3 Methods of Point Estimation

So far, we have discussed desirable properties of point estimators, how to distinguish "good" from "bad" or "better" estimators, based on how reliable they are in approximating the value of a population parameter. In all the procedures we analyzed and all the examples we discussed, the value of a point estimator $\overline{\theta}$ was given for a target parameter θ, based on sample variables $X_1, X_2, \ldots, X_n$, i.e. $\overline{\theta} = \overline{\theta}(X_1, X_2, \ldots, X_n)$. But *how* to *actually* find an estimator, an approximating value? Sometimes, such a value may be "guessed" from past experience or from observing many samples over time. But statisticians wanted more rigorous, more mathematical ways of producing a point estimator, which can then be analyzed from the various points of view discussed in the previous section. They developed a number of estimation techniques, each having certain optimal properties. This question will be addressed in this section.

We present two of the most popular methods of finding point estimators: the *method of moments* and the *method of maximum likelihood*. We will also discuss advantages and disadvantages of each method.

3.1 Method of Moments

This is one of the oldest and easiest methods for obtaining point estimators, first formalized by K. Pearson in the late 1800's.

Let us recall, for a population characteristic X, we define the *moments of order* k as

$$\nu_k \;=\; E\left(X^k\right) \;=\; \begin{cases} \displaystyle\sum_{i\in I} x_i^k p_i, & \text{if } X \text{ is discrete with pdf } X \begin{pmatrix} x_i \\ p_i \end{pmatrix}_{i\in I} \\ \\ \displaystyle\int_{\mathbb{R}} x^k f(x)\, dx, & \text{if } X \text{ is continuous with pdf } f:\mathbb{R}\to\mathbb{R}. \end{cases} \tag{3.1}$$

For a sample drawn from the distribution of X, i.e. sample variables $X_1, \ldots, X_n$ (iid), the *sample moments of order* k are defined by

$$\overline{\nu}_k \;=\; \frac{1}{n}\sum_{i=1}^{n} X_i^k. \tag{3.2}$$

Also, let us recall (from Proposition 2.6, Chapter 3, Lecture 4) that

$$E(\overline{\nu}_k) \;=\; \nu_k, \tag{3.3}$$

so $\overline{\nu}_k$ is an *unbiased* estimator for ν_k and

$$V(\overline{\nu}_k) \;=\; \frac{1}{n}\left(\nu_{2k} - \nu_k^2\right) \;\to\; 0, \text{ as } n \to \infty. \tag{3.4}$$

By (3.3)-(3.4), the sample moment of order k is an *absolutely correct* estimator for the population moment of the same order.

That is precisely the idea of this method. Since the theoretical (population) moments in (3.1) contain the target parameters that are to be estimated, while the sample moments in (3.2) are all known, computable from the sample data, simply set the two to be equal and solve the resulting system. To estimate k parameters, equate the first k population and sample moments:

$$\left\{\begin{array}{lcl} \nu_1 &=& \overline{\nu}_1 \\ \ldots & \ldots & \ldots \\ \nu_k &=& \overline{\nu}_k \end{array}\right. \tag{3.5}$$

The left-hand sides of these equations depend on the distribution parameters. The right-hand sides can be computed from data. The **method of moments estimator** is the solution of this $k \times k$ system of equations.

Remark 3.1. We state, without proof, the fact that an estimator $\overline{\theta}_n$ obtained by the method of moments converges *almost surely* to the target parameter that it estimates:

$$\overline{\theta}_n \overset{\text{a.s.}}{\to} \theta.$$

That implies the convergence *in probability*, $\overline{\theta}_n \overset{p}{\to} \theta$, so any method of moments estimator is a *consistent* estimator.

Example 3.2. Let X be a characteristic with pdf

$$f(x;\theta) = \frac{1}{\theta^2} x e^{-\frac{x}{\theta}},$$

for $x > 0$ and 0, otherwise, where $\theta > 0$ is unknown (this was used in Example 2.11, Lecture 6). Based on a random sample $X_1, \ldots, X_n$, find the method of moments estimator $\overline{\theta}$ for θ. For the sample data $\{2.3, 3.7, 1.44, 2.16\}$, find the numerical estimation of θ.

