7.2.3 **THE SAMPLE VARIANCE**

Let $X_1, X_2, \ldots, X_n$ be a random sample from a distribution with mean μ and variance σ^2. Let $\bar{X}$ and

Definition The statistic S^2, defined by

$$S^2 = \frac{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2}{n-1} \tag{7.26}$$

is called the sample variance. $S = \sqrt{S^2}$ is called the *sample standard deviation*.

To compute $E\left[S^2\right]$, we use an identity: For any numbers $x_1, x_2, \ldots, x_n$:

$$\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2 = \sum_{i=1}^{n} x_i^2 - n\bar{x}^2 \tag{7.27}$$

where $\bar{x} = \sum_{i=1}^{n} x_i / n$. It follows from this identity that

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

Taking expectations of both sides of the preceding yields, upon using the fact that for any random variable *U*, $E\left[U^2\right] = V[U] + \left(E[U]\right)^2$,

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

$$= nE\left[X_1^2\right] - nE\left[\bar{X}^2\right]$$

$$= nV\left[X_1\right] + n\left(E\left[X_1\right]\right)^2 - nV\left[\bar{X}\right] - n\left(E\left[\bar{X}\right]\right)^2$$

$$= n\sigma^2 + n\mu^2 - n\left(\sigma^2 / n\right) - n\mu^2$$

$$= (n-1)\sigma^2$$

$$E\left[S^2\right] = \sigma^2 \tag{7.28}$$

The expected value of the sample variance S^2 is equal to the population variance σ^2.

7.2.4 **SAMPLING DISTRIBUTIONS FROM A NORMAL POPULATION**

Let $X_1, X_2, \dots, X_n$ be a sample from a normal population having mean μ and variance σ^2. That is, they are independent and $X_i \sim N\left(\mu, \sigma^2\right)$, $i = 1, \dots, n$. Also let

$$\bar{X} = \sum_{i=1}^{n} X_i / n \tag{7.29}$$

and

$$S^2 = \frac{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2}{n-1} \tag{7.30}$$

denote the *sample mean* and *sample variance*, respectively. We would like to compute their distributions.

7.2.4a *Distribution of the Sample Mean*

Since the sum of independent normal random variables is normally distributed, it follows that $\bar{X}$ is normal with mean

$$E\left[\bar{X}\right] = \sum_{i=1}^{n} \frac{E\left[X_i\right]}{n} = \mu \tag{7.31}$$

and variance

$$V\left[\bar{X}\right] = \frac{1}{n^2} \sum_{i=1}^{n} V\left[X_i\right] = \sigma^2 / n \tag{7.32}$$

That is, $\bar{X}$, the average of the sample, is normal with a mean equal to the population mean but with a variance reduced by a factor of $1/n$. It follows from this that

$$\frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \tag{7.33}$$

is a standard normal random variable.

7.2.4b *Joint Distribution of* $\bar{X}$ *and* S^2

In this section, we not only obtain the distribution of the sample variance S^2, but we also discover a fundamental fact about normal samples—namely, that $\bar{X}$ and S^2 are independent with $(n-1)S^2/\sigma^2$ having a chi-square distribution with $n-1$ degrees of freedom.

To start, for numbers $x_1, x_2, \ldots, x_n$, let $y_i = x_i - \mu$, $i = 1, 2, \ldots, n$. Then as $\bar{y} = \bar{x} - \mu$, it follows from the identity

$$\sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n} y_i^2 - n\bar{y}^2$$

that

$$\sum_{i=1}^{n}(x_i - \mu - \bar{x} + \mu)^2 = \sum_{i=1}^{n}(x_i - \mu)^2 - n(\bar{x} - \mu)^2$$

$$\sum_{i=1}^{n}(x_i - \bar{x})^2 = \sum_{i=1}^{n}(x_i - \mu)^2 - n(\bar{x} - \mu)^2$$

Now, if $X_1, X_2, \ldots, X_n$ is a sample from a normal population having mean μ and variance σ^2, then we obtain from the preceding identity that

$$\frac{\sum_{i=1}^{n}(X_i - \mu)^2}{\sigma^2} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{\sigma^2} + \frac{n(\bar{X} - \mu)^2}{\sigma^2}$$

or, equivalently,

$$\sum_{i=1}^{n}\left(\frac{X_i - \mu}{\sigma}\right)^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{\sigma^2} + \left(\frac{\sqrt{n}(\bar{X} - \mu)}{\sigma}\right)^2 \qquad (7.34)$$

Because $(X_i - \mu)/\sigma$, $i = 1, 2, \ldots, n$ are independent standard normals, it follows that the left side of Eq (7.34) is a *chi-square* random variable with *n* degrees of freedom.

