{(1.96)^2 (10)^2}{(3.4)^2} = 33.232.$$

Thus, take $n = 34$.

EXAMPLE 5.5.6

Suppose that a local TV station in a city wants to conduct a survey to estimate support for the president's policies on the economy within 3% error with 95% confidence.

(a) How many people should the station survey if they have no information on the support level?

(b) Suppose they have an initial estimate that 70% of the people in the city support the economic policies of the president. How many people should the station survey?

Solution

Here $\alpha = 0.05$, *and thus* $z_{\alpha/2} = 1.96$. *Also,* $d = 0.03$.

(a) *With no information on* p, *we use the sample size formula:*

$$n = \frac{z_{\alpha/2}^2}{4d^2} = \frac{(1.96)^2}{4(0.03)^2} = 1067.1.$$

Hence, the TV station must survey 1068 people.

(b) *Because* $\tilde{p} = 0.7$, *the required sample size is calculated from:*

$$n = \frac{z_{\alpha/2}^2 \tilde{p}(1-\tilde{p})}{d^2}$$

$$= \frac{(1.96)^2 (0.70)(0.30)}{(0.03)^2} = 896.37.$$

Thus, the TV station must survey at least 897 people.

In practice, we should realize that one of the key factors of a good design is not sample size by itself; it is getting representative samples. Even if we have a very large sample size, if the sample is not representative of our target population, then sample size means nothing. Therefore, whenever possible, we should use random sampling procedures (or other appropriate sampling procedures) to ensure that our target population is properly represented.

5.5.3 Small-sample confidence intervals for μ

Now we will consider the problem of finding a CI for the true mean μ of a normal population when the variance σ^2 is unknown and obtaining a large sample is either impossible or impractical. Let $X_1, \ldots, X_n$ be a random sample from a normal population. We know that:

$$T = \frac{\sqrt{n}\,\dfrac{\overline{X}-\mu}{\sigma}}{\sqrt{(n-1)S^2/[\sigma^2(n-1)]}} = \frac{\overline{X}-\mu}{S/\sqrt{n}}$$

has a t distribution with $(n-1)$ degrees of freedom, irrespective of the value of σ^2. Thus, $\left(\overline{X}-\mu\right)/\left(S/\sqrt{n}\right)$ can be used as a pivot. Hence, for n small ($n < 30$) and σ^2 unknown, we have the following result.

Theorem 5.5.1 *If* $\overline{X}$ *and* S *are the sample mean and the sample standard deviation of a random sample of size* n *from a normal population, then:*

$$\overline{X} - t_{\alpha/2,n-1}\frac{S}{\sqrt{n}} < \mu < \overline{X} + t_{\alpha/2,n-1}\frac{S}{\sqrt{n}}$$

is a $(1-\alpha)100\%$ *CI for the population mean* μ.

Note that if the confidence coefficient, $1-\alpha$, and $\overline{X}$ and S remain the same, the confidence range $CR = \widehat{\theta}_U - \widehat{\theta}_L$ decreases as the sample size n increases, which means that we are closing in on the true parameter value of θ.

One can use the following procedure to find the CI for the mean when a small sample is from an approximately normal distribution.

Procedure to find small-sample confidence interval for μ

1. Calculate the values of $\overline{X}$ and S from the sample $X_1, \ldots, X_n$.
2. Using the t table, select two tail values, $-t_{\alpha/2}$ and $t_{\alpha/2}$.
3. The $(1-\alpha)100\%$ CI for μ is:

$$\left(\overline{X} - t_{\alpha/2,n-1}\frac{S}{\sqrt{n}}, \overline{X} + t_{\alpha/2,n-1}\frac{S}{\sqrt{n}}\right)$$

that is, $P\left(\overline{X} - t_{\alpha/2,n-1}\frac{S}{\sqrt{n}} \le \mu \le \overline{X} + t_{\alpha/2,n-1}\frac{S}{\sqrt{n}}\right) = 1-\alpha.$

4. *Conclusion*: We are $(1-\alpha)100\%$ confident that the true parameter μ lies in the interval $\left(\overline{X} - t_{\alpha/2,n-1}(S/\sqrt{n}), \overline{X} + t_{\alpha/2,n-1}(S/\sqrt{n})\right)$.
5. **Assumption:** The population is normal.

In practice, the first step in the previous procedure should include a test of normality (see Project 4C). A built-in test of normality is available in most of the statistical software packages. In Example 5.5.9, we show how this test is utilized. Even when the data fail the normality test, most statistical software will produce a CI based on normality or give an error report. We should understand that generally such answers are meaningless. In those cases, nonparametric methods (Chapter 12) such as the Wilcoxon rank sum method or bootstrap method (Chapter 13) will be more appropriate. For more discussion, refer to Section 14.4.1.

EXAMPLE 5.5.7

The following is a random data set from a normal population:

7.2 5.7 4.9 6.2 8.5 2.8

Construct a 95% CI for the population mean μ. Interpret.

Solution

The first step is to calculate mean and standard deviation of the sample. We compute as the mean $\bar{x} = 5.883$ *and as standard deviation,* $s = 1.959$*. For 5 degrees of freedom, and for* $\alpha = 0.05$*, from the* t *table,* $t_{0.025} = 2.571$*. Hence, a 95% CI for* μ *is:*

$$\left(\overline{x} - t_{\alpha/2,n-1}\frac{2}{\sqrt{n}}, \overline{x} + t_{\alpha/2,n-1}\frac{2}{\sqrt{n}}\right)$$

$$= \left(5.883 - 2.571\left(\frac{1.959}{\sqrt{6}}\right), 5.5883 + 2.571\left(\frac{1.959}{\sqrt{6}}\right)\right)$$

$$= (3.827,\ 7.939).$$

This can be interpreted as that we are 95% *confident that the true mean* μ *will be between* 3.827 *and* 7.939.

EXAMPLE 5.5.8

The scores of a random sample of 16 people who took the TOEFL (Test of English as a Foreign Language) had a mean of 540 and a standard deviation of 50. Construct a 95% CI for the population mean μ of the TOEFL score, assuming that the scores are normally distributed.

Solution

Because $n = 16$ *is small, using Theorem* 5.5.1 *with degrees of freedom* 15, *a* 95% *CI for* μ *is:*

$$\overline{x} \pm t_{\alpha/2,n-1}\frac{s}{\sqrt{n}} = 540 \pm 2.131\left(\frac{50}{\sqrt{16}}\right).$$

So the 95% *CI for the population mean* μ *of the TOEFL scores is* (513.36, 566.64).

A Dobson unit is the most basic measure used in ozone research. The unit is named after G.M.B. Dobson, one of the first scientists to investigate atmospheric ozone (between 1920 and 1960). He designed the Dobson spectrometer, the standard instrument used to measure ozone from the ground. The data in Example 5.5.9 represent the total ozone levels at randomly selected points on the Earth (represented by the pair (latitude, longitude)) on a particular day.

EXAMPLE 5.5.9

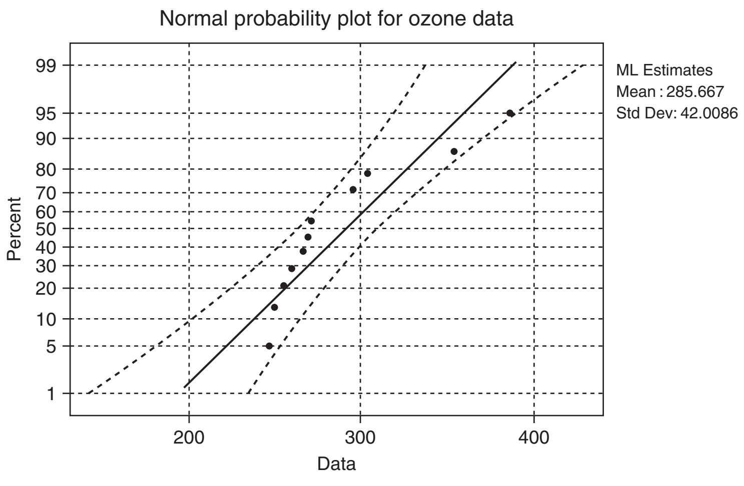
The following data represent the total ozone levels measured in Dobson units at randomly selected locations on the Earth on a particular day:

269	246	388	354	266	303
295	259	274	249	271	254

Can we say that the data are approximately normally distributed? Construct a 95% CI for the population mean μ of ozone levels on this day.

Solution

The following is the probability plot of these data created using Minitab.



Because all the data values lie within the bounds on the normal probability plot (see the discussion in Section 3.2.4), we can assume that the data have approximate normality. We have $\bar{x} = 285.7$ *and* $s = 43.9$. *Also,* $n = 12$. *For* $\alpha = 0.05$, $t_{0.025,\ 11} = 2.201$. *A 95% CI for* μ *is:*

$$\bar{x} \pm t_{\alpha/2,(n-1)}\frac{s}{\sqrt{n}} = 285.7 \pm 2.201\left(\frac{43.9}{\sqrt{12}}\right).$$

Hence, a 95% CI for μ*, the average ozone level over the Earth, lies in* (257.81, 313.59).

EXERCISES 5.5

5.5.1. A survey indicates that it is important to pay attention to truth in political advertising. Based on a survey of 1200 people, 35% indicated that they found political advertisements to be untrue; 60% say that they will not vote for candidates whose advertisements are judged to be untrue; and of this latter group, only 15% ever complained to the media or to the candidate about their dissatisfaction.

(a) Find a 95% CI for the percentage of people who find political advertising to be untrue.

(b) Find a 95% CI for the percentage of voters who will not vote for candidates whose advertisements are considered to be untrue.

(c) Find a 95% CI for the percentage of those who avoid voting for candidates whose advertisements are considered untrue and who have complained to the media or to the candidate about the falsehoods in commercials.

(d) For each case above, interpret the results and state any assumptions you have made.

5.5.2. Many mutual funds use an investment approach involving owning stocks whose price/earnings multiples (P/Es) are less than the P/E of the S&P 500. The following data give P/Es of 49 companies that a randomly selected mutual fund owns in a particular year.

6.8	5.6	8.5	8.5	8.4	7.5	9.3	9.4	7.8	7.1
9.9	9.6	9.0	9.4	13.7	16.6	9.1	10.1	10.6	11.1
8.9	11.7	12.8	11.5	12.0	10.6	11.1	6.4	12.3	12.3
11.4	9.9	14.3	11.5	11.8	13.3	12.8	13.7	13.9	12.9
14.2	14.0	15.5	16.9	18.0	17.9	21.8	18.4	34.3	

Find a 98% CI for the mean P/E multiples. Interpret the result and state any assumptions you have made.

5.5.3. Let $X_1, \ldots, X_n$ be a random sample from an $N(\mu, \sigma^2)$ distribution, with σ^2 known.

(a) Show that $\hat{\mu} = \bar{X}$ is an MLE of the population mean μ.