Solution. There is only one unknown parameter, θ, so we will have only one equation in system (3.5),

$$\begin{aligned} \nu_1 &= \overline{\nu}_1, \text{ i.e.} \\ E(X) &= \overline{X}. \end{aligned}$$

In our work in Example 2.11 (Lecture 6), we computed

$$E(X) = \int_{\mathbb{R}} xf(x)\,dx = 2\theta.$$

So, we solve the equation

$$2\theta = \overline{X},$$

to find the method of moments estimator

$$\overline{\theta} = \frac{1}{2}\overline{X}.$$

Notice that it is an unbiased estimator, since

$$E\left(\overline{\theta}\right) = E\left(\frac{1}{2}\overline{X}\right) = \frac{1}{2}E\left(\overline{X}\right) = \frac{1}{2}E(X) = \frac{1}{2}\cdot 2\theta = \theta.$$

For the sample data $x_1 = 2.3, x_2 = 3.7, x_3 = 1.44$ and $x_4 = 2.16$, we have

$$\overline{x} = \frac{x_1 + x_2 + x_3 + x_4}{4} = \frac{9.6}{4} = 2.4,$$

so the numerical value of our estimator is

$$\overline{\theta} = 1.2.$$

■

Example 3.3. Let us use the method of moments to estimate *both* parameters of the Normal $N(\mu, \sigma)$ distribution.

Solution. Now we have a characteristic X with pdf

$$f(x;\mu,\sigma) \;=\; \frac{1}{\sigma\sqrt{2\pi}} e^{-\dfrac{(x-\mu)^2}{2\sigma^2}},\; x \in \mathbb{R},$$

with $\mu \in \mathbb{R}$ and $\sigma > 0$, both unknown.

To estimate two parameters, we need two equations in system (3.5),

$$\begin{cases} \nu_1 &=& \overline{\nu}_1 \\ \nu_2 &=& \overline{\nu}_2 \end{cases}$$

In the first equation, we have

$$\begin{aligned} \nu_1 &= E(X) &= \mu \text{ and} \\ \overline{\nu}_1 &= \overline{X}, \end{aligned}$$

since for a Normal $N(\mu,\sigma)$ variable the first parameter is its expectation. We also know that the variance of a $N(\mu,\sigma)$ variable is equal to σ^2. But recall the computational formula for the variance (in general)

$$V(X) \;=\; E\left(X^2\right) - \left(E(X)\right)^2 \;=\; \nu_2 - \nu_1^2.$$

From here, we get

$$\nu_2 \;=\; V(X) + \nu_1^2 \;=\; \sigma^2 + \mu^2,$$

in this case.

So system (3.5) becomes

$$\begin{cases} \mu &=& \overline{X} \\ \mu^2 + \sigma^2 &=& \overline{\nu}_2 \end{cases},$$

a system of two equations in two unknowns, with solution

$$\begin{cases} \overline{\mu} &=& \overline{X} \\ \overline{\sigma} &=& \sqrt{\overline{\nu}_2 - \overline{X}^2} \;=\; \sqrt{\dfrac{1}{n}\sum_{i=1}^{n} X_i^2 - \left(\dfrac{1}{n}\sum_{i=1}^{n} X_i\right)^2} \end{cases}.$$

■

Remark 3.4. Method of moments estimates are typically easy to compute. However, on rare occasions, when k equations are not enough to estimate k parameters, higher moments (i.e. more

equations) can be considered.

3.2 Method of Maximum Likelihood

Maximum-likelihood estimation was first recommended, analyzed and then vastly popularized by R. A. Fisher in the 1920's, although it had been used earlier by Gauss and Laplace. For a fixed random sample from an underlying probability distribution, the maximum likelihood method picks the values of the population parameters that make the data "more likely" than any other values of the parameters would make them.

Let us illustrate it, first, with a simple example, to understand the underlying ideas.

Example 3.5. Suppose there are 5 balls in a box, black or white, the number of each being unknown. Suppose further, that we randomly select 3 of them, without replacement, and we get all three white. What would be a good estimate, $\overline{w}$, for the number of white balls in the box, w?

Solution. Obviously, $w \in \{3, 4, 5\}$.

If the true value was $w = 3$, then the probability of randomly selecting 3 white balls without replacement, would be (by the Hypergeometric model)

$$p_1 = \frac{C_3^3 C_2^0}{C_5^3} = \frac{1}{10}.$$

If the true value was $w = 4$, then the probability of us randomly selecting 3 white balls without replacement, would be

$$p_2 = \frac{C_4^3 C_1^0}{C_5^3} = \frac{4}{10}.$$

And, finally, if the true value was $w = 5$, then the probability of randomly selecting 3 white balls without replacement, would be

$$p_3 = \frac{C_5^3 C_0^0}{C_5^3} = 1.$$

So, it would seem reasonable to choose $\overline{w} = 5$ as our estimate for w, since this would *maximize* the probability of obtaining our observed sample.

