3 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 sections. 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.

3.1 Random Samples and Sample Functions

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, $x_1, \dots, x_n$, the observed values of the sample variables $X_1, \dots, X_n$.

Definition 3.1. *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 $X_1, \dots, X_n$, having the same distribution as X. The variables $X_1, \dots, X_n$, are called **sample variables** and their observed values $x_1, \dots, x_n$, are called **sample data**.*

Remark 3.2. The term *random sample* may refer to the objects selected, to the sample variables, or to the sample data. It is usually clear from the context which meaning is intended. In general, we use capital letters to denote sample variables and corresponding lowercase letters for their observed values, the sample data.

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

Definition 3.3. *A **sample function** or **statistic** is a random variable*

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

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

We will revisit now some sample numerical characteristics discussed in the previous sections 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.

3.2 Sample Mean

Definition 3.4. *The **sample mean** is the sample function defined by*

$$\overline{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i \tag{3.1}$$

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

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

Proposition 3.5. *Let X be a population characteristic with mean $E(X) = \mu$ and variance $V(X) = \sigma^2$. Then*

$$E\left(\overline{X}\right) = \mu \text{ and } V\left(\overline{X}\right) = \frac{\sigma^2}{n}. \tag{3.2}$$

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

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

□

Remark 3.6. As a consequence, the standard deviation of $\overline{X}$ is

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

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.

3.3 Sample Moments and Sample Variance

Definition 3.7. *The statistic*

$$\overline{\nu}_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k \tag{3.3}$$

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

The statistic

$$\overline{\mu}_k = \frac{1}{n}\sum_{i=1}^{n} (X_i - \overline{X})^k \tag{3.4}$$

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 3.8. 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 characteristics of these new sample functions.

Proposition 3.9. *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*

$$E\left(\overline{\nu}_k\right) = \nu_k \;\; \textit{and} \;\; V\left(\overline{\nu}_k\right) = \frac{1}{n}\left(\nu_{2k} - \nu_k^2\right). \tag{3.5}$$

Proof. First off, the condition that ν_{2k} exists for X ensures the fact that all theoretical moments of

X of order up to k also exist. The rest follows as before. We have

$$E\left(\overline{\nu}_k\right) = \frac{1}{n}\sum_{i=1}^{n} E(X_i^k) = \frac{1}{n}\sum_{i=1}^{n} E(X^k) = \frac{1}{n}\, n\nu_k = \nu_k$$

and

$$\begin{aligned} V\left(\overline{\nu}_k\right) &= \frac{1}{n^2}\sum_{i=1}^{n} V(X_i^k) &&= \frac{1}{n^2}\sum_{i=1}^{n} V(X^k) \\ &= \frac{1}{n^2}\, n\left(\nu_{2k} - \nu_k^2\right) &&= \frac{1}{n}\left(\nu_{2k} - \nu_k^2\right). \end{aligned}$$

□

Proposition 3.10. *Let X be a characteristic with variance $V(X) = \mu_2 = \sigma^2$ and for which the theoretical moment $\nu_4 = E\left(X^4\right)$ exists. Then*

$$E\left(\overline{\mu}_2\right) = \frac{n-1}{n}\,\sigma^2, \tag{3.6}$$
$$V\left(\overline{\mu}_2\right) = \frac{n-1}{n^3}\Big[(n-1)\mu_4 - (n-3)\sigma^4\Big].$$

Remark 3.11. 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 3.12. *The statistic*

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2 \tag{3.7}$$

*is called the **sample variance** and its value is* $\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2$.

*The statistic $s = \sqrt{s^2}$ is called the **sample standard deviation**.*

Remark 3.13. Notice that the sample central moment of order 2 is no longer equal to the sample variance, as we are used. In fact, we have

$$s^2 = \frac{n}{n-1}\,\overline{\mu}_2.$$

Then, by Proposition 3.10, we have for the sample variance

$$E\left(s^2\right) = \mu_2 = \sigma^2, \tag{3.8}$$

$$V\left(s^2\right) \quad = \quad \frac{1}{n(n-1)}\Big[(n-1)\mu_4 - (n-3)\sigma^4\Big]$$

and, again, the estimation of σ^2 by s^2 (or of σ by s) has a standard error that decreases as the sample size increases:

$$\mathrm{Std}(s^2) \;=\; \sqrt{\frac{1}{n(n-1)}\Big((n-1)\mu_4 - (n-3)\sigma^4\Big)} \;\longrightarrow\; 0, \;\; \text{as} \;\; n \to \infty.$$