(b) Show that:

$$P\left(\bar{X} - \frac{2\sigma}{\sqrt{n}} < \mu < \bar{X} + \frac{2\sigma}{\sqrt{n}}\right) = 0.954.$$

(c) Let

$$P\left(\bar{X} - \frac{k\sigma}{\sqrt{n}} < \mu < \bar{X} + \frac{k\sigma}{\sqrt{n}}\right) = 0.90.$$

Find k.

5.5.4. Let the observed mean of a sample of size 45 be $\bar{x} = 68.51$ from a distribution having variance 110. Find a 95% CI for the true mean μ and interpret the result and state any assumptions you have made.

5.5.5. In a random sample of 50 college seniors, 18 indicated that they were planning to pursue a graduate degree. Find a 98% CI for the true proportion of all college seniors planning to pursue a graduate degree, and interpret the result, and state any assumptions you have made.

5.5.6. DVD players coming off an assembly line are automatically checked to make sure they are not defective. The manufacturer wants an interval estimate of the percentage of DVD players that fail the testing procedure. Compute a 90% CI, based on a random sample of size 105 in which 17 DVD players failed the testing procedure. Also, interpret the result and state any assumptions you have made.

5.5.7. Studies have shown that the risk of developing coronary disease increases with the level of obesity, or accumulation of body fat. A study was conducted on the effect of exercise on losing weight. Fifty men who exercised lost an average of 11.4 lb, with a standard deviation of 4.5 lb. Construct a 95% CI for the mean weight loss through exercise. Interpret the result and state any assumptions you have made.

5.5.8. Basing findings on 60 successful pregnancies involving natural birth, an experimenter found that the mean pregnancy term was 274 days, with a standard deviation of 14 days. Construct a 99% CI for the true mean pregnancy term μ.

5.5.9. Let Y be the binomial random variable with parameter p and $n = 400$. If the observed value of Y is $y = 120$, find a 95% CI for p.

5.5.10. For a health screening in a large company, the diastolic and systolic blood pressures of all the employees were recorded. In a random sample of 150 employees, 12 were found to suffer from hypertension. Find 95% and 98% CIs for the proportion of the employees of this company with hypertension.

5.5.11. In a random sample of 500 items from a large lot of manufactured items, there were 40 defectives.

(a) Find a 90% CI for the true proportion of defectives in the lot.

(b) Is the assumption of normal approximation valid?

(c) Suppose we suspect that another lot has the same proportion of defectives as in the first lot. What should be the sample size if we want to estimate the true proportion within 0.01 with 90% confidence?

5.5.12. Pesticide concentrations in sediment from irrigation areas can provide information required to assess the exposure and fate of these chemicals in freshwater ecosystems and their likely impacts on the marine environment. In a study (Müller, J.F., et al., 2000. Pesticides in sediments from Queensland irrigation channels and drains. Mar. Pollut. Bull. 41 (7−12), 294−301), 103 sediment samples were collected from irrigation channels and drains in 11 agricultural areas of Queensland. In 74 of these samples, they detected DDT with concentration levels up to 840 ng/g dw. Obtain a 95% CI for the proportion of total number of sediments with detectable DDT.

5.5.13. Let $\overline{X}$ be the mean of a random sample of size n from an $N(\mu, 16)$ distribution. Find n such that $p\left(\overline{X} - 2 < \mu < \overline{X} + 2\right) = 0.95$.

5.5.14. Let X be a Poisson random variable with parameter λ. A sample of 150 observations from this population has a mean equal to 2.5. Construct a 98% CI for λ.

5.5.15. An opinion poll conducted in March of 1996 by a newspaper (*Tampa Tribune*) among eligible voters with a sample size 425 showed that the president, who was seeking reelection, had 45% support. Give a 95% and a 98% CI for the proportion of support for the president.

5.5.16. A random sample of 100 households located in a large city recorded the number of people living in each household, Y, and the monthly expenditure for food, X. The following summary statistics are given:

$$\sum_{i=1}^{100} Y_i = 340$$

$$\sum_{i=1}^{100} Y_i^2 = 1650$$

$$\sum_{i=1}^{100} X_i = 40,000$$

$$\sum_{i=1}^{100} X_i^2 = 44,000,000$$

(a) Form a 95% CI for the mean number of people living in a household in this city.

(b) Form a 95% CI for the mean monthly food expenses.

(c) For each case just given, interpret the results and state any assumptions you have made.

5.5.17. Let $X_1, \ldots, X_n$ be a random sample from an exponential distribution with parameter θ. A sample of 350 observations from this population has a mean equal to 3.75. Construct a 90% CI for θ.

5.5.18. Suppose a coin is tossed 100 times to estimate $p = P$ (*Heads*). It is observed that heads appeared 60 times. Find a 95% CI for p.

5.5.19. Suppose a population of women at least 35 years of age are pregnant with a fetus affected by Down syndrome. We are interested in testing positive on a noninvasive screening test for fetuses affected by Down syndrome by women at least 35 years of age. In an experiment, suppose 52 of 60 women tested positive. Obtain a 95% CI for the true proportion of women at least 35 years of age who are pregnant with a fetus affected by Down syndrome who will receive positive test results from this procedure.

5.5.20. **(a)** Let $X_1, \ldots, X_n$ be a random sample from a Poisson distribution with parameter λ. Derive a $(1 - \alpha)100\%$ large sample CI for λ.

(b) To date nodes in a phylogenetic tree, the mean path length (MPL) is used in estimating the relative age of a node. The following data represent the MPL for 39 nodes (Britton, T. et al., 2002. Phylogenetic dating with confidence intervals using mean path lengths. Mol. Phylogenet. Evol. 24, 58−65). Assume that the data (given in centimeters) follow a Poisson distribution with parameter λ:

65.2	47.0	38.2	13.5	18.0	25.6	16.3	14.0	23.2	18.8
7.5	13.3	11.0	54.9	22.0	50.1	32.6	26.0	13.0	9.0
7.2	4.7	4.5	41.1	45.8	37.0	8.5	30.5	29.3	13.8
7.7	5.5	24.1	12.5	22.3	19.0	9.5	4.7	3.0	

Obtain a 95% CI for λ and interpret.

5.5.21. A person plans to start an Internet service provider in a large city. The plan requires an estimate of the average number of minutes of Internet use of a household in a week. How many households must be (randomly) sampled to be 95% sure that the sample mean is within 15 minutes of the population mean? Assume that a pilot study estimated the value of $\sigma = 35$ minutes.

5.5.22. The fruit fly *Drosophila melanogaster* normally has a gray color. However, because of a mutation a good portion of them are black. A biologist eager to learn about the effects of mutation wants to collect a random sample to estimate the proportion of black fruit flies of this type within 1% error with 95% confidence.

(a) How many individual flies should the researcher capture if there is no information on the population proportion of black flies?

(b) Suppose the researcher has the initial estimate that 25% of the fruit fly *D. melanogaster* have been affected by this mutation. What is the sample size?

5.5.23. In a pharmacological experiment, 35 lab rats were not given water for 11 hours and were then permitted access to water for 1 hour. The amounts of water consumed (mL/h) are given in the following table:

10.6	13.3	15.5	10.7	9.6	12.1	11.8	10.9	9.9	13.2
9.3	11.7	9.9	13.0	12.3	11.0	13.1	11.0	12.5	13.9
14.1	14.8	15.1	12.8	14.0	7.1	14.1	12.7	9.6	12.5
9.0	12.7	13.6	12.5	12.6					

Obtain a 98% CI for the mean amount of water consumed.

5.5.24. In sociology, a social network is defined as the people you make frequent contact with, say, through Facebook. The personal network size for each adult in a random sample of 3000 adults was calculated. The sample had a mean personal network size of 190 with a known population standard deviation of 25. Find a 95% CI for the mean personal network size of all adults to see if we have a normal amount of friends in our network.

5.5.25. **(a)** How does the t distribution compare with the normal distribution?

(b) How does the difference affect the size of CIs constructed using z (normal approximation) relative to those constructed using the t distribution?

(c) Does sample size make a difference?

(d) What assumptions do we need to make in using the t distribution for the construction of a CI?

5.5.26. Use the t table to determine the values of $t_{\alpha/2}$ that would be used in the construction of a CI for a population mean in each of the following cases:
(a) $\alpha = 0.99$, $n = 20$
(b) $\alpha = 0.95$, $n = 18$
(c) $\alpha = 0.90$, $\mathrm{n} = 25$

5.5.27. Let $X_1, \ldots, X_n$ be a random sample from a normal population. A particular realization resulted in a sample mean of 20 with the sample standard deviation 4. Construct a 95% CI for μ when:
(a) $n = 5$, **(b)** $n = 10$, and **(c)** $n = 25$. What happens to the length of the CI as n changes?

5.5.28. In a large university, the following are the ages of 20 randomly chosen employees:

24	31	28	43	28	56	48	39	52	32
38	49	51	49	62	33	41	58	63	56

Assuming that the data came from a normal population, construct a 95% CI for the population mean μ of the ages of the employees of this university. Interpret your answer.

5.5.29. A random sample of size 26 is drawn from a population having a normal distribution. The sample mean and the sample standard deviation from the data are given, respectively, as $\bar{x} = -2.22$ and $s = 1.67$. Construct a 98% CI for the population mean μ and interpret.

5.5.30. A medication is suspected of causing an elevated heart rate in a certain group of high-risk patients. Twenty patients from the group were given the medication. The changes in heart rates were found to be as follows:

−1	8	5	10	2	12	7	9	1	3
4	6	4	12	11	2	−1	10	2	8

Construct a 98% CI for the mean change in heart rate. Assume that the population has a normal distribution. Interpret your answer.

5.5.31. Ten bearings made by a certain process have a mean diameter of 0.905 cm with a standard deviation of 0.0050 cm. Assuming that the data may be viewed as a random sample from a normal population, construct a 95% CI for the actual average diameter of bearings made by this process and interpret.

5.5.32. Air pollution in large US cities is monitored to see whether it conforms to requirements set by the Environmental Protection Agency. The following data, expressed as an air pollution index, give the air quality of a city for 10 randomly selected days:

57.3	58.1	58.7	66.7	58.6	61.9	59.0	64.4	62.6	64.9

Assuming that the data may be looked upon as a random sample from a normal population, construct a 95% CI for the actual average air pollution index for this city and interpret.

5.5.33. To find the average hemoglobin (Hb) level in children with chronic diarrhea, a random sample of 10 children with chronic diarrhea is selected from a city and their Hb levels (g/dL) are obtained as follows:

12.3	11.4	14.2	15.3	14.8	13.8	11.1	15.1	15.8	13.2

Assuming that the data may be looked upon as a random sample from a normal population, construct a 99% CI for the actual average Hb level in children with chronic diarrhea for this city and interpret. Draw a box plot and normal plot for these data, and comment.