Also, $\sqrt{n}(\bar{X} - \mu)/\sigma$ is a standard normal random variable and so its square is a *chi-square* random variable with *1* degree of freedom. Thus Eq (7.34) equates a chi-square random variable having *n* degrees of freedom to the sum of two random variables, one of which is chi-square with 1 degree of freedom. But the sum of two independent chi-square random variables is also chi-square with a degree of freedom equal to the sum of the two degrees of freedom. Thus, it would seem that there is a reasonable possibility that the two terms on the right side of Eq (7.34) are independent, with $\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2 / \sigma^2$ having a *chi-square* distribution with $n-1$ degrees of freedom.

Theorem If $X_1, X_2, \ldots, X_n$ is a sample from a normal population having mean μ and variance σ^2, then $\bar{X}$ and S^2 are independent random variables, with $\bar{X}$ being *normal* with mean μ and variance σ^2 / n and $(n-1)S^2/\sigma^2$ being *chi-square* with $n-1$ degrees of freedom.

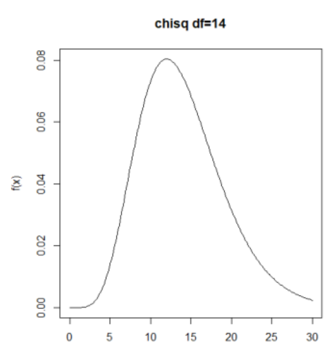
The Theorem not only provides the distributions of $\bar{X}$ and S^2 for a normal population but also establishes the important fact that they are independent. In fact, it turns out that this independence of $\bar{X}$ and S^2 is a unique property of the normal distribution.

Exercise 4: The time it takes a central processing unit to process a certain type of job is normally distributed with mean 20 seconds and standard deviation 3 seconds. If a sample of 15 such jobs is observed, what is the probability that the sample variance will exceed 12?

Since the sample is of size $n = 15$ and $\sigma^2 = 9$, write

$$\Pr\left(S^2 > 12\right) = \Pr\left(\frac{14S^2}{9} > \frac{14}{9} \times 12\right)$$

$$\Pr\left(\chi^2_{14} > 18.67\right) = 1 - \Pr\left(\chi^2_{14} < 18.67\right) = 1 - 0.8221 = 0.1779$$



```
> x<-seq(0,30,by=0.1)
> y<-dchisq(x,14)
> plot(x,y,main = 'chisq df=14')
> pchisq(18.67, 14)
[1] 0.8220542
```

Corollary Let $X_1, X_2, \ldots, X_n$ be a sample from a normal population with mean μ. If $\bar{X}$ denotes the sample mean and *S* the sample standard deviation, then

$$\sqrt{n}\,\frac{\bar{X}-\mu}{S} \approx t_{n-1} \tag{7.35}$$

That is, $\sqrt{n}\left(\bar{X}-\mu\right)/S$ has a *t*-distribution with $n-1$ degrees of freedom.

Proof

Recall that a *t*-random variable with *n* degrees of freedom is defined as the distribution of

$$\frac{Z}{\sqrt{\chi_n^2 / n}} \tag{7.36}$$

where *Z* is a standard normal random variable that is independent of χ_n^2, a chi-square random variable with *n* degrees of freedom. Because the last theorem gives that $\sqrt{n}\left(\bar{X}-\mu\right)/\sigma$ is a standard normal that is independent of $(n-1)S^2/\sigma^2$, which is chi-square with $n-1$ degrees of freedom, we can conclude that

$$\frac{\sqrt{n}\left(\bar{X}-\mu\right)/\sigma}{\sqrt{S^2/\sigma^2}} = \sqrt{n}\,\frac{\left(\bar{X}-\mu\right)}{S} \tag{7.37}$$

is a *t*-random variable with $n-1$ degrees of freedom.

Chapter 8 PARAMETER ESTIMATION

Bibliography: Sheldon Ross (2014)

8.1 INTRODUCTION

Let $X_1, X_2, \ldots, X_n$ be a random sample from a distribution F_θ that is specified up to a vector of unknown parameters θ. For instance, the sample could be from a Poisson distribution whose mean value is unknown; or it could be from a normal distribution having an unknown mean and variance. Whereas in probability theory it is usual to suppose that all of the parameters of a distribution are known, the opposite is true in statistics, where a *central problem* is to use the observed data to make inferences about the unknown parameters.