■

This, in essence, describes the *method of maximum likelihood* estimation. Now let us write it formally.

Recall that the probability of obtaining an observed sample is measured by the *likelihood function*

of a sample:

$$L(X_1, \ldots, X_n; \Theta) = \prod_{i=1}^{n} f(X_i; \Theta),$$

where now *all* unknown target parameters are contained in a vector $\Theta = (\theta_1, \ldots, \theta_l)$.
This method chooses the values of an estimator $\overline{\Theta} = (\overline{\theta}_1, \ldots, \overline{\theta}_l) = \overline{\Theta}(X_1, \ldots, X_n)$ that maximize the function $L(X_1, \ldots, X_n; \Theta)$. So, if L is differentiable with respect to each $\theta_1, \ldots, \theta_l$, we find the solutions of the maximum-likelihood system

$$\frac{\partial L(X_1, \ldots, X_n; \theta_1, \ldots, \theta_l)}{\partial \theta_j} = 0, \quad j = \overline{1, l}, \tag{3.6}$$

or, equivalently, but easier to compute, the maximum-likelihood equations

$$\frac{\partial \ln L(X_1, \ldots, X_n; \theta_1, \ldots, \theta_l)}{\partial \theta_j} = 0, \quad j = \overline{1, l}. \tag{3.7}$$

If the system (3.7) has a solution, then it is unique and it is called the **maximum likelihood (MLE) estimator**.

Example 3.6. Consider again the situation in Example 3.2, so a characteristic with pdf

$$f(x; \theta) = \frac{1}{\theta^2} x e^{-\frac{x}{\theta}},$$

for $x > 0$, with $\theta > 0$ is unknown. Based on a random sample $X_1, \ldots, X_n$, let us now find the MLE $\hat{\theta}$ for θ.

Solution. The likelihood function is given by

$$\begin{aligned} L(X_1, \ldots, X_n; \theta) &= \prod_{i=1}^{n} \frac{1}{\theta^2} X_i e^{-\frac{X_i}{\theta}} \\ &= \left(\prod_{i=1}^{n} X_i \right) \frac{1}{\theta^{2n}} e^{-\frac{1}{\theta} \sum_{i=1}^{n} X_i} \\ &= K \frac{1}{\theta^{2n}} e^{-\frac{n\overline{X}}{\theta}}, \end{aligned}$$

where $K = \prod_{i=1}^{n} x_i$ is a constant with respect to θ.

Take the logarithm, to make computations easier and differentiate it with respect to θ (the only unknown).

$$\begin{aligned} \ln L &= \ln K - 2n \ln \theta - \frac{n\overline{X}}{\theta} \\ \frac{\partial \ln L}{\partial \theta} &= -\frac{2n}{\theta} + \frac{n\overline{X}}{\theta^2}. \end{aligned}$$

Then system (3.7) becomes

$$-\frac{2n}{\theta} + \frac{n\overline{X}}{\theta^2} = 0,$$

whose solution is the MLE

$$\hat{\theta} = \frac{1}{2}\overline{X},$$

the same as the method of moments estimator $\hat{\theta}$.

■

Example 3.7. Let us also find the MLE's for the parameters of the Normal $N(\mu, \sigma)$ distribution

$$f(x; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \ x \in \mathbb{R},$$

with $\mu \in \mathbb{R}$ and $\sigma > 0$, both unknown.

Solution. We find the likelihood function and its logarithm:

$$\begin{aligned} L(X_1, \ldots, X_n; \mu, \sigma) &= \prod_{i=1}^{n} \left(\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(X_i-\mu)^2}{2\sigma^2}} \right) \\ &= \left(\frac{1}{\sigma\sqrt{2\pi}} \right)^n e^{-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (X_i - \mu)^2}, \\ \ln L(\mu, \sigma) &= -n \ln (\sigma\sqrt{2\pi}) - \frac{1}{2\sigma^2} \left(\sum_{i=1}^{n} X_i^2 - 2\mu \sum_{i=1}^{n} X_i + n\mu^2 \right) \\ &= -n \ln \sigma - n \ln (\sqrt{2\pi}) - \frac{1}{2\sigma^2} \left(\sum_{i=1}^{n} X_i^2 - 2n\mu\overline{X} + n\mu^2 \right). \end{aligned}$$