5.5.34. Suppose that you need to estimate the mean number of typographical errors per page in the rough draft of a 400-page book. A careful examination of 10 pages gives an average of six errors per page with a standard deviation of two errors. Assuming that the data may be looked upon as a random sample from a normal population, construct a 99% CI for the actual average number of errors per page in this book and interpret. In this problem, is the normal model appropriate?

5.5.35. Creatine kinase (CK) is found predominantly in muscle and is released into the circulation from muscular lesions. Therefore, serum CK activity has been theoretically expected to be useful as a marker in exercise physiology and sports medicine for the detection of muscle injury and overwork. The following data represent the peak CK activity (measured in IU/L) after 90 min of exercise in 15 healthy young men (Totsuka, M., et al., Break point of serum creatine kinase release after endurance exercise. http://jap.physiology.org/cgi/content/full/93/4/1280):

1112	722	689	251	196	185	128	102	166	178
775	694	514	244	208					

Construct a 95% CI for the mean peak CK activity.

5.5.36. A random sample of 20 observations gave the following summary statistics: $\sum x_i = 234$ and $\sum x_i^2 = 3048$. Assuming that the data may be looked upon as a random sample from a normal population, construct a 95% CI for the actual average, μ.

5.5.37. Let a random sample of size 17 from a normal population for which both mean μ and variance σ^2 are unknown yield $\bar{x} = 3.12$ and $s^2 = 1.04$. Determine a 99% CI for μ.

5.5.38. A random sample from a normal population yields the following 25 values:

90	87	121	96	106	107	89	107	83	92
117	93	98	120	97	109	78	87	99	79
104	85	91	107	89					

(a) Calculate an unbiased estimate $\widehat{\theta}$ of the population mean.
(b) Give an approximate 99% CI for the population mean.

5.5.39. The following are random data from a normal population:

3.3	3.3	4.7	2.6	6.4	4.7	1.7	4.5	5.0	3.0

Construct a 98% CI for the population mean μ.

5.5.40. The following data represent the rates (micrometers per hour) at which a razor cut made in the skin of anesthetized newts is closed by new cells:

28	20	21	39	32	23	18	31	14	23
18	22	28	24	33	12	23	21	25	25

(a) Can we say that the data are approximately normally distributed?
(b) Find a 95% CI for population mean rate μ for the new cells to close a razor cut made in the skin of anesthetized newts.
(c) Find a 99% CI for μ.
(d) Is the 95% CI wider or narrower than the 99% CI? Briefly explain why.

5.5.41. For a particular car, when the brake is applied at 62 mph, the following data give stopping distance (in feet) for 10 random trials on a dry surface (source: http://www.nhtsa.dot.gov/cars/testing/brakes/b.pdf):

146.9	148.4	149.4	148.6	150.3
147.5	147.5	149.3	148.4	145.5

(a) Can we say that the data are approximately normally distributed?
(b) Find a 95% CI for population mean stopping distance μ.

5.5.42. A pharmaceutical company tested a new medicine to be marketed for the treatment of a particular type of virus. To obtain an estimate of the mean recovery time, this medicine was tested on 15 volunteer patients, and the recovery time (in days) was recorded. The following data were obtained:

8	17	10	6	34	11	13	6	9	8
19	4	12	17	7					

(a) Obtain a 95% CI estimate of the mean recovery.
(b) What assumptions do we need to make? Test for these assumptions.

5.6 A confidence interval for the population variance

In this section we derive a CI for the population variance σ^2 based on the chi-square distribution (χ^2 distribution). Recall that the χ^2 distribution, like the Student t distribution, is indexed by a parameter called the degrees of freedom. However, the χ^2 distribution is not symmetric and covers positive values only, and hence, it cannot be used to describe a random variable that assumes negative values. Let $X_1, \ldots, X_n$ be normally distributed with mean μ and variance σ^2, with both μ and σ unknown. We know that:

$$\frac{\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2}{\sigma^2}=\frac{(n-1)S^2}{\sigma^2}$$

has a χ^2 distribution with $(n-1)$ degrees of freedom irrespective of σ^2. Hence, it can be used as a pivot. We now find two numbers, χ^2_L and χ^2_U, such that:

$$P\left(\chi^2_L\le\frac{(n-1)S^2}{\sigma^2}\le\chi^2_U\right)=1-\alpha.$$

The foregoing inequality can be rewritten as:

$$P\left(\frac{(n-1)S^2}{\chi^2_U}\le\sigma^2\le\frac{(n-1)S^2}{\chi^2_L}\right)=1-\alpha.$$

Hence, a $(1-\alpha)100\%$ CI for σ^2 is given by $\left((n-1)S^2\big/\chi^2_U,(n-1)S^2\big/\chi^2_L\right)$ For convenience, we take the areas to the right of $\chi^2_U=\chi^2_{\alpha/2}$ and to the left of $\chi^2_L=\chi^2_{1-\alpha/2}$ to be both equal to $\alpha/2$; see Fig. 5.6. Using the chi-square table we can find the values of $\chi^2_{\alpha/2}$ and $\chi^2_{1-\alpha/2}$. Then, we have the following result.

Theorem 5.6.1 *If $\overline{X}$ and S are the mean and standard deviation of a random sample of size n from a normal population, then:*

$$P\left(\frac{(n-1)S^2}{\chi^2_{\alpha/2}}\le\sigma^2\le\frac{(n-1)S^2}{\chi^2_{1-\alpha/2}}\right)=1-\alpha,$$

where the χ^2 distribution has $(n-1)$ degrees of freedom.

That is, we are $(1-\alpha)100\%$ confident that the population variance σ^2 falls in the interval $\left((n-1)S^2\big/\chi^2_{\alpha/2},\ (n-1)S^2\big/\chi^2_{1-\alpha/2}\right)$

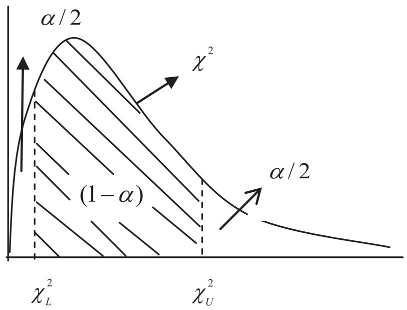


FIGURE 5.6 Chi-square density with equal area on both sides of the confidence interval.

EXAMPLE 5.6.1

A random sample of size 21 from a normal population gave a standard deviation of 9. Determine a 90% CI for σ^2.

Solution

Here $n=21$ *and* $s^2=81$. *From the* χ^2 *table with* 20 *degrees of freedom,* $\chi^2_{0.05}=31.4104$ *and* $\chi^2_{0.95}=10.8508$. *Therefore, a* 90% *CI for* σ^2 *is obtained from:*

$$\left(\frac{(n-1)S^2}{\chi^2_{\alpha/2}}, \frac{(n-1)S^2}{\chi^2_{1-\alpha/2}}\right).$$

Thus, we get:

$$\frac{(20)(81)}{31.4104} < \sigma^2 < \frac{(20)(81)}{10.8508},$$

or we are 90% *confident that* $51.575 < \sigma^2 < 149.298$.

We can summarize the steps for obtaining the CI for the true variance as follows.

Procedure to find confidence interval for σ^2

1. Calculate $\overline{x}$ and s^2 from the sample $x_1, \ldots, x_n$.
2. Find $\chi^2_U = \chi^2_{\alpha/2}$, and $\chi^2_L = \chi^2_{1-\alpha/2}$ using the χ^2 square table with $(n-1)$ degrees of freedom.
3. Compute the $(1-\alpha)100\%$ CI for the population variance σ^2 as $\left((n-1)s^2 \Big/ \chi^2_{\alpha/2}, (n-1)s^2 \Big/ \chi^2_{1-\alpha/2}\right)$, where the χ^2 values are with $(n-1)$ degrees of freedom.

Assumption: The population is normal.

EXAMPLE 5.6.2

The following data represent cholesterol levels (in mg/dL) of 10 randomly selected patients from a large hospital on a particular day:

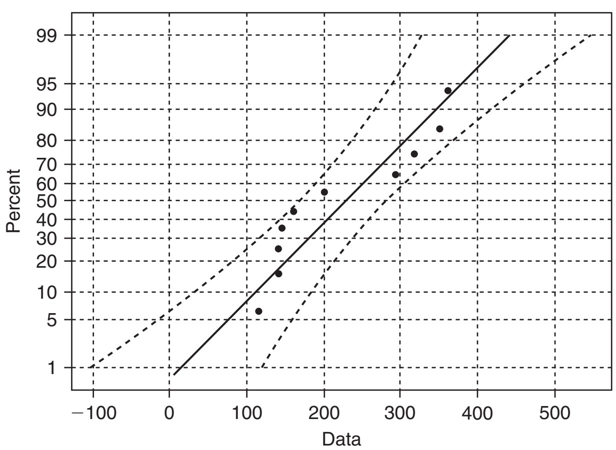
360 352 294 160 146 142 318 200 142 116

Determine a 95% CI for σ^2.

Solution

From the data, we can get $\overline{x} = 223$ *and standard deviation* $s = 96.9$. *The following probability graph is obtained via Minitab.*

Normal probability plot for cholesterol levels



Even though the scattergram does not appear to follow a straight line, the data are still within the band, so we can assume approximate normality for the data. (In situations like this, it is more appropriate to use nonparametric tests explained in Chapter 12.) A box plot of the data shows that there are no outliers. From the χ^2 *table,* $\chi^2_{0.025}(9) = 19.023$ *and* $\chi^2_{0.975}(9) = 2.70$. *Therefore a* 90% *CI for* σ^2 *is obtained from:*

$$\left(\frac{(n-1)S^2}{\chi^2_{\alpha/2}(n-1)}, \frac{(n-1)S^2}{\chi^2_{1-\alpha/2}(n-1)}\right).$$

Thus, we get:

$$\frac{(9)(96.9)^2}{19.023} < \sigma^2 < \frac{(9)(96.9)^2}{2.70},$$

or we are 95% confident that $4442.3 < \sigma^2 < 31{,}299$. *Note that the numbers look very large, but it is the value of variance. By taking the square root of the numbers on the both sides, we can also get a CI for the standard deviation* σ.

As remarked in the previous exercise, in general, to find a $(1-\alpha)100\%$ *CI for the true population standard deviation,* σ, *take the square roots of the end points of the CI of the variance.*

EXERCISES 5.6

5.6.1. A random sample of size 20 is drawn from a population having a normal distribution. The sample mean and the sample standard deviation from the data are given, respectively, as $\bar{x} = -2.2$ and $s = 1.42$. Construct a 90% CI for the population variance σ^2 and interpret.