In this chapter we present the *maximum likelihood method* for determining estimators of unknown parameters. The estimates so obtained are called *point estimates*, because they specify a single quantity as an estimate of θ. Next, we consider the problem of obtaining *interval estimates*. In this case, rather than specifying a certain value as our estimate of θ, we specify an interval in which we estimate that θ lies. Additionally, we consider the question of how much confidence we can attach to such an interval estimate.

In a optional Section, we consider the problem of determining an estimate of an unknown parameter when there is some prior information available. This is the Bayesian approach, which supposes that prior to observing the data, information about θ is always available to the decision maker.

8.2 MAXIMUM LIKELIHOOD ESTIMATORS

Any *statistic* used to estimate the value of an unknown parameter θ is called an *estimator* of θ. The observed value of the estimator is called the *estimate*. For instance, the usual estimator of the mean of a normal population, based on a sample $X_1, X_2, \ldots, X_n$ from that population, is the sample mean $\bar{X} = \sum_{i=1}^{n} X_i / n$. If a sample of size 3 yields the data $X_1 = 6$, $X_2 = 7$, $X_3 = 8$, then the estimate of the population mean, resulting from the estimator $\bar{X}$, is the value 7.

Suppose that the random variables $X_1, X_2, \ldots, X_n$, whose joint distribution is assumed given except for an unknown parameter θ, are to be observed. The problem of interest is to use the observed values to estimate θ. For example, the X_i's might be independent, *exponential* random variables each having the same unknown mean θ. In this case, the joint density function of the random variables would be given by

$$f(x_1, x_2, \ldots, x_n) = f_{X_1}(x_1) f_{X_2}(x_2) \cdots f_{X_n}(x_n) \qquad (8.1)$$

$$= \frac{1}{\theta} e^{-\frac{x_1}{\theta}} \frac{1}{\theta} e^{-\frac{x_2}{\theta}} \cdots \frac{1}{\theta} e^{-\frac{x_n}{\theta}} \ , \ 0 < x_i < \infty \ , \ i = 1, \ldots, n$$

$$= \frac{1}{\theta^n} e^{-\sum_{i=1}^{n} x_i / \theta} \ , \ 0 < x_i < \infty \ , \ i = 1, \ldots, n \qquad (8.2)$$

and the objective would be to estimate θ from the observed data $X_1, X_2, \ldots, X_n$.

A particular type of estimator, known as the *maximum likelihood estimator*, is widely used in statistics. It is obtained by reasoning as follows.

Let $f(x_1, \ldots, x_n | \theta)$ denote the *joint probability function* of the random variables. Because θ is assumed unknown, we also write f as a function of θ. Now since $f(x_1, \ldots, x_n | \theta)$ represents the likelihood that the values $x_1, x_2, \ldots, x_n$ will be observed when θ is the true value of the parameter, it would seem that a reasonable estimate of θ would be that value yielding the largest likelihood of the observed values. In other words, the maximum likelihood estimate $\hat{\theta}$ is defined to be that value of θ maximizing $f(x_1, \ldots, x_n | \theta)$ where $x_1, x_2, \ldots, x_n$ are the observed values. The function $f(x_1, \ldots, x_n | \theta)$ is often referred to as the *likelihood function* of θ.

In determining the maximizing value of θ, it is often useful to use the fact that $f(x_1, \ldots, x_n | \theta)$ and $\ln\left[f(x_1, \ldots, x_n | \theta) \right]$ have their maximum at the same value of θ. Hence, we may also obtain $\hat{\theta}$ by maximizing $\ln\left[f(x_1, \ldots, x_n | \theta) \right]$.

Exercise 1 (Maximum Likelihood Estimator of a Bernoulli Parameter) Suppose that n independent trials, each of which is a success with probability p, are performed. What is the maximum likelihood estimator of p?