3.4 Sample Proportions

Definition 3.14. *Assume a subpopulation A of a population consists of items that have a certain attribute. The* ***population proportion*** *is then the probability*

$$p \quad = \quad P(i \in A), \tag{3.9}$$

i.e. the probability for a randomly selected item i to have this attribute.
The ***sample proportion*** *is*

$$\overline{p} \quad = \quad \frac{\text{number of sampled items from } A}{n}, \tag{3.10}$$

where n is the sample size.

Proposition 3.15. *Let p be a population proportion. Then*

$$E\left(\overline{p}\right) = p, \;\; V\left(\overline{p}\right) = \frac{p(1-p)}{n} = \frac{pq}{n} \;\; \textit{and} \;\; \sigma\left(\overline{p}\right) = \sqrt{\frac{pq}{n}}, \tag{3.11}$$

where $q = 1 - p$.

Proof. We use the indicator random variable

$$X_i \quad = \quad \begin{cases} 1, & i \in A \\ 0, & i \notin A \end{cases}.$$

Then $X_i \in Bern(p)$ and, so, we know that $E(X_i) = p$ and $V(X_i) = pq$, for every $i = 1, \ldots, n$. But notice that $\overline{p} = \frac{1}{n}\sum_{i=1}^{n} X_i$, i.e. the sample mean of the sample $X_1, \ldots, X_n$. Thus, by Proposition

3.5, we have

$$\begin{aligned} E(\overline{p}) &= p, \\ V(\overline{p}) &= \frac{pq}{n}, \\ \sigma(\overline{p}) &= \sqrt{\frac{pq}{n}}. \end{aligned}$$

□

3.5 Sample Functions for Comparing Two Populations

It will be necessary sometimes to compare characteristics of two populations. For that, we will need results on sample functions referring to both collections. Assume we have two characteristics $X_{(1)}$ and $X_{(2)}$, relative to two populations. We draw from both populations independent random samples of sizes n_1 and n_2, respectively. Denote the two sets of random variables by

$$X_{11}, \ldots, X_{1n_1} \text{ and } X_{21}, \ldots, X_{2n_2}.$$

Then we have two sample means and two sample variances, given by

$$\overline{X}_1 = \frac{1}{n_1}\sum_{i=1}^{n_1} X_{1i}, \quad \overline{X}_2 = \frac{1}{n_2}\sum_{j=1}^{n_2} X_{2j}$$

and

$$s_1^2 = \frac{1}{n_1 - 1}\sum_{i=1}^{n_1}\left(X_{1i} - \overline{X}_1\right)^2, \;\; s_2^2 = \frac{1}{n_2 - 1}\sum_{j=1}^{n_2}\left(X_{2j} - \overline{X}_2\right)^2,$$

respectively. In addition, denote by

$$s_p^2 = \frac{\sum_{i=1}^{n_1}\left(X_{1i} - \overline{X}_1\right)^2 + \sum_{j=1}^{n_2}\left(X_{2j} - \overline{X}_2\right)^2}{n_1 + n_2 - 2} = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

the **pooled variance** of the two samples, i.e. a variance that considers (pools) the data from both samples.

In inferential Statistics, when comparing the means of two populations, we will look at their difference and try to estimate it. Regarding that, we have the following result.

Proposition 3.16. *Let $X_{(1)}, X_{(2)}$ be two population characteristics with means $E(X_{(i)}) = \mu_i$ and variances $V(X_{(i)}) = \sigma_i^2, i = 1, 2$. Then*

$$\begin{aligned} E\left(\overline{X}_1 - \overline{X}_2\right) &= \mu_1 - \mu_2, \\ V\left(\overline{X}_1 - \overline{X}_2\right) &= \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}. \end{aligned} \tag{3.12}$$

In a similar fashion, we can compare two population proportions. Again, the random variable of interest is their difference.