5.6.2. A medicine is suspected of causing an elevated heart rate in a certain group of high-risk patients. Twenty patients from the group were given the medicine. The changes in heart rates were found to be as follows:

−1	8	5	10	2	12	7	9	1	3
4	6	4	12	11	2	−1	10	2	8

Construct a 95% CI for the variance of change in heart rate. Assume that the population has a normal distribution and interpret.

5.6.3. Air pollution in large US cities is monitored to see whether it conforms to requirements set by the Environmental Protection Agency. The following data, expressed as an air pollution index, give the air quality of a city for 10 randomly selected days:

56.23	57.12	57.7	65.80	59.40
62.90	58.00	64.56	63.92	63.45

Assuming that the data may be viewed as a random sample from a normal population, construct a 99% CI for the actual variance of the air pollution index for this city and interpret.

5.6.4. A random sample of 25 observations gave the following summary statistics: $\sum x_i = 234$ and $\sum x_i^2 = 3048$. Assuming that the data can be looked upon as a random sample from a normal population, construct a 95% CI for the actual variance, σ^2.

5.6.5. Let a random sample of size 18 from a normal population with both mean μ and variance σ^2 unknown yield $\bar{x} = 2.27$ and $s^2 = 1.02$. Determine a 99% CI for σ^2.

5.6.6. Suppose we want to study contaminated fish in a river. It is important for the study to know the size of the variance σ^2 in the fish weights. The 25 samples of fish in the study produced the following summary statistics: $\bar{x} = 1030.5g$, and standard deviation $s = 200.6$ g. Construct a 95% CI for the true variation in weights of contaminated fish in this river.

5.6.7. A random sample from a normal population yields the following 25 values:

90	87	121	96	106	107	89	107	83	92
117	93	98	120	97	109	78	87	99	79
104	85	91	107	89					

(a) Calculate an unbiased estimate $\widehat{\sigma}^2$ of the population variance.
(b) Give approximate 99% CI for the population variance.
(c) Interpret your results and state any assumptions you made to solve the problem.

5.6.8. It is known that some brands of peanut butter contain impurities within an acceptable level. A test conducted on 11 randomly selected jars of a certain brand of peanut butter resulted in the following percentages of impurities:

1.9	2.7	2.1	2.8	2.3	3.6	1.4	1.8	2.1	3.2	2.0

Construct a 95% CI for the average percentage of impurities in this brand of peanut butter.
Give an approximate 95% CI for the population variance.
Interpret your results and test for normality.

5.6.9. The following data represent the maximal head measurements (across the top of the skull) in millimeters of 15 Etruscans (inhabitants of ancient Etruria):

152	147	126	140	135	139	149	140
142	147	132	148	146	143	137	

Calculate an unbiased estimate $\widehat{\sigma}^2$ of the population variance.
Give approximate 95% CI for the population variance.
Interpret your results and test for normality.

5.6.10. The rates of return (rounded to the nearest percentage) for 25 clients of a financial firm are given in the following table:

13	11	28	6	−4	15	13	6	11	11
3	12	20	3	16	16	15	8	20	15
4	1	12	2	−9					

Find a 98% CI for the variance σ^2 of rates of return. Use this to find the CI for the population standard deviation, σ.

5.6.11. To test the precision of a new type of blood sugar monitor for diabetic patients, 20 randomly selected monitors of this type were used. A blood sample with 120 mg/dL was tested in each of these monitors, and the resulting readings are given in the following table:

117	116	121	120	122	117	120	120	118	119
118	123	119	123	119	122	118	122	121	120

(a) Obtain a 99% CI for the variance σ^2.
(b) Is it reasonable to assume that the data follow a normal distribution?

5.7 Confidence interval concerning two population parameters

In the earlier sections we studied the confidence limits of true parameters from samples from a single population. Now, we consider the interval estimation based on samples from two populations. Our aim is to obtain a CI for the parameters of interest based on two independent samples taken from these two populations.

Let $X_{11}, \ldots, X_{1n_1}$ be a random sample from a normal distribution with mean μ_1 and variance σ_1^2, and let $X_{21}, \ldots, X_{2n_2}$ be a random sample from a normal distribution with mean μ_2 and variance σ_2^2. Let $\overline{X}_1 = (1/n_1)\sum_{i=1}^{n_1} X_{1i}$ and $\overline{X}_2 = (1/n_2)\sum_{i=1}^{n_2} X_{2i}$. We will assume that the two samples are independent. Then $\overline{X}_1$ and $\overline{X}_2$ are independent. The distribution of $\overline{X}_1 - \overline{X}_2$ is $N\left(\mu_1 - \mu_2, (1/n_1)\sigma_1^2 + (1/n_2)\sigma_2^2\right)$. Now, as in the one-sample case, the CI for $\mu_1 - \mu_2$ is obtained as follows.

Large-sample confidence interval for the difference of two means

(i) σ_1, σ_2 are known. The $(1-\alpha)100\%$ large sample CI for $\mu_1 - \mu_2$ is given by:

$$\left(\overline{X}_1 - \overline{X}_2\right) \pm z_{\alpha/2}\sqrt{\left(\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)}.$$

(ii) If σ_1 and σ_2 are not known, σ_1 and σ_2 can be replaced by the respective sample standard deviations S_1 and S_2 when $n_i \geq 30$, $i = 1, 2$. Thus, we can write:

$$p\left(\left(\overline{X}_1 - \overline{X}_2\right) - z_{\alpha/2}\sqrt{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)} \leq \mu_1 - \mu_2 \leq \left(\overline{X}_1 - \overline{X}_2\right) + z_{\alpha/2}\sqrt{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)}\right) = 1 - \alpha.$$

Assumptions: The population is normal, and the samples are independent.

EXAMPLE 5.7.1

A study of two kinds of machine failures shows that 58 failures of the first kind took an average of 79.7 minutes to repair with a standard deviation of 18.4 minutes, whereas 71 failures of the second kind took on average 87.3 minutes to repair with a standard deviation of 19.5 minutes. Find a 99% CI for the difference between the true average amounts of time it takes to repair failures of the two kinds of machines.

Solution

Here, $n_1 = 58$, $n_2 = 71$, $\bar{x}_1 = 79.7$, $s_1 = 18.4$, $\bar{x}_2 = 87.3$, *and* $s_2 = 19.5$. *Then the 99% CI for* $\mu_1 - \mu_2$ *is given by:*

$$(79.7 - 87.3) \pm 2.575\sqrt{\frac{(18.4)^2}{58} + \frac{(19.5)^2}{71}}.$$

That is, we are 99% certain that $\mu_1 - \mu_2$ *is located in the interval* $(-16.215, 1.0149)$. *Note that* $-16.215 < \mu_1 - \mu_2 < 1.0149$ *means that more than 90% of the length of this interval is negative. Thus, we can conclude that* μ_2 *dominates* μ_1, *that is,* $\mu_2 > \mu_1$ *more than 90% of the time.*

In the small-sample case, the problem of constructing CIs for the difference of the means from the two normal populations with unknown variances can be a difficult one. However, if we assume that the two populations have a common but unknown variance, say $\sigma_1^2 = \sigma_2^2 = \sigma^2$, we can obtain an estimate of the variance by pooling the two sample data sets. Define the pooled sample variance S_p^2 as:

$$S_p^2 = \frac{\sum_{i=1}^{n_1}\left(X_{1i} - \overline{X}_1\right)^2 + \sum_{i=1}^{n_2}\left(X_{2i} - \overline{X}_2\right)^2}{n_1 + n_2 - 2}$$

$$= \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}.$$

Now, when the two samples are independent,

$$T = \frac{\left(\overline{X}_1 - \overline{X}_2\right) - (\mu_1 - \mu_2)}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

has a t distribution with $n_1 + n_2 - 2$ degrees of freedom. We summarize the CI for $\mu_1 - \mu_2$ below.

Small-sample confidence interval for the difference of two means $\left(\sigma_1^2 = \sigma_2^2\right)$

The small-sample $(1 - \alpha)100\%$ CI for $\mu_1 - \mu_2$ is:

$$\left(\overline{X}_1 - \overline{X}_2\right) \pm t_{\alpha/2,(n_1+n_2-2)} S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}.$$

Assumption: The samples are independent from two normal populations with equal variances.

EXAMPLE 5.7.2

Independent random samples from two normal populations with equal variances produced the following data:

Sample 1:	1.2	3.1	1.7	2.8	3
Sample 2:	4.2	2.7	3.6	3.9	

(a) Calculate the pooled estimate of σ^2.

(b) Obtain a 90% CI for $\mu_1 - \mu_2$.

Solution

(a) *We have* $n_1 = 5$ *and* $n_2 = 4$. *Also,*

$$\overline{x}_1 = 2.36, \quad s_1^2 = 0.733$$
$$\overline{x}_2 = 3.6, \quad s_2^2 = 0.42.$$

Hence,

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} = 0.5989.$$

(b) *For the confidence coefficient 0.90,* $\alpha = 0.10$, *and from the* t *table,* $t_{0.05,7} = 1.895$. *Thus, a 90% CI for* $\mu_1 - \mu_2$ *is:*

$$\left(\overline{X}_1 - \overline{X}_2\right) \pm t_{\alpha/2,(n_1+n_2-2)s_p}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

$$= (2.36 - 3.6) \pm 1.895\sqrt{0.5989\left(\frac{1}{5} + \frac{1}{4}\right)}$$

$$= -1.24 \pm 0.98 = (-2.22, -0.26).$$

Here, μ_2 *dominates* μ_1 *uniformly. Note that we can decrease the confidence range,* -2.22 *to* 0.26, *by increasing* n_1 *and* n_2 *with* $1 - \alpha = 0.90$ *to remain the same. This means that we are closing on the unknown true value of* $\mu_1 - \mu_2$.

In the small-sample case, if the equality of the variances cannot be reasonably assumed, that is, $\sigma_1^2 \neq \sigma_2^2$, we can still use the previous procedure, except that we use the following degrees of freedom in obtaining the t value from the table. Let

$$v = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2 - 1}}.$$

The number given in this formula is always rounded down for the degrees of freedom. Hence, in this case, a small-sample $(1 - \alpha)100\%$ CI for $\mu_1 - \mu_2$ is given by:

$$\left(\overline{X}_1 - \overline{X}_2\right) \pm t_{\alpha/2,v}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}},$$

where the t distribution has v degrees of freedom as given previously.

EXAMPLE 5.7.3

Assume that two populations are normally distributed with unknown and unequal variances. Two independent samples are taken with the following summary statistics:

$$n_1 = 16 \quad \overline{x}_1 = 20.17 \quad s_1 = 4.3$$
$$n_2 = 11 \quad \overline{x}_2 = 19.23 \quad s_2 = 3.8$$

Construct a 95% CI for $\mu_1 - \mu_2$.

