Chapter 6. Statistical Inference

1 Sample Theory

In inferential Statistics, we will have the following situation: we are interested in studying a characteristic (a random variable) X, relative to a population P of (known or unknown) size N. The difficulty or even the impossibility of studying the entire population, as well as the merits of choosing and studying a random sample from which to make inferences about the population of interest, have already been discussed in the previous chapter. Now, we want to give a more rigorous and precise definition of a random sample, in the framework of random variables, one that can then employ probability theory techniques for making inferences.

We choose n objects from the population and actually study $X_i,\ i = \overline{1,n}$, the characteristic of interest *for the i^{th} object selected*. Since the n objects were randomly selected, it makes sense that for $i = \overline{1,n}$, X_i is a random variable, one that has *the same* distribution (pdf) as X, the characteristic relative to the entire population. Furthermore, these random variables are independent, since the value assumed by one of them has no effect on the values assumed by the others. Once the n objects have been selected, we will have n numerical values available, $x_1, \ldots, x_n$, the observed values of $X_1, \ldots, X_n$.

Definition 1.1. *A **random sample of size n** from the distribution of X, a characteristic relative to a population P, is a collection of n independent random variables $X_1, \ldots, X_n$, having the same distribution as X. The variables $X_1, \ldots, X_n$, are called **sample variables** and their observed values $x_1, \ldots, x_n$, are called **sample data**.*

We are able now to define sample functions, or statistics, in the more precise context of random variables.

Definition 1.2. *A **sample function** or **statistic** is a random variable*

$$Y_n = h_n(X_1, \ldots, X_n),$$

where $h_n : \mathbb{R}^n \to \mathbb{R}$ is a measurable function. The value of the sample function Y_n is $y_n = h_n(x_1, \ldots, x_n)$.

We will revisit now some sample numerical characteristics discussed in the previous chapter and define them as sample functions. That means they will have a pdf, a cdf, a mean value, variance, standard deviation, etc. A sample function will, in general, be an approximation for the corresponding population characteristic. In that context, the standard deviation of the sample function is usually referred to as the **standard error**.

In what follows, $\{X_1, \dots, X_n\}$ denotes a sample of size n drawn from the distribution of some population characteristic X.

Definition 1.3. *The **sample mean** is the sample function defined by*

$$\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i \tag{1.1}$$

and its value is $\overline{x} = \dfrac{1}{n} \displaystyle\sum_{i=1}^{n} x_i$.

Now that the sample mean is defined as a random variable, we can discuss its distribution and its numerical characteristics.

Proposition 1.4. *Let X be a characteristic with $E(X) = \mu$ and $V(X) = \sigma^2$. Then*

$$E\left(\overline{X}\right) = \mu \ \text{ and } \ V\left(\overline{X}\right) = \frac{\sigma^2}{n}. \tag{1.2}$$

Moreover, if $X \in N(\mu, \sigma)$, then $\overline{X} \in N\left(\mu, \dfrac{\sigma}{\sqrt{n}}\right)$.

Proof. Since $X_1, \dots, X_n$ are identically distributed, with the same distribution as X, $E(X_i) = E(X) = \mu$ and $V(X_i) = V(X) = \sigma^2$, $\forall i = \overline{1, n}$. Then, by the usual properties of expectation, we have

$$E\left(\overline{X}\right) = E\left(\frac{1}{n} \sum_{i=1}^{n} X_i\right) = \frac{1}{n} \sum_{i=1}^{n} E(X_i) = \frac{1}{n}\, n\mu = \mu.$$

Further, since $X_1, \dots, X_n$ are also independent, by the properties of variance, it follows that

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

The last part follows from the fact that $\overline{X}$ is a linear combination of independent, Normally distributed random variables.

$\square$

Remark 1.5. As a consequence, the standard deviation of $\overline{X}$ is

$$\mathrm{Std}(\overline{X}) \;=\; \sqrt{V(\overline{X})} \;=\; \frac{\sigma}{\sqrt{n}}.$$

So, when estimating the population mean μ from a sample of size n by the sample mean $\overline{X}$, the *standard error* of the estimate is $\sigma/\sqrt{n}$, which oftentimes is estimated by $s/\sqrt{n}$. Either way, notice that as n increases and tends to ∞, the standard error decreases and approaches 0. That means that the larger the sample on which we base our estimate, the more accurate the approximation.