The data consist of the values of $X_1, X_2, \ldots, X_n$ where

$$X_i = \begin{cases} 1, & \text{if trial } i \text{ is a succes} \\ 0, & \text{otherwise} \end{cases}$$

$$\Pr(X_i = 1) = p \qquad \Pr(X_i = 0) = 1 - p$$

which can be succinctly expressed as

$$\Pr(X_i = x) = p^x (1-p)^{1-x} \quad , \quad x = 0,1$$

Hence, by the assumed independence of the trials, the likelihood (that is, the joint probability function) of the data is given by

$$f(x_1,\ldots,x_n | p) = \Pr(X_1 = x_1,\ldots,X_n = x_n | p)$$

$$= p^{x_1}(1-p)^{1-x_1} \ldots p^{x_n}(1-p)^{1-x_n}$$

$$= p^{\sum_{i=1}^{n} x_i} (1-p)^{n - \sum_{i=1}^{n} x_i} \quad , \quad x_i = 0,1 \quad , \quad i = 1,\ldots,n$$

To determine the value of p that maximizes the likelihood, first take logs to obtain

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

Differentiation yields

$$\frac{d}{dp} \ln f(x_1,\ldots,x_n | p) = \frac{\sum_{i=1}^{n} x_i}{p} - \frac{\left(n - \sum_{i=1}^{n} x_i\right)}{1-p}$$

Upon equating to zero and solving, we obtain that the maximum likelihood estimate $\hat{p}$ satisfies

$$\frac{\sum_{i=1}^{n} x_i}{\hat{p}} = \frac{n - \sum_{i=1}^{n} x_i}{1-\hat{p}} \quad \Rightarrow \quad (1-\hat{p}) \sum_{i=1}^{n} x_i = \hat{p}\left(n - \sum_{i=1}^{n} x_i\right) \quad \Rightarrow \quad \hat{p} = \frac{\sum_{i=1}^{n} x_i}{n}$$

Hence, the maximum likelihood estimator of the unknown mean of a Bernoulli distribution is given by

$$\hat{p} = \frac{\sum_{i=1}^{n} X_i}{n} \qquad (8.3)$$

For example, suppose that each RAM (random access memory) chip produced by a certain manufacturer is, independently, of acceptable quality with probability *p*. Then if out of a sample of 1,000 tested 921 are acceptable, it follows that the maximum likelihood estimate of *p* is 0.921.

Exercise 2 (Maximum Likelihood Estimator of a Poisson Parameter) Suppose $X_1, X_2, \ldots, X_n$ are independent Poisson random variables each having mean λ. Determine the maximum likelihood estimator of λ.

The likelihood function is given by

$$f\left(x_1, x_2, \ldots, x_n | \lambda\right) = \frac{e^{-\lambda}\lambda^{x_1}}{x_1!}\frac{e^{-\lambda}\lambda^{x_2}}{x_2!}\cdots\frac{e^{-\lambda}\lambda^{x_n}}{x_n!}$$

$$= \frac{e^{-n\lambda}\lambda^{\sum_{i=1}^{n} x_i}}{x_1!x_2!\ldots x_n!}$$

$$\ln f\left(x_1, x_2, \ldots, x_n | \lambda\right) = -n\lambda + \sum_{i=1}^{n} x_i \ln\lambda - \ln\prod_{i=1}^{n} x_i!$$

$$\frac{d}{d\lambda}\ln f\left(x_1, x_2, \ldots, x_n | \lambda\right) = -n + \frac{\sum_{i=1}^{n} x_i}{\lambda}$$

By equating to zero, we obtain that the maximum likelihood estimate

$$\hat{\lambda} = \frac{\sum_{i=1}^{n} x_i}{n}$$

and so the maximum likelihood estimator is given by

$$\hat{\lambda} = \frac{\sum_{i=1}^{n} X_i}{n} \qquad (8.4)$$

For example, suppose that the number of people who enter a certain retail establishment in any day is a Poisson random variable having an unknown mean λ, which must be estimated. If after 20 days a total of 857 people have entered the establishment, then the maximum likelihood estimate of λ is 857/20 = 42.85. That is, we estimate that on average, 42.85 customers will enter the establishment on a given day.