Solution

First let us compute the degrees of freedom,

$$v=\frac{\left(\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1-1}+\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1-1}}=\frac{\left(\frac{(4.3)^2}{16}+\frac{(3.8)^2}{11}\right)}{\frac{\left(\frac{(4.3)^2}{16}\right)^2}{15}+\frac{\left(\frac{(3.8)^2}{11}\right)^2}{110}}=23.312.$$

Hence, $v = 23$, and $t_{0.025,23} = 2.069$.
Now a 95% *CI for* $\mu_1 - \mu_2$ *is:*

$$(\overline{x}_1-\overline{x}_2)\pm t_{\alpha/2,v}\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}=(20.17-19.23)$$

$$\pm(2.069)\sqrt{\frac{(4.3)^2}{16}+\frac{(3.8)^2}{11}}$$

which gives the 95% CI as:

$$-2.3106<\mu_1-\mu_2<4.1906.$$

In a real-world problem, how do we determine if $\sigma_1^2 = \sigma_2^2$, or $\sigma_1^2 \neq \sigma_2^2$, so that we can select one of the two methods just given? In Chapter 14, we discuss a procedure that determines the homogeneity of the variances (i.e., whether $\sigma_1^2 = \sigma_2^2$). For the time being a good indication is to look at the point estimators of σ_1^2 and σ_2^2, namely, S_1^2 and S_2^2. If the point estimators are fairly close to each other, then we can select $\sigma_1^2 = \sigma_2^2$. Otherwise, $\sigma_1^2 \neq \sigma_2^2$. For a more general method of testing for equality of variances, we refer to Section 14.4.3.

We now give a procedure for a large-sample CI for the difference of the true proportions, $p_1 - p_2$, in two binomial distributed populations.

Large-sample confidence interval for $p_1 - p_2$

The $(1-\alpha)100\%$ large-sample CI for $p_1 - p_2$ is given by:

$$\left(\widehat{p}_1-\widehat{p}_2\right)\pm z_{\alpha/2}\sqrt{\left(\frac{\widehat{p}_1(1-\widehat{p}_1)}{n_1}+\frac{\widehat{p}_2(1-\widehat{p}_2)}{n_2}\right)},$$

where $\widehat{p}_1$ and $\widehat{p}_2$ are the point estimators of p_1 and p_2. This approximation is applicable if $\widehat{p}_i n_i \geq 5, i = 1, 2$ and $\left(1-\widehat{p}_i\right)n_i \geq 5, i = 1, 2$. The two samples are independent.

EXAMPLE 5.7.4

Iron deficiency, the most common nutritional deficiency worldwide, has negative effects on work capacity and on motor and mental development. In a 1999–2000 survey by the National Health and Nutrition Examination Survey, iron deficiency was detected in 58 of 573 white, non-Hispanic females (10% rounded to whole number) and 95 of 498 (19% rounded to whole number) black, non-Hispanic females (source: http://www.cdc.gov/mmwr/preview/mmwrhtml/mm5140a1.htm). Let p_1 be the proportion of black, non-Hispanic females with iron deficiency and let p_2 be the proportion of white, non-Hispanic females with iron deficiency. Obtain a 95% CI for $p_1 - p_2$.

Solution

Here, $n_1 = 573$ and $n_2 = 498$. *Also,* $\widehat{p}_1 = \frac{58}{573} = 0.10122 \approx 0.1$, *and* $\widehat{p}_2 = \frac{95}{498} = 0.1907 \approx 0.19$. *For* $\alpha = 0.05$, $z_{0.015} = 1.96$.
Hence, a 95% CI for $p_1 - p_2$ *is:*

$$\left(\widehat{p}_1-\widehat{p}_2\right)\pm z_{\alpha/2}\sqrt{\left(\frac{\widehat{p}_1(1-\widehat{p}_1)}{n_1}+\frac{\widehat{p}_2(1-\widehat{p}_2)}{n_2}\right)}$$

$$=(0.1-0.19)\pm(1.96)\sqrt{\frac{(0.1)(0.9)}{573}+\frac{(0.19)(0.81)}{498}}$$

$$=(-0.13232,-0.047685).$$

Here, the true difference of $p_1 - p_2$ *is located in the negative portion of the real line, which tells us that the true proportion of black, non-Hispanic females with iron deficiency is larger than the proportion of white, non-Hispanic females with iron deficiency.*

There are situations in applied problems that make it necessary to study and compare the true variances of two independent normal distributions. For this purpose, we will find a CI for the ratio σ_1^2/σ_2^2 using the F distribution. Let $X_1, ..., X_{n_1}$ and $Y_1, ..., Y_{n_2}$ be independent samples of size n_1 and n_2 from two normal distributions $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$, respectively. Let S_1^2 and S_2^2 be the variances of the two random samples. The CI for the ratio σ_1^2/σ_2^2 is given as follows.

A $(1-\alpha)100\%$ confidence interval for $\frac{\sigma_1^2}{\sigma_2^2}$

A $(1-\alpha)100\%$ CI for σ_1^2/σ_2^2 is given by:

$$\left(\left(\frac{S_1^2}{S_2^2}\right)\left(\frac{1}{F_{n_1-1,n_2-1,1-\alpha/2}}\right), \left(\frac{S_1^2}{S_2^2}\right)\left(\frac{1}{F_{n_1-1,n_2-1,(\alpha/2)}}\right)\right).$$

That is,

$$P\left(\left(\frac{S_1^2}{S_2^2}\right)\left(\frac{1}{F_{n_1-1,n_2-1,1-\alpha/2}}\right) \le \frac{\sigma_1^2}{\sigma_2^2} \le \left(\frac{S_1^2}{S_2^2}\right)\left(\frac{1}{F_{n_1-1,n_2-1,(\alpha/2)}}\right)\right)$$

$$= 1 - \alpha.$$

Assumptions: These two populations are normal, and the samples are independent.

Note that we can also write a $(1-\alpha)100\%$ CI for σ_1^2/σ_2^2 in the form:

$$\left(\left(\frac{S_1^2}{S_2^2}\right)\left(\frac{1}{F_{n_1-1,n_2-1,\ 1-\alpha/2}}\right), \left(\frac{S_1^2}{S_2^2}\right)F_{n_2-1,n_1-1,\ 1-\alpha/2}\right).$$

The following example illustrates how to find the CI for σ_1^2/σ_2^2.

EXAMPLE 5.7.5

Assuming that two populations are normally distributed, two independent random samples are taken with the following summary statistics:

$$\begin{array}{lll} n_1 = 21 & \overline{x}_1 = 20.17 & s_1 = 4.3 \\ n_2 = 16 & \overline{x}_2 = 19.23 & s_2 = 3.8 \end{array}$$

Construct a 95% CI for σ_1^2/σ_2^2.

Solution

Here, $n_1 = 21$*,* $n_2 = 16$*, and* $\alpha = 0.05$*. Using the* F *table, we have:*

$$F_{n_1-1,n_2-1,1-\alpha/2} = F(20, 15, 0.975) = 2.76$$

and

$$F_{n_2-1,n_1-1,\ 1-\alpha/2} = F(15, 20, 0.975) = 2.57.$$

A 95% CI for σ_1^2/σ_2^2 *is:*

$$\left(\left(\frac{S_1^2}{S_2^2}\right)\left(\frac{1}{F_{n_1-1,n_2-1,\ 1-\alpha/2}}\right), \left(\frac{S_1^2}{S_2^2}\right)F_{n_2-1,n_1-1,\ 1-\alpha/2}\right)$$

$$= \left(\left(\frac{(4.3)^2}{(3.8)^2}\right)\left(\frac{1}{2.76}\right), \left(\frac{(4.3)^2}{(3.8)^2}\right)(2.57)\right) = (0.46394,\ 3.2908).$$

That is, we are 95% confident that the ratio of true variance, σ_1^2/σ_2^2*, is located in the interval that implies a 95% CI* (0.46394, 3.2908)*.*

EXERCISES 5.7

5.7.1. A study was conducted to compare two different procedures for assembling components. Both procedures were implemented and run for a month to allow employees to learn each procedure. Then each was observed for 10 days with the following results. Values are number of components assembled per day:

Procedure I	115	101	113	64	104	97	114	96	87	93
Procedure II	86	99	100	78	97	111	102	94	88	99

Construct a 98% CI for the difference in the mean number of components assembled by the two methods. Assume that the data for each procedure are from approximately normal populations with a common variance. Interpret the result.

5.7.2. A study was conducted to see the differences between oxygen consumption rates for male runners from a college who had been trained by two different methods, one involving continuous training for a period of time each day and the other involving intermittent training of about the same overall duration. The means, standard deviations, and sample sizes are shown in the following table:

Continuous training	$n_1 = 15$	$\bar{x}_1 = 46.28$	$s_1 = 6.3$
Intermittent training	$n_2 = 7$	$\bar{x}_2 = 42.34$	$s_2 = 7.8$

If the measurements are assumed to come from normally distributed populations with equal variances, estimate the difference between the population means, with confidence coefficient 0.95, and interpret.

5.7.3. Studies have shown that the risk of developing coronary disease increases with the level of obesity. A study comparing two methods of losing weight, diet alone and exercise alone, was conducted on 87 men over a 1-year period. Forty-two men dieted and lost an average of 16.0 lb over the year, with a standard deviation of 5.6 lb. Forty-five men who exercised lost an average of 10.6 lb, with a standard deviation of 7.9 lb. Construct a 99% CI for the difference in the mean weight loss by these two methods. State any assumptions you made and interpret the result you obtained.

5.7.4. The following information was obtained from two independent samples selected from two normally distributed populations with unknown but equal variances:

Sample 1	14	15	12	13	6	14	11	12	17	19	23		
Sample 2	16	18	12	20	15	19	15	22	20	18	23	12	20

Construct a 95% CI for the difference between the population means and interpret.

5.7.5. In the academic year 2001–02, two random samples of 25 male professors and 23 female professors from a large university produced a mean salary for male professors of \$58,550 with a standard deviation of \$4000; the mean for female professors was \$53,700 with a standard deviation of \$3200. Construct a 90% CI for the difference between the population mean salaries. Assume that the salaries of male and female professors are both normally distributed with equal standard deviations. Interpret the result.

5.7.6. Let the random variables X_1 and X_2 follow binomial distributions that have parameters $n_1 = 100$, $n_2 = 75$. Let $x_1 = 35$ and $x_2 = 27$ be observed values of X_1 and X_2. Let p_1 and p_2 be the true proportions. Determine an appropriate 95% CI for $p_1 - p_2$.