Proposition 3.17. *Assume we have two population proportions p_1 and p_2. From each population we draw independent samples of size n_1 and n_2, respectively, which yield the population proportions $\overline{p}_1$ and $\overline{p}_2$. Then*

$$\begin{aligned} E\left(\overline{p}_1 - \overline{p}_2\right) &= p_1 - p_2, \\ V\left(\overline{p}_1 - \overline{p}_2\right) &= \frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}, \end{aligned} \tag{3.13}$$

with $q_i = 1 - p_1, i = 1, 2$.

Summary of Notations

Notations of the sample functions and their corresponding population characteristics.

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	$\mu_k = E\left[(X - E(X))^k\right]$	$\overline{\mu}_k = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^k$
Proportion	$p = P(i \in A)$	$\overline{p} = \frac{\text{number of } X_i \text{ from } A}{n}$

Table 1: Notations

Chapter 4. Inferential Statistics

Populations are characterized by parameters. The goal of Inferential Statistics is to make inferences (estimates) about one or more population parameters on the basis of a sample.

1 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.

In this chapter, we discuss how

- *to estimate parameters of the distribution*. The methods in the previous chapter mostly concern measure of location (mean, median, quantiles) and variability (variance, standard deviation, interquartile range). As we know, this does not cover all possible parameters, and thus, we still lack a general methodology of estimation.

- *to construct confidence intervals*. Any estimator, computed from a collected random sample instead of the whole population, is understood as only an approximation of the corresponding parameter. Instead of one estimator that is subject to a sampling error, it is often more reasonable to produce an interval that will contain the true population parameter with a certain known high probability.

- *to test hypotheses*. That is, we shall use the collected sample to verify statements and claims about the population. As a result of each test, a statement is either rejected on basis of the observed data or accepted (not rejected). Sampling error in this analysis results in a possibility of wrongfully accepting or rejecting the hypothesis; however, we can design tests to control the probability of such errors.

Results of such statistical analysis are used for making decisions under uncertainty, developing optimal strategies, forecasting, evaluating and controlling performance and so on.

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. If X is discrete, then f represents the probability distribution function, while if X is continuous, f is the probability density function.

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. The notations introduced in the previous chapter for some sample functions still stand.

Definition 1.1. *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*. 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.

2 The Normal, Student (T) and Fisher (F) Distributions

2.1 Normal Distribution $N(\mu, \sigma)$

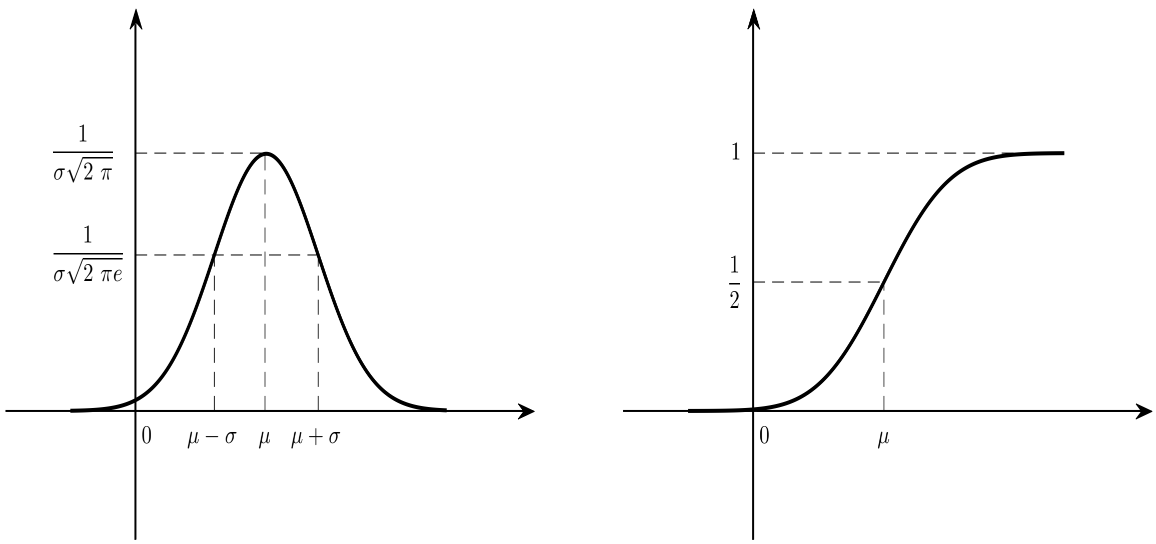
The Normal distribution is, by far, the most important distribution, underlying many of the modern statistical methods used in data analysis. It was first described in the late 1700's by De Moivre, as a limiting case for the Binomial distribution (when n, the number of trials, becomes infinite), but did not get much attention. Half a century later, both Laplace and Gauss (independently of each other) rediscovered it in conjunction with the behavior of errors in astronomical measurements. It is also referred to as the "Gaussian" distribution.