Exercise 3 (Maximum Likelihood Estimator in a Normal Population) Suppose $X_1, X_2, \ldots, X_n$ are independent, normal random variables each with unknown mean μ and unknown standard deviation σ. The joint density is given by

$$f\left(x_1, x_2, \ldots, x_n \middle| \mu, \sigma\right) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}$$

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

$$\ln f\left(x_1, x_2, \ldots, x_n \middle| \mu, \sigma\right) = -\frac{n}{2}\ln(2\pi) - n\ln\sigma - \frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2\sigma^2}$$

In order to find the value of μ and σ maximizing the foregoing, we compute

$$\frac{\partial}{\partial \mu} \ln f\left(x_1, x_2, \ldots, x_n \middle| \mu, \sigma\right) = \frac{\sum_{i=1}^{n}(x_i-\mu)}{\sigma^2}$$

$$\frac{\partial}{\partial \sigma} \ln f\left(x_1, x_2, \ldots, x_n \middle| \mu, \sigma\right) = -\frac{n}{\sigma} + \frac{\sum_{i=1}^{n}(x_i-\mu)^2}{\sigma^3}$$

Equating these equations to zero yields that

$$\hat{\mu} = \frac{\sum_{i=1}^{n} x_i}{n}$$

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

Hence, the maximum likelihood estimators of μ and σ are given, respectively, by

$$\bar{X} \quad \text{and} \quad \left[\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2/n\right]^{1/2} \tag{8.5}$$

It should be noted that the maximum likelihood estimator of the standard deviation σ differs from the sample standard deviation

$$S=\left[\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2/\left(n-1\right)\right]^{1/2} \tag{8.6}$$

8.3 INTERVAL ESTIMATES

Suppose that $X_1, X_2, \ldots, X_n$ is a sample from a normal population having unknown mean μ and known variance σ^2. It has been shown that $\bar{X}=\sum X_i/n$ is the maximum likelihood estimator for μ. However, we don't expect that the sample mean $\bar{X}$ will exactly equal μ, but rather that it will "be close". Hence, rather than a point estimate, it is sometimes more valuable to be able to specify an interval for which we have a certain degree of confidence that μ lies within. To obtain such an interval estimator, we make use of *the probability distribution of the point estimator*.

In the foregoing, since the point estimator $\bar{X}$ is normal with mean μ and variance σ^2/n, it follows that

$$\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}=\sqrt{n}\frac{\left(\bar{X}-\mu\right)}{\sigma} \tag{8.7}$$

has a *standard normal distribution*. Therefore,

$$\Pr\left(-1.96<\sqrt{n}\frac{\left(\bar{X}-\mu\right)}{\sigma}<1.96\right)=0.95 \tag{8.8}$$

```
pnorm(1.96,0,1)-pnorm (-1.96,0,1)
[1] 0.9500042
```

$$\Pr\left(-1.96\frac{\sigma}{\sqrt{n}}<\bar{X}-\mu<1.96\frac{\sigma}{\sqrt{n}}\right)=0.95$$

Multiplying through by −1 yields the equivalent statement

$$\Pr\left(-1.96\frac{\sigma}{\sqrt{n}} < \mu - \bar{X} < 1.96\frac{\sigma}{\sqrt{n}}\right) = 0.95 \qquad \Big| + \bar{X}$$

$$\Pr\left(\bar{X} - 1.96\frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + 1.96\frac{\sigma}{\sqrt{n}}\right) = 0.95 \qquad (8.9)$$

That is, 95 percent of the time the value of the sample average $\bar{X}$ will be such that the distance between it and the mean μ will be less than $1.96\sigma/\sqrt{n}$. If we now observe the sample and it turns out that $\bar{X} = \bar{x}$, then we say that "with 95 percent confidence"

$$\bar{x} - 1.96\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + 1.96\frac{\sigma}{\sqrt{n}} \qquad (8.10)$$

That is, "with 95 percent confidence" we assert that the true mean lies within $1.96\sigma/\sqrt{n}$ of the observed sample mean. The interval

$$\left(\bar{x} - 1.96\frac{\sigma}{\sqrt{n}},\ \bar{x} + 1.96\frac{\sigma}{\sqrt{n}}\right) \qquad (8.11)$$

is called a *95 percent confidence interval* estimate of μ.

Exercise 4 Suppose that when a signal having value μ is transmitted from location *A* the value received at location *B* is normally distributed with mean μ and variance 4. That is, if μ is sent, then the value received is $\mu + N$ where *N*, representing noise, is normal with mean 0 and variance 4. To reduce error, suppose the same value is sent 9 times. If the successive values received are 5, 8.5, 12, 15, 7, 9, 7.5, 6.5, 10.5, let us construct a 95 percent confidence interval for μ.

Since

$$\bar{x} = \frac{81}{9} = 9$$

It follows, under the assumption that the values received are independent, that a 95 percent confidence interval for μ is

$$\left(9 - 1.96\frac{\sigma}{3},\ 9 + 1.96\frac{\sigma}{3}\right) = (7.69,\ 10.31)$$

Hence, we are "95 percent confident" that the true message value lies between 7.69 and 10.31.