Corollary 1.6. *Let X be a characteristic with $E(X)=\mu$ and $V(X)=\sigma^2$ and for $n\in\mathbb{N}$, let*

$$Z_n = \frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}}.$$

Then the variable Z_n converges in distribution to a Standard Normal variable, as $n\to\infty$, i.e. $F_{Z_n} \overset{n\to\infty}{\longrightarrow} F_Z = \Phi$. Moreover, if $X\in N(\mu,\sigma)$, then the statement is true for every $n\in\mathbb{N}$.

Proof. This is a direct consequence of the Central Limit Theorem (CLT).

□

Definition 1.7. *The statistic*

$$\overline{\nu}_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k \tag{1.3}$$

is called the ***sample moment of order k*** *and its value is* $\frac{1}{n}\sum_{i=1}^{n} x_i^k$.

The statistic

$$\overline{\mu}_k = \frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^k \tag{1.4}$$

is called the ***sample central moment of order k*** *and its value is* $\frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})^k$.

Remark 1.8. Just like for theoretical (population) moments, we have

$$\begin{aligned}\overline{\nu}_1 &= \overline{X},\\ \overline{\mu}_1 &= 0,\\ \overline{\mu}_2 &= \overline{\nu}_2-\overline{\nu}_1^2.\end{aligned}$$

Next we discuss the distributions and characteristics of these new sample functions.

Proposition 1.9. *Let X be a characteristic with the property that for $k \in \mathbb{N}$, the theoretical moment $\nu_{2k} = \nu_{2k}(X) = E\left(X^{2k}\right)$ exists. Then*

$$E\left(\overline{\nu}_k\right) = \nu_k \ \text{ and } \ V\left(\overline{\nu}_k\right) = \frac{1}{n}\left(\nu_{2k} - \nu_k^2\right). \tag{1.5}$$

Corollary 1.10. *Let X be a characteristic as in Proposition 1.9 and for $n \in \mathbb{N}$, let*

$$Z_n = \frac{\overline{\nu}_k - \nu_k}{\sqrt{\dfrac{\nu_{2k} - \nu_k^2}{n}}}.$$

Then $Z_n \xrightarrow{d} Z$, as $n \to \infty$.

Proposition 1.11. *Let X be a characteristic with $V(X) = \mu_2 = \sigma^2$ and for which the theoretical moment $\nu_4 = E\left(X^4\right)$ exists. Then*

$$\begin{aligned} E\left(\overline{\mu}_2\right) &= \frac{n-1}{n}\,\sigma^2, \\ V\left(\overline{\mu}_2\right) &= \frac{n-1}{n^3}\Big[(n-1)\mu_4 - (n-3)\sigma^4\Big]. \end{aligned} \tag{1.6}$$

Remark 1.12. Notice that the sample central moment of order 2 is the first statistic whose expected value *is not* the corresponding population function, in this case the theoretical variance. This is the motivation for the next definition.

Definition 1.13. *The statistic*

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2 \tag{1.7}$$

*is called the **sample variance** and its value is $\dfrac{1}{n-1}\displaystyle\sum_{i=1}^{n}(x_i - \overline{x})^2$.*
*The statistic $s = \sqrt{s^2}$ is called the **sample standard deviation**.*

Remark 1.14. With this definition, we have for the sample variance

$$E\left(s^2\right) \;=\; E\left(\frac{n}{n-1}\overline{\mu}_2\right) \;=\; \mu_2 \;=\; \sigma^2. \tag{1.8}$$

So, for the rest of this chapter, we will use these notations:

Function	Population (theoretical)	Sample
Mean	$\mu = E(X)$	$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$
Variance	$\sigma^2 = V(X)$	$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$
Standard deviation	$\sigma = \sqrt{V(X)}$	$s = \sqrt{s^2}$
Moment of order k	$\nu_k = E\left(X^k\right)$	$\overline{\nu}_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$
Central moment of order k	$\mu_k = E\left[(X - E(X))^k\right]$	$\overline{\mu}_k = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^k$

Table 1: Notations

2 Estimation; Basic Notions

We will refer to the parameter to be estimated as the **target parameter** and denote it by θ.
Two types of estimation will be considered: **point estimate**, when the result of the estimation is one single value and **interval estimate**, when the estimate is an interval enclosing the value of the target parameter. In either case, the actual estimation is accomplished by an **estimator**, a rule, a formula, or a procedure that leads us to the value of an estimate, based on the data from a sample.