5.7.7. The following information is obtained from two independent samples selected from two populations:

$$n_1 = 40 \quad \bar{x}_1 = 28.4 \quad s_1 = 4.1$$
$$n_2 = 32 \quad \bar{x}_2 = 25.6 \quad s_2 = 4.5$$

(a) What is the MLE of $\mu_1 - \mu_2$?
(b) Construct a 99% CI for $\mu_1 - \mu_2$.

5.7.8. To compare the mean Hb levels of well-nourished and undernourished groups of children, random samples from each of these groups yielded the following summary:

	Number of children	Sample mean	Sample standard deviation
Well nourished	95	11.2	0.9
Undernourished	75	9.8	1.2

Construct a 95% CI for the true difference of means, $\mu_1 - \mu_2$.

5.7.9. In a certain part of a city, the average price of homes in 2000 was \$148,822, and in 2001 it was \$155,908. Suppose these means were based on a random sample of 100 homes in 1997 and 150 homes in 1998 and that the sample standard deviations of sale prices were \$21,000 for 2000 and \$23,000 for 2001. Find a 98% CI for the difference in the two population means.

5.7.10. Two independent samples from a normal population are taken with the following summary statistics:

$$n_1 = 16 \quad \bar{x}_1 = 2.4 \quad s_1 = 0.1$$
$$n_2 = 11 \quad \bar{x}_2 = 2.6 \quad s_2 = 0.5$$

Construct a 95% CI for σ_1^2/σ_2^2.

5.7.11. The following information was obtained from two independent samples selected from two normally distributed populations:

Sample 1	35	36	33	34	27	35	32	33	38	40	44		
Sample 2	37	39	33	41	36	40	36	43	41	39	44	33	41

Construct a 90% CI for σ_1^2/σ_2^2.

5.7.12. The management of a supermarket wanted to study the spending habits of its male and female customers. A random sample of 16 male customers who shopped at this supermarket showed that they spent an average of \$55 with a standard deviation of \$12. Another random sample of 25 female customers showed that they spent \$85 with a standard deviation of \$20.50. Assuming that the amounts spent at this supermarket by all its male and female customers were approximately normally distributed, construct a 90% CI for the ratio of variance in spending for males and females, σ_1^2/σ_2^2.

5.7.13. An experiment is conducted comparing the effectiveness of a new method of teaching algebra for eighth-grade students. Twelve gifted and 12 average students are taught using this method. Their scores on a final exam are shown in the following table:

Average	58	69	55	65	88	52	99	76	45	86	55	79
Gifted	77	86	84	93	77	91	87	95	68	78	74	58

(a) Compute the 95% CI on the difference between the means of the students being taught by this new method.
(b) Construct a 95% CI for the ratio of variance in test scores for average and gifted students, σ_1^2/σ_2^2.
(c) What are the assumptions you made in (a) and (b)? Are these assumptions justified?

5.7.14. Assume that two populations have the same variance σ^2. If a sample of size n_1 produced a variance S_1^2 from population I and a sample of size n_2 produced a variance S_2^2 from population II, show that the pooled variance,

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

is an unbiased estimator of σ^2. Show that $\left(S_1^2 + S_2^2\right)/2$ is also an unbiased estimator of σ^2. Which of the two estimators would you prefer? Give reasons for your choice.

5.8 Chapter summary

In this chapter we have discussed the basic concepts of estimation, both point estimation and interval estimation. Two methods of finding point estimators were described—the method of moments and the method of maximum likelihood. Some desirable properties of the point estimators that we have discussed are unbiasedness and sufficiency. Unbiasedness guards against consistently producing under- or overestimates of the parameter in repeated sampling. A sufficient estimator is a "good" estimator of the population parameter θ in the sense that it depends on fewer data values. Later, this chapter discusses the concept of interval estimation. A $(1-\alpha)100\%$ CI for an unknown parameter θ is computed from sample data. The so-called pivotal method is introduced for deriving a CI. Large-sample and small-sample CIs are derived for population mean μ. CIs in the case of two samples are also discussed. In addition, CIs for variance and ratio of variances are derived.

We will now list some of the key definitions introduced in this chapter.

- Method of moments
- Likelihood function
- Maximum likelihood equations
- Unbiased estimator
- MSE
- MVUE
- Sufficient estimator
- Jointly sufficient
- Upper and lower confidence limits
- Confidence coefficient
- A $(1-\alpha)100\%$ CI for θ
- Interval estimation
- CI

In this chapter, we have also learned the following important concepts and procedures:

- The method of moments procedure
- Procedure to find MLE
- Procedure to verify
- Pivotal method
- Procedure to find a CI for θ using the pivot
- Procedure to find a large-sample CI for θ
- Procedure to find a small-sample CI for μ
- Procedure to find a CI for the population variance σ^2
- Large-sample CI for the difference of the means
- Small-sample CI for the difference of two means $\left(\sigma_1^2 = \sigma_2^2\right)$
- Small-sample CI for the difference of two means $\left(\sigma_1^2 \neq \sigma_2^2\right)$
- Large-sample CI for $p_1 - p_2$
- A $(1-\alpha)100\%$ CI for σ_1^2/σ_2^2

5.9 Computer examples

5.9.1 Examples using R

It should be noted that for the problems where you are generating random samples your answers will vary!

EXAMPLE 5.9.1

Descriptive point estimates

Generate 50 sample points from an $N(4,4)$ distribution and find the descriptive statistics. Obtain an unbiased and sufficient estimate of μ.

R-code

```
sample=rnorm(50,4,4);
summary(sample);
sd(sample);
sd(sample)/sqrt(length(sample));
```

Standard error of the mean

Output

Your output will be unique since the samples are generated randomly; take notice of standard error.

```
Min.     1st Qu.   Median    Mean 3rd   Qu.     Max.
−4.292   1.105     4.012     3.865      6.478   14.790
```

Notice this is an estimate;
we know that the population
mean is 4 as we defined it.

Notice this is an estimate;
we know that the population
mean is 4 as we defined it.

```
4.288085
0.6064268
```

Standard deviation

Standard error of the mean

EXAMPLE 5.9.2

Uniform maximum likelihood

Generate 35 samples from a $U(0,5)$ distribution and, using the descriptive statistics command, find the maximum likelihood estimate for these data.

Solution

We know that for a random sample $X_1, \ldots, X_n$ *from* $U(0, \theta)$*, the MLE* $\widehat{\theta} = max(X_i) = X_{(n)}$*, the nth order statistic. We can use the following steps to obtain the estimate.*

R-code

```
sample=runif(35,0,5);
summary(sample);
```

Output

Your output will be unique since the samples are generated randomly.

```
Min.     1st Qu.   Median   Mean     3rd Qu.   Max.
0.1155   1.5710    2.9520   2.7620   4.0920    4.9900
```

The *MLE* of the data 4.99

EXAMPLE 5.9.3

Confidence interval

Obtain a 95% CI for μ using the following data:

Sample *(x)*: 7.227 5.7383 4.9369 6.238 8.4876 2.7618

This example assumes you have stored your data into variable *x*. Please modify code appropriately.

R-code

```
t.test(x,conf.level=0.95);
```

Output

```
One Sample t-test.
data: x
t = 7.3399, df = 5, p-value = 0.0007365
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
3.832566 7.963967
sample estimates:
mean of x
5.898267
```

Notice the interval
rather than point estimate.

EXAMPLE 5.9.4

Confidence interval

For the following data obtain a 98% CI for μ:

Sample (x): 6.8 5.6 8.5 8.5 8.4 7.5 9.3 9.4 7.8 7.1 9.9 9.6 9.0 13.7 9.4 16.6 9.1 10.1 10.6 11.1 8.9 11.7 12.8 11.5 10.6 12.0 11.1 6.4 12.3 12.3 11.4 9.9 15.5 14.3 11.5 13.3 11.8 12.8 13.7 13.9 12.9 14.2 14.0

This example assumes you have stored the data into variable *x*. Please modify your code appropriately.

R-code

```
t.test(x,conf.level=0.98);
```

Output

```
One Sample t-test
data: x
t = 27.7762, df = 42, p-value < 2.2e-16
alternative hypothesis: true mean is not equal to 0
98 percent confidence interval:
9.910598 11.801030
sample estimates:
mean of x
10.85581
```

Notice the interval rather than point estimate.

EXAMPLE 5.9.5

Confidence interval

For the following data, find a 90% CI for $\mu_1 - \mu_2$ using the following data:

Sample *(x)*: 1.2 3.1 1.7 2.8 3.0

Sample *(y)*: 4.2 2.7 3.6 3.9

This example assumes you have stored your data into variables *x* and *y*. Please modify your code appropriately.

R-code

```
t.test(x,y,conf.level=0.90);
```

Output

```
Welch Two Sample t-test
data: x and y
t = −2.4721, df = 6.996, p-value = 0.04272
alternative hypothesis: true difference in means is not equal to 0
90 percent confidence interval:
−2.1903896 −0.2896104
sample estimates:
mean of x mean of y
      2.36      3.60
```

90% Confidence Interval

5.9.2 Minitab examples

EXAMPLE 5.9.6

Generate 50 sample points from an $N(4, 4)$ distribution and find the descriptive statistics. Obtain an unbiased and sufficient estimate of μ.

Solution

Because we know that the sample mean $\bar{x}$ is an unbiased and sufficient estimate of the population mean μ, we need to find only the sample mean of the generated data.

Calc > ***Random Data*** > ***Normal*** … > Type ***50*** in ***Generate __ rows of data*** > ***Store in column(s):*** type ***C1*** > type in ***Mean: 4.0*** and in ***Standard deviation: 2.0*** > click ***OK***.

EXAMPLE 5.9.7

Generate 35 samples from a $U(0, 5)$ distribution and, using the descriptive statistics command, find the maximum likelihood estimate for these data.

Solution

We know that for a random sample $X_1, \ldots, X_n$ *from* $U(0, \theta)$, *the MLE* $\widehat{\theta} = \max(X_i) = X_{(n)}$, *the* n*th order statistic. We can use the following steps to obtain the estimate.*

Calc > ***Random Data*** > ***Uniform*** ... > Type ***35*** in ***Generate __ rows of data*** > ***Store in column(s):*** type ***C1*** > type in ***Lower end point: 0.0*** and in ***Upper end point: 5.0*** > click ***OK***.

EXAMPLE 5.9.8

(**Small Sample**) Using Minitab, obtain a 95% CI for μ using the following data:

7.227 5.7383 4.9369 6.238 8.4876 2.7618

Solution

Use the following commands.

Enter the data in ***C1.*** *Then,*

Stat > ***Basic Statistics*** > ***1-sample t*** ..., in ***variables:*** enter ***C1***, click ***Confidence interval***, in ***Level*** default value is **95**, if any other value, enter that value, and click ***OK***.