The maximum likelihood system will consist of two equations

$$\begin{cases} \dfrac{\partial \ln L(\mu,\sigma)}{\partial \mu} &= 0 \\ \dfrac{\partial \ln L(\mu,\sigma)}{\partial \sigma} &= 0, \end{cases}$$

i.e.

$$\begin{cases} -\dfrac{1}{2\sigma^2}\left(-2n\overline{X}+2n\mu\right) &= 0 \\ -n\dfrac{1}{\sigma}+\dfrac{1}{\sigma^3}\left(\displaystyle\sum_{i=1}^{n} X_i^2 - 2n\mu\overline{X}+n\mu^2\right) &= 0, \end{cases}$$

From the first equation, we get

$$\hat{\mu} = \overline{X}.$$

Substituting that into the second equation, we find

$$\begin{aligned} n &= \frac{1}{\sigma^2}\left(\sum_{i=1}^{n} X_i^2 - 2n\hat{\mu}\overline{X}+n\hat{\mu}^2\right), \\ \sigma^2 &= \frac{1}{n}\left(\sum_{i=1}^{n} X_i^2 - 2n\overline{X}^2+n\overline{X}^2\right) \\ &= \overline{\nu}_2 - \overline{X}^2 \\ \hat{\sigma} &= \sqrt{\overline{\nu}_2 - \overline{X}^2} \end{aligned}$$

So, again, the MLE's coincide with the method of moments estimators.

■

Remark 3.8. In both our examples, the two methods yielded the same point estimator. That is not always the case. If they differ, the natural question is: which one is better? In some respects, when estimating parameters of a known family of probability distributions, the method of moments is superseded by Fisher's method of maximum likelihood, because maximum likelihood estimators have higher probability of being close to the quantities to be estimated. However, in some cases, the likelihood equations (3.7) may be intractable without computers, whereas the method of moments estimators can be quickly and easily calculated by hand as seen above. Estimates by the method of moments may be used as the first approximation to the solutions of the likelihood equations (3.7), and successive improved approximations may then be found by the Newton method. In this way,

the method of moments and the method of maximum likelihood are symbiotic. In some cases, infrequent with large samples but not so infrequent with small samples, the estimates given by the method of moments are outside of the parameter space and it does not make sense to rely on them then. That problem never arises in the method of maximum likelihood. Also, estimates by the method of moments are not necessarily sufficient statistics, i.e., they sometimes fail to take into account all relevant information in the sample.

4 Estimation by Confidence Intervals

4.1 Basic Concepts

So far, point estimators provided one single value, $\overline{\theta}$, to estimate the value of an unknown parameter θ, but little measure of the accuracy of the estimate. In contrast, an **interval estimator** specifies a *range* of values, within which the parameter is estimated to lie. More specifically, the sample will be used to produce *two* sample functions, $\overline{\theta}_L(X_1, \ldots, X_n) < \overline{\theta}_U(X_1, \ldots, X_n)$, with values $\overline{\theta}_L = \overline{\theta}_L(x_1, \ldots, x_n)$, $\overline{\theta}_U = \overline{\theta}_U(x_1, \ldots, x_n)$, respectively, such that for a given $\alpha \in (0, 1)$,

$$P(\overline{\theta}_L \leq \theta \leq \overline{\theta}_U) \quad = \quad 1 - \alpha. \tag{4.1}$$

Then
– the range $(\overline{\theta}_L, \overline{\theta}_U)$ is called a **confidence interval (CI)**, more specifically, a $100(1-\alpha)\%$ confidence interval,
– the values $\overline{\theta}_L, \overline{\theta}_U$ are called (lower and upper) **confidence limits**,
– the quantity $1 - \alpha$ is called **confidence level** or **confidence coefficient** and
– the value α is called **significance level**.

Remark 4.1.
1. It may seem a little peculiar that we use $1 - \alpha$ instead of simply α in (4.1), since both values are in $(0, 1)$, but the reasons are in close connection with *hypothesis testing* and will be revealed in the next sections.
2. The condition (4.1) *does not* uniquely determine a $100(1-\alpha)\%$ CI.
3. Evidently, the smaller α and the length of the interval $\overline{\theta}_U - \overline{\theta}_L$ are, the better the estimate for θ. Unfortunately, as we will see, as the confidence level increases, so does the length of the CI, thus, reducing accuracy.