A random variable X has a **Normal** distribution (norm) with parameters $\mu \in \mathbb{R}$ and $\sigma > 0$, if its pdf is

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad x \in \mathbb{R}. \tag{2.1}$$

The cdf of a Normal variable is then given by

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{x-\mu}{\sigma}} e^{-\frac{t^2}{2}} dt. \tag{2.2}$$



(a) Density Function (pdf) (b) Cumulative Distribution Function (cdf)

Fig. 1: Normal Distribution

The graph of the Normal density is a symmetric, bell-shaped curve (known as "Gauss's bell" or "Gauss's bell curve") centered at the value of the first parameter μ, as can be seen in Figure 1(a). The graph of the cdf of a Normally distributed random variable is given in Figure 1(b) and this is approximately what the graph of the cdf of *any* continuous random variable looks like.

Remark 2.1.

1. There is an important particular case of a Normal distribution, namely $N(0, 1)$, called the **Standard (or Reduced) Normal Distribution**. A variable having a Standard Normal distribution is usually denoted by Z. The density and cdf of Z are given by

$$f_Z(x) = \frac{1}{\sqrt{2\pi}} e^{-\dfrac{x^2}{2}}, \quad x \in \mathbb{R} \quad \text{and} \quad F_Z(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\dfrac{t^2}{2}}\, dt. \tag{2.3}$$

The function F_Z given in (2.3) is known as *Laplace's function* (or *the error function*) and its values can be found in tables or can be computed by most mathematical software.

3. As noticed from (2.2) and (2.3), there is a relationship between the cdf of any Normal $N(\mu, \sigma)$

variable X and that of a Standard Normal variable Z, namely

$$F_X(x) = F_Z\left(\frac{x-\mu}{\sigma}\right) .$$

Asymptotic Normality

By the Central Limit Theorem, the sum of observations, and therefore, the sample mean have approximately Normal distribution if they are computed from a large sample. That is, the distribution of $\overline{X}$ is approximately $N\Big(\mu, \dfrac{\sigma}{\sqrt{n}}\Big)$ and that of

$$Z = \frac{\overline{X}-\mu}{\dfrac{\sigma}{\sqrt{n}}}$$

(which is the *reduced* variable of $\overline{X}$) is approximately Standard Normal ($N(0,1)$) as $n \to \infty$. This property is called *asymptotic normality*. The same is true for other statistics, e.g. the difference of means:

$$Z = \frac{\overline{X}_1 - \overline{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}} \longrightarrow N(0,1), \text{ as } n_1, n_2 \to \infty.$$

Quantiles

In many inferential statistical procedures, we will need to use *quantiles*. Recall that a quantile of a given order $\alpha \in (0,1)$ for a random variable X with cdf F, is a value q_α with the property that

$$F(q_\alpha) = P(X \le q_\alpha) = \alpha, \; q_\alpha = F^{-1}(\alpha),$$

i.e., that the area under the graph of the pdf, to the *left* of q_α is α (see Figure 2).