EXAMPLE 5.9.9

(Large Sample) For the data:

6.8	5.6	8.5	8.5	8.4	7.5	9.3	9.4	7.8	7.1	9.9
9.6	9.0	13.7	9.4	16.6	9.1	10.1	10.6	11.1	8.9	11.7
12.8	11.5	10.6	12.0	11.1	6.4	12.3	12.3	11.4	9.9	15.5
14.3	11.5	13.3	11.8	12.8	13.7	13.9	12.9	14.2	14.0	

obtain a 98% CI for μ.

Solution

Enter the data in ***C1.*** *Then click:*

Stat > ***Basic Statistics*** > ***1-Sample Z*** ... >, in ***Variables:*** type ***C1*** > click ***Confidence interval***, and enter ***98*** in ***Level:***> enter ***5*** in ***Sigma:*** > ***OK***.

EXAMPLE 5.9.10

For the following data, find a 90% CI for $\mu_1 - \mu_2$:

Sample 1	1.2	3.1	1.7	2.8	3.0
Sample 2	4.2	2.7	3.6	3.9	

Solution

Enter sample 1 in ***C1*** *and sample 2 in* ***C2.*** *Then click:*

Stat > ***Basic Statistics*** > ***2-Sample t*** ... > click ***Sample*** in different columns > in ***First:*** enter ***C1*** and in ***Second:*** enter ***C2*** > enter ***90*** in ***Confidence Level:*** (if equality of variance can be assumed, click ***Assume equal variances***) > ***OK***.

5.9.3 SPSS examples

EXAMPLE 5.9.11

Consider the data:

66 74 79 80 77 78 65 79 81 69

Using SPSS, obtain a 99% CI for μ.

Solution

One easy way to obtain the CI in SPSS is to use the hypothesis testing procedure. The procedure is as follows: First enter the data in ***C1.*** *Then click:*

Analyze > ***Compare Means*** > ***One-sample t Test*** …, > Move **var00001** to ***Test Variable(s)***, and Click ***Options*** …, and enter ***99*** in ***Confidence interval:***, click ***Continue***, and ***OK***.

Note that the default value is 95%.

5.9.4 SAS examples

We will not give the output in this section.

EXAMPLE 5.9.12

The following data give P/Es for a particular year of 49 mutual fund companies owned by a randomly selected mutual fund:

6.8	5.6	8.5	8.5	8.4	7.5	9.3	9.4	7.8	7.1
9.9	9.6	9.0	16.6	9.1	10.1	10.6	11.1	8.9	11.7
12.8	11.5	12.0	10.6	11.1	6.4	11.4	9.9	14.3	11.5
11.8	13.3	13.9	12.9	14.2	14.0	15.5	17.9	21.8	18.4
34.3	13.7	12.3	18.0	9.4	12.3	16.9	12.8	13.7	

Find a 98% CI for the mean P/E multiples. Use SAS procedures.

Solution

We could use the following procedure.

```
DATA peratio;
INPUT patio @@;
DATALINES;
6.8   5.6   8.5   8.5   8.4   7.5   9.3   9.4   7.8
7.1   9.9   9.6   9.0   9.4   13.7  16.6  9.1   10.1  10.6
11.1  8.9   11.7  12.8  11.5  12.0  10.6  11.1  6.4   12.3
12.3  11.4  9.9   14.3  11.5  11.8  13.3  12.8  13.7  13.9
      12.9
14.2  14.0  15.5  16.9  18.0  17.9  21.8  18.4  34.3
;
PROC MEANS  data = peratio  lclm uclm  alpha  =  0.02;
var peratio;
RUN;
```

EXERCISES 5.9

5.9.1. Using any of the software packages (R, Minitab, SPSS, or SAS), obtain CIs for at least one data set taken from each section of this chapter.

5.10 Projects for Chapter 5

5.10.1 Asymptotic properties

In general, we do not have a single sample with one estimator of the unknown parameter θ. Rather, we will have a general formula that defines an estimator for any sample size. This gives a sequence of estimators of θ:

$$\hat{\theta} = h_n(X_1, \ldots, X_n), \quad n = 1, \ 2, \ldots..$$

In this case, we can define the following asymptotic properties:

(i) The sequence of estimators is said to be *asymptotically unbiased* for θ if $bias(\hat{\theta}_n) \to 0$ as $n \to \infty$.
(ii) Suppose $(\hat{\theta}_n)$ and $\hat{y}_n$ are two sequences of estimators that are asymptotically unbiased for θ. The *asymptotic relative efficiency* of $\hat{\theta}_n$ to $\hat{y}_n$ is defined by:

$$\lim_{n} \frac{Var(\hat{\theta}_n)}{Var(\hat{y}_n)}.$$

(a) Show that $\hat{\theta}_n$ is asymptotically unbiased if and only if:

$$E(\hat{\theta}_n) \to \theta \text{ as } n \to \infty.$$

(b) Let $X_1, \ldots, X_n$ be a random sample from a distribution with unknown mean μ and variance σ^2. It is known that the method of moments estimators for μ and σ^2 are, respectively, the sample mean $\overline{X}$ and $S'^2_n = (1/n) \Sigma_{i=1}^{n} (X_i - \overline{X})^2 = ((n-1)/n) S_n^2$, where S_n^2 is the sample variance.
  (i) Show that S'^2_n is an asymptotically unbiased estimator of σ^2.
  (ii) Show that the asymptotic relative efficiency of S'^2_n to S_n^2 is 1.
  (iii) Show that $MSE\,(S'^2_n) < MSE(S_n^2)$. Thus, (S_n^2) is unbiased but (S'^2_n) has a smaller MSE. However, it should be noted that the difference is very small and approaches zero as n becomes large.

5.10.2 Robust estimation

The estimators derived in this chapter are for particular parameters of a presumed underlying family of distributions. However, if the choice of the underlying family of distributions is based on past experience, there is a possibility that the true population will be slightly different from the model used to derive the estimators. Formally, a statistical procedure is *robust* if its behavior is relatively insensitive to deviations from the assumptions on which it is based. If the behavior of an estimator is taken as its variance, a given estimator may have minimum variance for the distribution used, but it may not be very good for the actual distribution. Hence, it is desirable for the derived estimators to have small variance over a range of distributions. We call such estimators *robust estimators.* The following illustrates how the variance of an estimator can be affected by deviations from the presumed underlying population model.

Consider estimating the mean of a standard normal distribution. Let $X_1, \ldots, X_n$ be a random sample from a standard normal distribution. Suppose the population actually follows a contaminated normal distribution. That is, for $0 \le \delta \le 1$, $(1 - \delta)100\%$ of the observations come from an $N(0, 1)$ distribution and the remaining $(\delta)100\%$ of observations come from an $N(0, 5)$ distribution. We already know that the MVUE of the mean μ of an uncontaminated normal distribution is the sample mean. A less effective alternative would be the sample median.

(a) Conduct a simulation study with sample size n that takes, say, 5000 random samples of 100 observations each. Find the mean and median. Also find the sample variance of each. For various values of δ, say 0.0, 0.01, 0.05, 0.1, 0.2, 0.3, and 0.4, create a table of variances of sample mean and sample variance. Compare the variances as the value of δ increases.
(b) The aim of robust estimation is to derive estimators with variance near that of the sample mean when the distribution is standard normal while having the variance remain relatively stable as δ increases. One such estimator is the $\alpha -$ *trimmed mean.* Let $0 \le \alpha \le 0.5$, and define $k = [n\alpha]$, where $[x]$ is the greatest integer that is less than or equal to x. For the ordered sample, discard the k highest and lowest observations and find the mean of the remaining $n - k$ observations. That is, let $X_{(1)} \le X_{(2)} \le \ldots < X_{(n)}$ be the ordered sample, and define:

$$\overline{X_\alpha} = \frac{X_{(1+k)} \le X_{(2+k)} \le \ldots \le X_{(n+k)}}{n - 2k}.$$

For the values of δ and the samples in (a), compute the mean and the 0.05-, 0.1-, 0.25-, and 0.5-trimmed means. Discuss the robustness.

5.10.3 Numerical unbiasedness and consistency

(a) Run the simulation of a normal experiment with increasing sample size. Numerically show the unbiased and consistent properties of the sample mean. Run the experiment at least up until $n = 1000$.

(b) Repeat the experiment of (a), now with an exponential distribution.

5.10.4 Averaged squared errors

Generate 25 samples of size 40 from a normal population with $\mu = 10$ and $\sigma^2 = 4$. For each of the 25 samples:

(a) Compute: $\bar{x}$, $s^2 = \frac{\sum_{i=1}^{40}(x_i - \bar{x})^2}{39}$, $s_1^2 = \frac{\sum_{i=1}^{40}(x_i - \bar{x})^2}{40}$, and $s_2^2 = \frac{\sum_{i=1}^{40}(x_i - \bar{x})^2}{41}$.

(b) Compute the average squared error (ASE) for each of the estimates s^2, s_1^2, and s_2^2 as follows:

Let $K^{s^2} = \left[\left[\sum_{i=1}^{K}(x_i - \bar{x})^2\right] \Big/ 39\right]$ for $K = 1, 2, \ldots, 25$ and K^{s^2} be the sample variance for the Kth sample. Then, the ASE is:

$$ASE = \frac{\sum_{i=1}^{25}\left(K^{S^2} - \sigma^2\right)^2}{25}.$$

Repeat this procedure for the other two estimators. Compare the three ASEs and check which has the smallest ASE.

(c) Repeat (a) and (b) with a sample size of 15.

5.10.5 Alternate method of estimating the mean and variance

(a) Consider the following alternative method of estimating μ and σ^2. We sample sequentially, and at each stage we compute the estimates of μ and σ^2 as follows:

Let $X_1, \ldots, X_n, X_{n+1}$ be the sample values.

Compute:

$$\overline{X}_n = \frac{\sum_{i=1}^{n} X_i}{n}, \quad \overline{X}_{n+1} = \frac{\sum_{i=1}^{n+1} X_i}{n+1}, S_n^2 = \frac{\sum_{i=1}^{n}\left(X_i - \overline{X}_n\right)^2}{n-1}, \quad \text{and}$$

$$S_{n+1}^2 = \frac{\sum_{i=1}^{n+1}\left(X_i - \overline{X}_n\right)^2}{n}.$$

The sequential procedure is stopped when:

$$\left|S_n^2 - S_{n+1}^2\right| \leq 0.01.$$

This will also determine the sample size.

(b) Compare the sample sizes and estimates in Sections 5.10.4 and 5.10.5(a) to see if the sequential procedure has an advantage over ASEs in 5.10.4.