To produce a CI estimate for θ, we need a *pivotal quantity*, i.e. a statistic S that satisfies two

conditions:
$-$ $S = S(X_1, \ldots, X_n; \theta)$ is a function of the sample measurements and the unknown parameter θ, this being the *only* unknown,
$-$ the distribution of S is known and does not depend on θ.

Let us illustrate the idea of the pivotal method, by a simple example.

Example 4.2. Suppose we have a single observation X_1, from an Exponential $Exp(1/\theta)$ distribution, with $\theta > 0$ unknown. We want to use X_1 to find a 90% CI for θ.

Solution. The pdf of X_1 is given by

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

Let $S = \theta X_1$. Then it can be easily shown that S also has an Exponential distribution, $Exp(1)$, with pdf

$$f_S(x) \;=\; e^{-x},\ x > 0.$$

Then its cdf is (by integration)

$$F_S(x) \;=\; 1 - e^{-x},\ x > 0,$$

and we have

$$P(a \le S \le b) \;=\; F_S(b) - F_S(a) \;=\; e^{-a} - e^{-b}.$$

So choose $a, b > 0$ such that $e^{-a} - e^{-b} = 0.9$. Obviously, there are infinitely many such pairs. Let's say we choose $e^{-a} = 0.95$ and $e^{-b} = 0.05$, i.e. $a = 0.051, b = 2.996$. Then we can use this to find a CI for θ.

$$\begin{aligned} 0.9 &= P(0.051 \le S \le 2.996) \\ &= P(0.051 \le \theta X_1 \le 2.996) \\ &= P\left(\frac{0.051}{X_1} \le \theta \le \frac{2.996}{X_1}\right) \end{aligned}$$

So, we found a 90% CI for θ, $\left[\dfrac{0.051}{X_1}, \dfrac{2.996}{X_1}\right]$.

■

We used the variable S, which was related to θ, but for which the distribution was *completely* known and thus we could find confidence interval limits. Once those were found for S, we went backwards and found the limits for θ. This is the idea of the pivotal method.

4.2 Confidence Intervals, General Framework

We will use the pivotal method to find $100(1-\alpha)\%$ CI's, as in the previous example. Depending on which population parameter we wish to estimate, the expression and the pdf of the pivot will change, but the principles will stay the same. So, we start with the case where the pivot has a $N(0,1)$ distribution, so we can better understand the ideas.

Let θ be a target parameter and let $\overline{\theta}$ be an unbiased estimator for θ ($E(\overline{\theta}) = \theta$), with standard error $\sigma_{\overline{\theta}}$, such that, under certain conditions, it is know that

$$Z \;=\; \frac{\overline{\theta}-\theta}{\sigma_{\overline{\theta}}} \;\left(= \;\frac{\overline{\theta}-E(\overline{\theta})}{\sigma(\overline{\theta})}\right) \tag{4.2}$$

has an approximately Standard Normal $N(0,1)$ distribution. We can use Z as a pivotal quantity to construct a $100(1-\alpha)\%$ CI for estimating θ. Just as in the previous example, since the pdf of Z is known, we can choose two values, Z_L, Z_U such that for a given $\alpha \in (0,1)$,

$$P(Z_L \leq Z \leq Z_U) \;=\; 1-\alpha. \tag{4.3}$$

How to choose them? Of course, there are infinitely many possibilities. Recall that for continuous random variables, the probability in (4.3) is an *area*, namely the area under the graph of the pdf and above the x-axis, between the values Z_L and Z_U. Basically, the values Z_L and Z_U should be chosen so that that area is $1-\alpha$. We will take advantage of the symmetry of the Standard Normal pdf and choose the two values so that the area $1-\alpha$ is in "the middle". That means (since the total area under the graph is 1) the two portions left on the two sides, both should have an area of $\dfrac{\alpha}{2}$, as seen in Figure 1.