The interval in Eq (8.10) is called a *two-sided confidence interval*. Sometimes, however, we are interested in determining a value so that we can assert with, say, 95 percent confidence, that μ is at least as large as that value.

To determine such a value, note that if *Z* is a standard normal random variable then

$$\Pr\left(Z < 1.645\right) = 0.95$$

```
pnorm(1.645,0,1)
[1] 0.9500151
```

$$\Pr\left(\sqrt{n}\frac{\left(\bar{X}-\mu\right)}{\sigma} < 1.645\right) = 0.95 \qquad (8.12)$$

$$\Pr\left(\bar{X}-\mu < 1.645\frac{\sigma}{\sqrt{n}}\right) = 0.95$$

$$\Pr\left(\bar{X}-1.645\frac{\sigma}{\sqrt{n}} < \mu\right) = 0.95 \qquad (8.13)$$

Thus, a 95 percent *one-sided upper confidence interval* for μ is

$$\left(\bar{x}-1.645\frac{\sigma}{\sqrt{n}}, \infty\right) \qquad (8.14)$$

where $\bar{x}$ is the observed value of the sample mean.

A *one-sided lower confidence interval* is obtained similarly; when the observed value of the sample mean is $\bar{x}$, then the 95 percent *one-sided lower confidence interval* for μ is

$$\left(-\infty, \bar{x}+1.645\frac{\sigma}{\sqrt{n}}\right) \qquad (8.15)$$

Exercise 5 Determine the upper and lower 95 percent confidence interval estimates of μ in Exercise 4.

Since $1.645\frac{\sigma}{\sqrt{n}} = 1.645\frac{2}{\sqrt{9}} = 1.097$

the 95 percent *upper* confidence interval is $\left(9-1.097,\ \infty\right) = \left(7.903,\ \infty\right)$

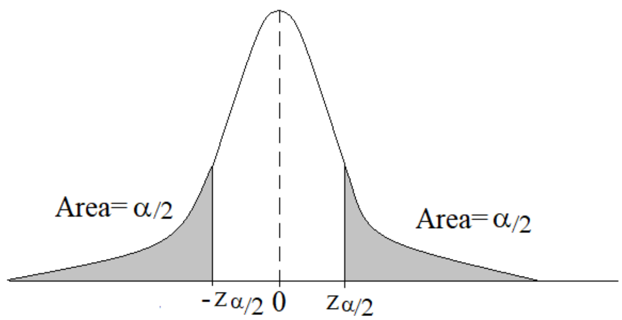
and the 95 percent *lower* confidence interval is $\left(-\infty,\ 9+1.097\right) = \left(-\infty,\ 10.097\right)$

We can also obtain confidence intervals of any specified *level of confidence*. To do so, let z_α be such that

$$\Pr\left(Z > z_\alpha\right) = \alpha \ , \ \alpha \in \left(0,1\right) \tag{8.16}$$

when Z is a standard normal random variable. But this implies (see Figure) that for any α

$$\Pr\left(-z_{\alpha/2} < Z < z_{\alpha/2}\right) = 1-\alpha \tag{8.17}$$



$$\Pr\left(-z_{\alpha/2} < \sqrt{n}\,\frac{\left(\bar{X}-\mu\right)}{\sigma} < z_{\alpha/2}\right) = 1-\alpha \tag{8.18}$$

$$\Pr\left(-z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \bar{X}-\mu < z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = 1-\alpha$$

$$\Pr\left(-z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu-\bar{X} < z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = 1-\alpha$$

$$\Pr\left(\bar{X}-z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{X}+z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = 1-\alpha \tag{8.19}$$

Hence, a $100(1-\alpha)$ *percent two-sided confidence interval* for μ is

$$\mu \in \left(\bar{x}-z_{\alpha/2}\frac{\sigma}{\sqrt{n}},\ \bar{x}+z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) \tag{8.20}$$

where $\bar{x}$ is the observed sample mean.