Throughout this chapter, we consider a characteristic X (relative to a population), whose pdf $f(x;\theta)$ depends on the parameter θ, which is to be estimated.

As before, we consider a random sample of size n, i.e. sample variables $X_1, \dots, X_n$, which are **independent and identically distributed (iid)**, having the same pdf as X.

A **point estimator** for (the estimation of) the target parameter θ is a sample function (statistic)

$$\overline{\theta} = \overline{\theta}(X_1, X_2, \dots, X_n).$$

Other notations may be used, such as $\hat{\theta}$ or $\tilde{\theta}$.

Each statistic is a random variable because it is computed from random data. It has a so-called *sampling distribution* (a pdf). Each statistic estimates the corresponding population parameter and adds certain information about the distribution of X, the variable of interest. The value of the point estimator, the **point estimate**, is the actual approximation of the unknown parameter.

Many different point estimators may be obtained for the same target parameter. Some are considered "good", others "not so good", some "better" than others. We need some criteria to decide on one estimator versus another.

For one thing, it is highly desirable that the sampling distribution of an estimator $\overline{\theta}$ to be "clustered" around the target parameter. In simple terms, we *expect* that the value the point estimator provides to be the actual value of the parameter it estimates. This justifies the following notion.

Definition 2.1. *A point estimator $\overline{\theta}$ is called an **unbiased** estimator for θ if*

$$E(\overline{\theta}) \quad = \quad \theta. \tag{2.1}$$

*The **bias** of $\overline{\theta}$ is the value $B = E(\overline{\theta}) - \theta$.*

Unbiasedness means that in the long-run, collecting a large number of samples and computing $\overline{\theta}$ from each of them, on the average, we hit the unknown parameter θ exactly. In other words, in a long run, unbiased estimators neither underestimate, nor overestimate the parameter.

Example 2.2.
1. Recall from Proposition 1.4 that for the sample mean, as a random variable, we have $E(\overline{X}) = \mu$. Thus the sample mean is an unbiased estimator for the population mean.
2. Also, by Proposition 1.9, the sample moment of order k is an unbiased estimator for the theoretical moment of order k. 3. By Proposition Proposition 1.11, the sample central moment of order 2 *is not* an unbiased estimator for the population central moment of order 2 (or it is a *biased* estimator), since

$$E(\overline{\mu}_2) = \frac{n-2}{n}\mu_2 \neq \mu_2 = \sigma^2.$$

3. However, the sample variance

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

is an unbiased estimator for the population variance, since $E(s^2) = \sigma^2$. That was the main reason for the way the sample variance was defined.

Another desirable trait for a point estimator is that its values do not vary too much from the value of the target parameter. So we need to evaluate variability of computed statistics and especially parameter estimators. That can be accomplished by computing the following statistic.

Definition 2.3. *The **standard error** of an estimator $\overline{\theta}$, denoted by $\sigma_{\overline{\theta}}$, is its standard deviation*

$$\sigma_{\overline{\theta}} = \sigma(\overline{\theta}) = \mathrm{Std}(\overline{\theta}) = \sqrt{V(\overline{\theta})}.$$

Both population and sample variances are measured in squared units. Therefore, it is convenient to have standard deviations that are comparable with our variable of interest, X. As a measure of variability, standard errors show precision and reliability of estimators. They show how much estimators of the same target parameter θ can vary if they are computed from different samples. Ideally, we would like to deal with unbiased or nearly unbiased estimators that have *low* standard error.