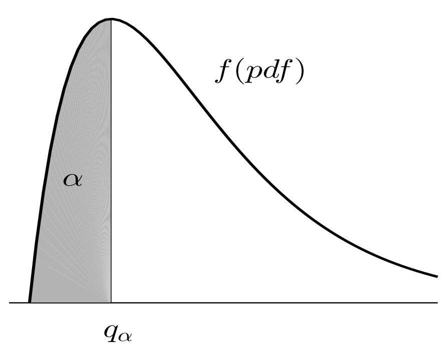


Fig. 2: Quantile of order $\alpha \in (0,1)$

For *symmetric* distributions, the symmetry is reflected in the computation of quantiles. By symmetry, we have

$$\begin{aligned} P(X \leq -q_\alpha) &\stackrel{\text{sym}}{=} P(X \geq q_\alpha) \;=\; 1 - P(X \leq q_\alpha) \\ &= 1 - \alpha \;=\; P(X \leq q_{1-\alpha}), \end{aligned}$$

therefore,

$$q_{1-\alpha} = -q_\alpha, \;\; \forall \alpha \in (0,1). \tag{2.4}$$

This is certainly the case for the Standard Normal distribution (see Figure 3):

$$z_{1-\alpha} = -z_\alpha, \;\; \forall \alpha \in (0,1).$$

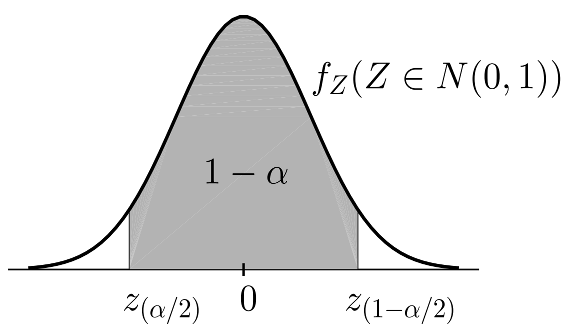


Fig. 3: Quantiles for the $N(0,1)$ distribution

2.2 Student (T) Distribution

The **Student (T)** distribution appeared as a necessity, when the sample size was small and asymptotic normality could not be used. It was developed in the early 1900's by W. S. Gosset under the pseudonym "Student". It has one parameter, denoted by n or ν or simply, df and it stands for "number of degrees of freedom". The T-distribution is symmetric and bell-shaped, like the Normal one, only it is narrower. Since it is symmetric, its quantiles also satisfy relation (2.4) (see Figure 4):

$$t_{1-\alpha} = -t_\alpha, \;\; \forall \alpha \in (0,1).$$

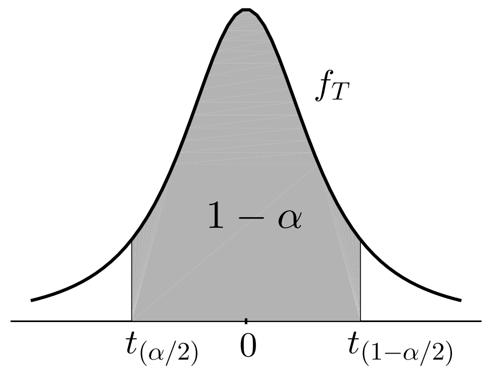


Fig. 4: Student T Distribution pdf and quantiles

2.3 Fisher (F) Distribution

In many cases, two population *variances* need to be compared. Such inference is used for the comparison of accuracy, stability, uncertainty, or risks arising in two populations.

Comparison of variances can be accomplished using the **Fisher-Snedecor (F) distribution**. This distribution was first considered in 1918 by a famous English statistician and biologist, Sir Ronald Fisher (1890-1962) and developed and formalized in 1934 by an American mathematician George Snedecor (1881-1974). A random variable X follows a *Fisher (F) distribution* with parameters $m, n \in \mathbb{N}$ (degrees of freedom), if its density function is

$$f(x) \;=\; \frac{1}{\beta(\frac{m}{2}, \frac{n}{2})} \left(\frac{m}{n}\right)^{\frac{m}{2}} x^{\frac{m}{2}-1} \left(1 + \frac{m}{n}x\right)^{-\frac{m+n}{2}}, \; x > 0,$$

where $\beta(a, b) = \int\limits_0^1 x^{a-1}(1-x)^{b-1}dx,\ a, b > 0$, is Euler's Beta function. Its density has a right-skewed shape (see Figure 5). Since this is asymmetric, we no longer have relation (2.4) between its quantiles. However, there is an important property of the F distribution:

Proposition 2.2. *If the variable X has a $F(m, n)$ distribution, then its reciprocal $\frac{1}{X}$ has a $F(n, m)$ distribution. As a consequence, the following relation holds for F-quantiles:*

$$f_{1-\alpha,m,n} \;=\; \frac{1}{f_{\alpha,n,m}},\ \forall\alpha \in (0, 1), \tag{2.5}$$

where the quantile $f_{1-\alpha,m,n}$ refers to the $F(m, n)$ distribution and $f_{\alpha,n,m}$ is for $F(n, m)$.