Similarly, knowing that $Z = \sqrt{n}\frac{(\bar{X} - \mu)}{\sigma}$ is a standard normal random variable, along with the identities

$$\Pr(Z > z_\alpha) = \alpha \qquad (8.21)$$

and

$$\Pr(Z < -z_\alpha) = \alpha \qquad (8.22)$$

results in one-sided confidence intervals of any desired level of confidence. Specifically, we obtain that

$$\left(\bar{x} - z_\alpha \frac{\sigma}{\sqrt{n}},\ \infty\right) \qquad (8.23)$$

and $$\left(-\infty,\ \bar{x} + z_\alpha \frac{\sigma}{\sqrt{n}}\right) \qquad (8.24)$$

are, respectively, $100(1-\alpha)$ percent *one-sided upper* and $100(1-\alpha)$ percent *one-sided lower* confidence intervals for μ.

Exercise 6 Use the data of Exercise 4 ($\sigma^2 = 4$, $n = 9$, $\bar{x} = 9$) to obtain a 99 percent confidence interval estimate of μ, along with 99 percent one-sided upper and lower intervals.

Since $\alpha = 0.01$ and $z_{\alpha/2} = z_{0.005} = 2.58$,

```
qnorm(0.005,0,1,lower.tail = FALSE)
[1] 2.575829
```

$$2.58\frac{\sigma}{\sqrt{n}} = \frac{5.16}{3} = 1.72$$

it follows that a 99 percent confidence interval for μ is

$$9 \pm 1.72$$

That is, the 99 percent confidence interval estimate is $(7.28,\ 10.72)$.

Also, since $z_{0.01} = 2.33$,

```
qnorm(0.01,0,1,lower.tail = FALSE)
[1] 2.326348
```

a 99 percent upper confidence interval is

$$\left(9-2.33\times\frac{2}{3},\ \infty\right)=(7.447,\ \infty)$$

Similarly, a 99 percent lower confidence interval is

$$\left(-\infty,\ 9+2.33\times\frac{2}{3}\right)=(-\infty,\ 10.553)$$

Sometimes we are interested in a two-sided confidence interval of a certain level, say $1-\alpha$, and the problem is to choose the sample size *n* so that the interval is of a certain size. For instance, suppose that we want to compute an interval of length 0.1 that we can assert, with 99 percent confidence, contains μ. How large need *n* be? To solve this, note that as $z_{0.005}=2.58$ it follows that the 99 percent confidence interval for μ from a sample of size *n* is

$$\left(\overline{x}-2.58\frac{\sigma}{\sqrt{n}},\overline{x}+2.58\frac{\sigma}{\sqrt{n}}\right)$$

Hence, its length is

$$5.16\frac{\sigma}{\sqrt{n}}$$

Thus, to make the length of the interval equal to 0.1, we must choose

$$5.16\frac{\sigma}{\sqrt{n}}=0.1$$

$$n=(51.6\sigma)^2$$

REMARK

The interpretation of “a $100(1-\alpha)$ percent confidence interval” can be confusing.

It should be noted that we are not asserting that the probability that

$$\mu\in\left(\overline{x}-1.96\sigma/\sqrt{n}\ ,\ \ \overline{x}+1.96\sigma/\sqrt{n}\right)$$

is 0.95, for there are no random variables involved in this assertion. What we are asserting is that *the technique utilized to obtain this interval* is such that 95 percent of the time that it is employed it will result in an interval in which μ lies. In other words, before the data are observed we can assert that with probability 0.95 the interval that will be obtained will contain μ, whereas after the data are obtained we can only assert that the resultant interval indeed contains μ "with confidence 0.95."

Exercise 7 From past experience it is known that the weights of salmon grown at a commercial hatchery are normal with a mean that varies from season to season but with a standard deviation that remains fixed at 0.3 pounds. If we want to be 95 percent certain that our estimate of the present season's mean weight of a salmon is correct to within ±0.1 pounds, how large a sample is needed?

A 95 percent confidence interval estimate for the unknown mean μ, based on a sample of size *n*, is

$$\mu \in \left(\bar{x} - 1.96\sigma/\sqrt{n}\ ,\ \ \bar{x} + 1.96\sigma/\sqrt{n}\right)$$

Because the estimate $\bar{x}$ is within $1.96\sigma/\sqrt{n} = 0.588/\sqrt{n}$ of any point in the interval, it follows that we can be 95 percent certain that $\bar{x}$ is within 0.1 of μ provided that

$$\frac{0.588}{\sqrt{n}} \le 0.1$$

$$\sqrt{n} \ge 5.88 \ \ \text{or} \ \ n \ge 34.57$$

That is, a sample size of 35 or larger will suffice.