3 Estimation by Confidence Intervals

3.1 Confidence Intervals, General Framework

Point estimators provide 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 $\alpha \in (0, 1)$,

$$P(\overline{\theta}_L \le \theta \le \overline{\theta}_U) \;=\; 1 - \alpha. \tag{3.1}$$

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 3.1.
1. It may seem a little peculiar that we use $1-\alpha$ instead of simply α in (3.1), 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 (3.1) *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.

We will use the pivotal method to find $100(1-\alpha)\%$ CI's. 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 known that

$$Z \;=\; \frac{\overline{\theta}-\theta}{\sigma_{\overline{\theta}}} \;\left(= \;\frac{\overline{\theta}-E(\overline{\theta})}{\sigma(\overline{\theta})}\right) \tag{3.2}$$

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. Since the pdf of Z is known, we can choose two values, Z_L, Z_U such that for a given $\alpha \in (0,1)$,

$$P(Z_L \leq Z \leq Z_U) \;=\; 1-\alpha. \tag{3.3}$$

How to choose them? Of course, there are infinitely many possibilities. Recall *quantiles*. A quantile of a given order $\alpha \in (0,1)$ for a random variable X, is a value q_α with the property that

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

i.e., that the area under the graph of the pdf, to the *left* of q_α is α (Figure 1).

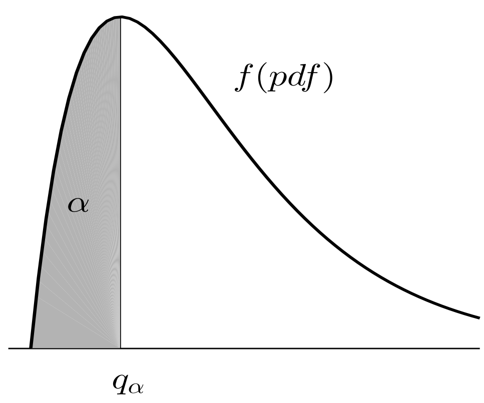

Fig. 1: Quantile of order $\alpha \in (0,1)$

Recall that for continuous random variables, the probability in (3.3) 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 2.

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 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 (3.3).

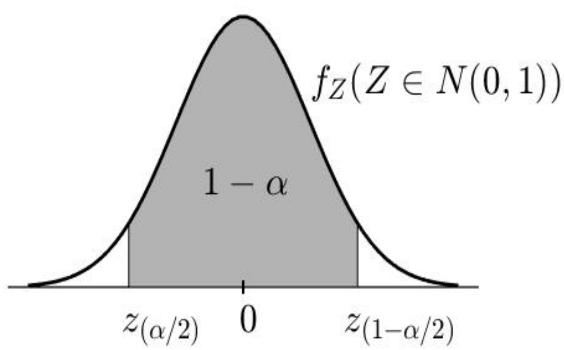

Fig. 2: Confidence interval for the $N(0,1)$ distribution

From here, we proceed to rewrite the inequality inside, until we get the limits of the CI for θ. We have

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

$$= \quad P\left(\overline{\theta} - \sigma_{\overline{\theta}} \cdot z_{1-\frac{\alpha}{2}} \leq \theta \leq \overline{\theta} - \sigma_{\overline{\theta}} \cdot z_{\frac{\alpha}{2}}\right),$$

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

$$\left[\overline{\theta} - \sigma_{\overline{\theta}} \cdot z_{1-\frac{\alpha}{2}},\ \overline{\theta} - \sigma_{\overline{\theta}} \cdot z_{\frac{\alpha}{2}}\right]. \tag{3.4}$$

Remark 3.2.
1. Since the Standard Normal distribution is symmetric about the origin, $z_{\frac{\alpha}{2}} = -z_{1-\frac{\alpha}{2}}$ 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.
3. In what follows, 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.

3.2 Confidence Intervals for the Mean and Variance of One Population

Let X be a population characteristic, with mean $\mu = E(X)$ and variance $V(X) = \sigma^2$, whose pdf depends on a parameter θ, $f(x;\theta)$. Let $X_1, X_2, \ldots, X_n$ be a sample drawn from the pdf of X.
The formulas for finding confidence intervals for the mean μ and variance σ^2 are based on the following results (which follow either from properties of random variables, or are the consequence of some CLT).