So what should the values be? Recall *quantiles*. A quantile of a given order $\beta \in (0,1)$ for a random variable X, is a value q_β with the property that

$$F(q_\beta) \;=\; P(X \leq q_\beta) \;=\; \beta,$$

i.e., that the area under the graph of the pdf, to the *left* of q_β is β.
Since for Z_L we want the area to its left to be $\alpha/2$, we choose it to be the quantile of order $\alpha/2$ for

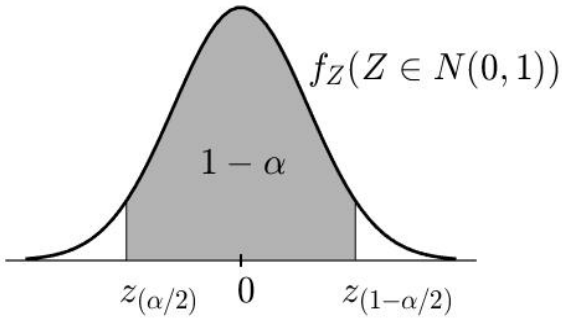

Fig. 1: Confidence Interval for the $N(0,1)$ distribution

Z,

$$Z_L \;=\; z_{\alpha/2}.$$

For the value Z_U, the area to its *right* should be $\alpha/2$, which means the area to the left is $1-\alpha/2$. Thus, we choose

$$Z_U \;=\; z_{1-\alpha/2}.$$

Indeed, now we have

$$P(z_{\alpha/2} \le Z \le z_{1-\alpha/2}) \;=\; 1-\alpha,$$

as in (4.2).

From here, we proceed (as in the previous example) to rewrite the inequality inside, until we get the limits of the CI for θ. We have

$$\begin{aligned} 1-\alpha &= P\left(z_{\frac{\alpha}{2}} \le \frac{\overline{\theta}-\theta}{\sigma_{\overline{\theta}}} \le z_{1-\frac{\alpha}{2}}\right) \\ &= P\left(\sigma_{\overline{\theta}} \cdot z_{\frac{\alpha}{2}} \le \overline{\theta}-\theta \le \sigma_{\overline{\theta}} \cdot z_{1-\frac{\alpha}{2}}\right) \\ &= P\left(-\sigma_{\overline{\theta}} \cdot z_{1-\frac{\alpha}{2}} \le \theta-\overline{\theta} \le -\sigma_{\overline{\theta}} \cdot z_{\frac{\alpha}{2}}\right) \\ &= P\left(\overline{\theta}-\sigma_{\overline{\theta}} \cdot z_{1-\frac{\alpha}{2}} \le \theta \le \overline{\theta}-\sigma_{\overline{\theta}} \cdot z_{\frac{\alpha}{2}}\right), \end{aligned}$$

so the $100(1-\alpha)\%$ CI for θ is given by

$$\left[\overline{\theta}-\sigma_{\overline{\theta}} \cdot z_{1-\frac{\alpha}{2}},\; \overline{\theta}-\sigma_{\overline{\theta}} \cdot z_{\frac{\alpha}{2}}\right]. \tag{4.4}$$

Remark 4.3.

1. Since the Standard Normal distribution is symmetric about the origin, we have $z_{\frac{\alpha}{2}} = -z_{1-\frac{\alpha}{2}}$ and the CI can be written as

$$\left[\overline{\theta} - \sigma_{\overline{\theta}} \cdot z_{1-\frac{\alpha}{2}},\, \overline{\theta} + \sigma_{\overline{\theta}} \cdot z_{1-\frac{\alpha}{2}}\right] \quad \text{or} \quad \left[\overline{\theta} + \sigma_{\overline{\theta}} \cdot z_{\frac{\alpha}{2}},\, \overline{\theta} - \sigma_{\overline{\theta}} \cdot z_{\frac{\alpha}{2}}\right].$$

2. The CI we determined is a **two-sided CI**, because it gives bounds on both sides. A two-sided CI is not always the most appropriate for the estimation of a parameter θ. It may be more relevant to make a statement simply about how *large* or how *small* the parameter might be, i.e. to find confidence intervals of the form $(-\infty, \overline{\theta}_U]$ and $[\overline{\theta}_L, \infty)$, respectively, such that the probability that θ is in the CI is $1-\alpha$. These are called **one-sided confidence intervals** and they can be found the same way, using quantiles of an appropriate order. Then, we can find a $100(1-\alpha)\%$ *upper confidence interval* for θ, as

$$\left(-\infty, \overline{\theta} - \sigma_{\overline{\theta}} \cdot z_{\alpha}\right]$$

and a $100(1-\alpha)\%$ *lower confidence interval* for θ,

$$\left[\overline{\theta} - \sigma_{\overline{\theta}} \cdot z_{1-\alpha}, \infty\right).$$

3. As mentioned earlier, for estimating various population parameters, the pivot will be different, but the procedure of finding the CI will be the same, even when the distribution of the pivot *is not* symmetric.