5.10.6 Newton–Raphson in one dimension

For a given function $g(x)$, suppose we need to solve $g(\theta) = 0$. Using the first-order Taylor expansion, $g(\theta) \approx g(x) + (\theta - x)g'(x)$, where $g'(x) = \frac{dg}{dx}$, and setting $g(\theta) = 0$, we get $\theta \approx x - \frac{g(x)}{g'(x)}$. Thus, starting with an initial guess solution x, the guess is updated by θ using the previous formula. This derivation is the basis for the Newton–Raphson iterative method for obtaining the solution of $g(\theta) = 0$. This is given by:

$$\theta_{(n+1)} = \theta_n - \frac{g(\theta_n)}{g'(\theta_n)}, \quad n \geq 0,$$

where θ_n is the value of θ at the nth iteration, starting with the initial guess, θ_0. For a good approximation of the solution, the choice of θ_0 is important. The convergence of this algorithm cannot be guaranteed.

For the MLE, we want to find a solution to:

$$g(\theta) = \frac{dL}{d\theta} = 0,$$

where $L = L(\theta)$ is the likelihood function of the random sample $X_1, \ldots, X_n$. An iterative algorithm for finding the MLE can be given by:

$$\theta_{(n+1)} = \theta_n - \frac{\frac{dL}{d\theta}(\theta_n)}{\frac{d^2L}{d\theta^2}(\theta_n)}, \quad n \geq 0.$$

Write a computer program to find the MLE of a for a gamma distribution with parameters α and β.

5.10.7 The empirical distribution function

In this project, we use an estimation procedure that estimates the whole distribution function, F, of a random variable X. We now define the empirical distribution.

The *empirical distribution function* for a random sample $X_1, \ldots, X_n$ from a distribution F is the function defined by:

$$F_n(x) = \frac{1}{n}\#\{i,\ 1 \leq i \leq n: X_i \leq x\}.$$

It can be shown that $nF_n(x)$ is a binomial random variable with:

$$E[F_n(x)] = F(x) \text{ and } Var\,[F_n(x)] = \frac{1}{n}F(x)[1 - F(x)].$$

Also, by the strong law of large numbers, for each real number x,

$$\lim_{n \to \infty} F_n\ (x) = F(x) \text{ with probability } 1.$$

One of the tests to determine whether a random sample comes from a specific distribution is the Kolmogorov−Smirnov (K−S) test. The K−S test is based on the maximum distance between the empirical distribution function and the actual cdf of this specific distribution (such as, say, the normal distribution).

Using the method of Project 4A (or using any statistical software), generate 100 sample points from a normal distribution with mean 2 and variance 9. Graph the empirical distribution function for this sample. Compare this graph with the graph of the $N(2, 9)$ distribution.

5.10.8 Simulation of coverage of the small confidence intervals for μ

(a) Generate 25 samples of size 15 from a normal population with $\mu = 10$ and $\sigma^2 = 4$. Using a statistical package (such as Minitab), compute the 95% CIs for each of the samples using the small-sample formula. From your output, determine the proportion of the 25 intervals that cover the true mean $\mu = 10$.

(b) What would you expect if the sample size is increased to 100? Would the width of the interval increase or decrease? Would you expect more or fewer of these intervals to contain the true mean 10? Check your answers with actual computation.

(c) Repeat with 20 samples of size 10.

5.10.9 Confidence intervals based on sampling distributions

If we want to obtain a $(1 - \alpha)100\%$ CI for θ, begin with an estimator $\widehat{\theta}$ of θ and determine its sampling distribution. Now select two probability levels, α_1 and α_2, so that $\alpha = \alpha_1 + \alpha_2$. Generally we let $\alpha_1 = \alpha_2$. Take a sample and calculate the value of $\widehat{\theta}$, say $\widehat{\theta} = k$. Now we need to determine the values of the upper and lower confidence limits. Find a value θ_L such that:

$$p\left(\widehat{\theta} \geq k\right) = \alpha_1$$

and θ_U such that:

$$p\left(\widehat{\theta} \leq k\right) = \alpha_2.$$

Then a $(1-\alpha)100\%$ CI for θ will be:

$$\theta_L < \theta < \theta_U.$$

(a) Let $X_1, \ldots, X_n$ be a random sample from a $U(0,\theta)$ distribution. Obtain a $(1-\alpha)100\%$ CI for θ, using the method of sampling distribution.

(b) Let X have a binomial distribution with parameters n and p. First show that there is no quantity that satisfies the conditions of a pivotal quantity. Then, using the method of sampling distributions, obtain a $(1-\alpha)100\%$ CI for p.

5.10.10 Large-sample confidence intervals: general case

The method of finding a CI for a parameter θ that we described in this chapter depends on our ability to find the pivotal quantity. We have seen that such a quantity may not exist. In those cases, the method of sampling distribution described in the previous project could be used. However, this method can involve some difficult calculations. For large samples, we can utilize the following procedure, which is based on the asymptotic distribution of MLEs. Under fairly general conditions, the MLEs have a limiting distribution that is normal. Also, MLEs are asymptotically efficient. Hence, for a large sample the MLE $\widehat{\theta}$ of θ will have approximately normal distribution with mean θ. Also, if the Cramér–Rao lower bound exists, the limiting variance of $\widehat{\theta}$ will be:

$$\sigma_{\hat{\theta}}^2 = \frac{1}{E\left[\left(\dfrac{\partial \ \ln L}{\partial \theta}\right)^2\right]}.$$

Hence,

$$Z = \frac{\widehat{\theta} - \theta}{\sigma_{\hat{\theta}}} \sim N(0,1).$$

Then a large-sample $(1-\alpha)100\%$ CI is obtained from the probability statement:

$$P\left(-z_{\alpha/2} < \frac{\widehat{\theta} - \theta}{\sigma_{\hat{\theta}}} < z_{\alpha/2}\right) \approx 1 - \alpha.$$

We summarize the procedure to construct large-sample CIs.

1. Determine the MLE, $\widehat{\theta}$, of θ. Also find the MLEs of all other unknown parameters.
2. Obtain the variance $\sigma_{\hat{\theta}}$ (if possible directly, otherwise by using the Cramér–Rao lower bound).
3. In the expression for $\sigma_{\hat{\theta}}$, substitute $\widehat{\theta}$ for θ. Replace all other unknown parameters with its MLE. Let the resulting quantity be denoted by $s_{\hat{\theta}}$.
4. Now construct a $(1-\alpha)100\%$ CI for θ from:

$$\widehat{\theta} - z_{\alpha/2} s_{\hat{q}} < \theta < \widehat{\theta} + z_{\alpha/2} s_{\hat{q}}.$$

(a) Using the foregoing procedure, show that a large-sample $(1-\alpha)100\%$ CI for the parameter p in a binomial distribution based on n trials is:

$$\widehat{p} - z_{\alpha/2}\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}} < p < \widehat{p} + z_{\alpha/2}\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}.$$

(b) Let $X_1, \ldots, X_n$ be a random sample from a normal population with parameters μ and σ^2. Derive a large-sample CI for σ^2 using the above procedure.

(c) Let $X_1, \ldots, X_n$ be a random sample from a population with a pdf:

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

Derive a large-sample CI for θ.

5.10.11 Prediction interval for an observation from a normal population

In many cases, we may be interested in predicting future observations from a population, rather than making an inference. A $(1-\alpha)100\%$ prediction interval for a future observation X is an interval of the form (X_L, X_U) such that $P(X_L < X < X_U) = 1-\alpha$. Similar to CIs, we can also define one-sided prediction intervals. Assume that the population is normal with known variance σ^2. Let $X_1, \ldots, X_n$ be a random sample from this population. Then the sampling distribution of the difference $X-\overline{X}$ (we use $\overline{X}$ to denote $\overline{X}_n$) is normal with mean 0 and variance $\sigma^2+\sigma^2_{\overline{X}} = (1+(1/n))\sigma^2$. Then a $(1-\alpha)100\%$ prediction interval for X is given by:

$$\left(\overline{X} - z_{\alpha/2}\sqrt{\left(1+\frac{1}{n}\right)\sigma^2},\ \overline{X} + z_{\alpha/2}\sqrt{\left(1+\frac{1}{n}\right)\sigma^2}\right).$$

Thus, we are $(1-\alpha)100\%$ confident that the next observation, X_{n+1}, will lie in this interval. As in CIs, if the sample size is large, replace σ by sample standard deviation s.

In cases where both μ and σ are not known, and the sample size is small (so that the CLT cannot be applied), it can be shown that $\left[\left(X_{n+1}-\overline{X}_n\right)/\left(S_n\sqrt{1+(1/n)}\right)\right]$ has a t distribution with $(n-1)$ degrees of freedom. Thus, a $(1-\alpha)100\%$ prediction interval for X_{n+1} is given by:

$$\left(\overline{X} - t_{\alpha/2,n-1}\sqrt{(1+(1/n))S^2},\ \overline{X} + t_{\alpha/2,n-1}\sqrt{(1+(1/n))S^2}\right).$$

A standard measure of the capacity of lungs to expel air in breathing is called forced expiratory volume (FEV). The FEV1 is the volume exhaled during the first second of a forced expiratory maneuver started from the level of total lung capacity. The following data (M. Bland, *An Introduction to Medical Statistics*, Oxford University Press, 1995) represent FEV measurements (in liters) from 57 male medical students:

4.47	3.10	4.50	4.90	3.50	4.14	4.32	4.80	3.10	4.68
4.47	3.57	2.85	5.10	5.20	4.80	5.10	4.30	4.70	4.08
3.48	4.20	3.70	5.30	4.71	4.10	4.30	3.39	3.69	4.44
5.00	4.50	4.20	4.16	3.70	3.83	3.90	4.47	3.30	5.43
3.42	3.60	3.20	4.56	4.78	3.60	3.96	3.19	2.85	3.04
3.78	3.75	4.05	3.54	4.14	2.98	3.54			

Obtain a 95% prediction interval for a future observation X_{n+1}.

5.10.12 Empirical distribution function as estimator for cumulative distribution function

In Chapter 3, we saw that probabilistically, the cdf defined as $F(x) = P(X \le x),$ for all $x \in (-\infty, \infty)$. The question is, given a random sample $X_1, \ldots, X_n$ with common cdf $F(x)$, how do we estimate the cdf, $F(x)$? The empirical distribution function (EDF) is the "data analogue" of cdf of a random variable. The EDF is defined as:

$$\widehat{F}_n(x) = \frac{\text{number of elements in the sample } \le \text{x}}{n} = \frac{1}{n}\sum_{i=1}^{n} I_{(X_i \le x)},$$

where I_A is the indicator of event A. Note that EDF is a step function that jumps $\frac{1}{n}$ at each $X_i = x_i$. We can see that EDF describes the data in more detail than the histogram. Using the strong law of large numbers, it can be shown that $\widehat{F}_n(x) \to F(x)$, with probability 1. We refer the reader to look for other interesting properties of EDF.

Using R, create 100 data values from each of the uniform and normal distributions. Draw the theoretical cdf and EDF on the same graph for each of the distributions, respectively.