Proposition 3.3. *Assume $X \in N(\mu, \sigma)$. Then*

$Z = \dfrac{\overline{X} - \mu}{\dfrac{\sigma}{\sqrt{n}}} \in N(0,1),\ T = \dfrac{\overline{X} - \mu}{\dfrac{s}{\sqrt{n}}} \in T(n-1)$ *and*

$$V = \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \overline{X})^2 = \frac{(n-1)\,s^2}{\sigma^2} \in \chi^2(n-1).$$

Proposition 3.4. *If the sample size is large enough ($n > 30$), then*

$$Z = \frac{\overline{X} - \mu}{\dfrac{\sigma}{\sqrt{n}}} \in N(0,1) \text{ and } T = \frac{\overline{X} - \mu}{\dfrac{s}{\sqrt{n}}} \in T(n-1).$$

CI for the mean, known variance

If either $X \in N(\mu, \sigma)$ or the sample is large enough ($n > 30$) and σ is known, then by Propositions 3.3 and 3.4, we can use the pivot

$$Z = \frac{\overline{X} - \mu}{\dfrac{\sigma}{\sqrt{n}}} \in N(0,1).$$

The procedure will go *exactly* as described in the previous section, with $\theta = \mu, \overline{\theta} = \overline{X}, \sigma_{\overline{\theta}} = \dfrac{\sigma}{\sqrt{n}}$.

The $100(1-\alpha)\%$ CI for the mean is given by

$$\mu \;\in\; \left[\overline{X} - z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}},\; \overline{X} - z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right]. \tag{3.5}$$

Since $N(0,1)$ is symmetric (and one quantile is the negative of the other), we can write it in short as

$$\overline{X} \pm z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} \quad \text{or} \quad \overline{X} \mp z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}. \tag{3.6}$$

CI for the mean, unknown variance

In practice, it is somewhat unreasonable to expect to know the value of σ, if the value of μ is unknown. We can find CI's for the mean, without knowing the variance. If either $X \in N(\mu, \sigma)$ or the sample is large enough ($n > 30$), then by Propositions 3.3 and 3.4, we can use the pivot

$$T = \frac{\overline{X} - \mu}{\dfrac{s}{\sqrt{n}}} \in T(n-1).$$

The same computations as before will lead to the $100(1-\alpha)\%$ CI for the mean:

$$\mu \;\in\; \left[\overline{X} - t_{1-\frac{\alpha}{2}}\frac{s}{\sqrt{n}},\; \overline{X} - t_{\frac{\alpha}{2}}\frac{s}{\sqrt{n}}\right]. \tag{3.7}$$

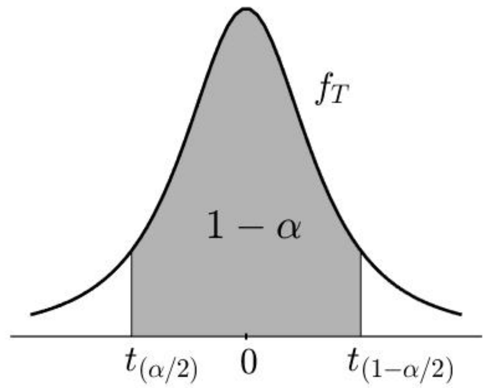

Fig. 3: Confidence interval for the T distribution

Notice that we change the notations for the quantiles, according to the pdf of the pivot (z for $N(0,1)$, t for $T(n-1)$, etc.). The Student $T(n-1)$ is also symmetric (see Figure 3), so again, we can write the CI in short as

$$\overline{X} \pm t_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}} \quad \text{or} \quad \overline{X} \mp t_{1-\frac{\alpha}{2}} \frac{s}{\sqrt{n}}. \tag{3.8}$$

CI for the variance

By Proposition 3.3, if $X \in N(\mu, \sigma)$, then we can use the pivot

$$V = \frac{(n-1)\, s^2}{\sigma^2} \in \chi^2(n-1).$$

Let us see how to do that. Even though the $\chi^2(n-1)$ is not symmetric (see Figure 4), so we cannot really talk about the "middle" for the area, we can still use the quantiles as before. So, we have:

$$\begin{aligned} 1-\alpha &= P\left(\chi^2_{\frac{\alpha}{2}} \le V \le \chi^2_{1-\frac{\alpha}{2}}\right) \\ &= P\left(\chi^2_{\frac{\alpha}{2}} \le \frac{(n-1)\, s^2}{\sigma^2} \le \chi^2_{1-\frac{\alpha}{2}}\right) \\ &= P\left(\frac{1}{\chi^2_{1-\frac{\alpha}{2}}} \le \frac{\sigma^2}{(n-1)\, s^2} \le \frac{1}{\chi^2_{\frac{\alpha}{2}}}\right) \\ &= P\left(\frac{(n-1)\, s^2}{\chi^2_{1-\frac{\alpha}{2}}} \le \sigma^2 \le \frac{(n-1)\, s^2}{\chi^2_{\frac{\alpha}{2}}}\right). \end{aligned}$$

Thus, a $100(1-\alpha)\%$ CI for the variance is

$$\sigma^2 \in \left[\frac{(n-1)\,s^2}{\chi^2_{1-\frac{\alpha}{2}}},\ \frac{(n-1)\,s^2}{\chi^2_{\frac{\alpha}{2}}}\right] \tag{3.9}$$

and one for the standard deviation is

$$\sigma \in \left[\sqrt{\frac{(n-1)\,s^2}{\chi^2_{1-\frac{\alpha}{2}}}},\ \sqrt{\frac{(n-1)\,s^2}{\chi^2_{\frac{\alpha}{2}}}}\right] \tag{3.10}$$

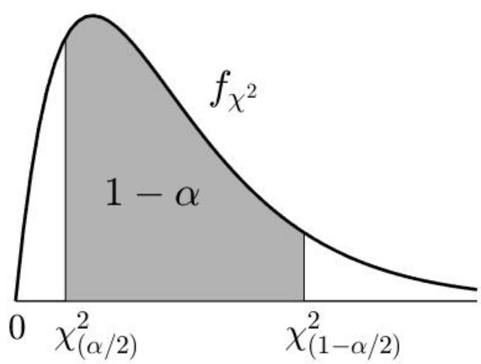

Fig. 4: Confidence interval for the χ^2 distribution

Remark 3.5.
1. Remember, "χ^2_α" is just a notation for the quantile of order α for the $\chi^2(n-1)$ distribution, it *does not* mean you have to take the square of it!
2. Since the $\chi^2(n-1)$ is no longer symmetric, there is no relationship between the two quantiles, we have to use *both* and there is no shorter writing for the CI for the variance than the one in (3.9) (or (3.10) for the standard deviation).

Example 3.6. The time spent for finding a parking space downtown Cluj-Napoca during the week was recorded for 64 drivers. The average and variance were found to be 15 minutes and 256 minutes, respectively. Find a 95% confidence interval for the true average time spent to find a parking spot during the week in downtown Cluj-Napoca.

Solution. The population here is the set of times spent to find a parking space downtown Cluj by *all* people who need to park downtown. We want to estimate its *average*, so the mean μ.

For our sample, $n = 64$, $\overline{X} = 15$ and $s^2 = 256$. To attain a confidence level of $1 - \alpha = 0.95$, we need $\alpha = 0.05$ and $\alpha/2 = 0.025$. Since σ is not known, we use formula (3.7) (or, actually, (3.8)). The quantiles for the $T(63)$ distribution are $t_{0.025} = -1.9983$, $t_{0.975} = 1.9983$ and the 95% CI for the mean is

$$\left[\overline{X} \pm t_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}}\right] = [11.0034,\ 18.9966].$$

So

$$\mu \in [11.0034,\ 18.9966]\,,$$

with probability 0.95. The interpretation is that 95% of the drivers spend, on average, between 11.0034 and 18.9966 minutes trying to find a parking space downtown Cluj-Napoca on weekdays.

$\blacksquare$