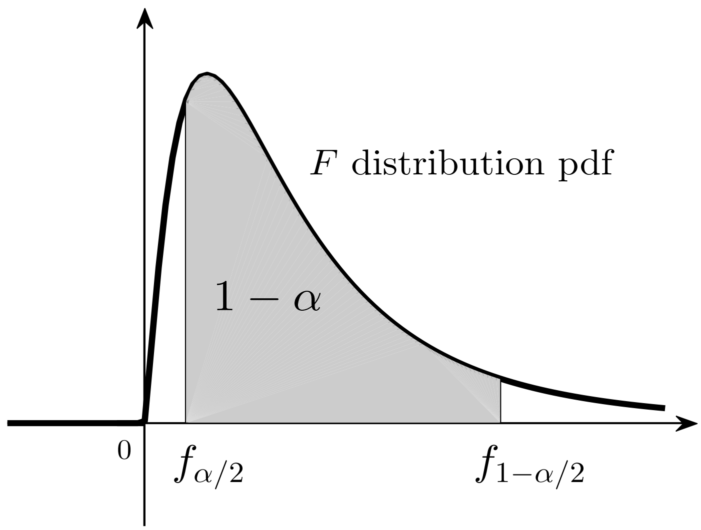


Fig. 5: Fisher (F) Distribution pdf and quantiles

3 Estimation by Confidence Intervals

3.1 Basic Concepts; General Framework

Unlike point estimators (that provide one single value), 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 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. 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 (possibly asymptotically) $N(0,1)$ distribution, so we can better understand the ideas.

Let θ be a target parameter and let $\overline{\theta}$ be a point estimator for θ such that $E(\overline{\theta}) = \theta$ (that means an *unbiased* estimator), 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 $\alpha \in (0,1)$,

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

How to choose them? Of course, there are infinitely many possibilities. Recall that for continuous random variables, the probability in (3.3) represents 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 3.

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).

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

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

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

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

Remark 3.2.
1. By the symmetry of $N(0,1)$ (and, hence, (2.4)) the CI can be written in short 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. 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.

One-sided confidence intervals

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.

• *Lower confidence interval* for θ
We want to find θ_U such that $P(\theta \le \theta_U) = 1-\alpha$. We have, successively.

$$\begin{aligned} 1-\alpha &= P(\theta \le \theta_U) \;=\; P(-\theta \ge -\theta_U) \\ &= P\Big(\frac{\overline{\theta}-\theta}{\sigma_{\overline{\theta}}} \ge \frac{\overline{\theta}-\theta_U}{\sigma_{\overline{\theta}}}\Big) \\ &= P\Big(Z \ge \frac{\overline{\theta}-\theta_U}{\sigma_{\overline{\theta}}}\Big). \end{aligned}$$

But we know that $P(Z \ge z_\alpha) = 1-\alpha$, so, by equating $\dfrac{\overline{\theta}-\theta_U}{\sigma_{\overline{\theta}}} = z_\alpha$, we get $\theta_U = \overline{\theta} - \sigma_{\overline{\theta}} \cdot z_\alpha$ and the lower CI

$$\big(-\infty, \overline{\theta} - \sigma_{\overline{\theta}} \cdot z_\alpha\big] \;=\; \big(-\infty, \overline{\theta} + \sigma_{\overline{\theta}} \cdot z_{1-\alpha}\big]\,,$$

the last equality coming from the symmetry of the quantiles $z_{1-\alpha} = -z_\alpha$.

• *Upper confidence interval* for θ
Similarly, to find θ_L such that $P(\theta \ge \theta_L) = 1-\alpha$, we use

$$\begin{aligned} 1-\alpha &= P(\theta \ge \theta_L) \;=\; P(-\theta \le -\theta_L) \\ &= P\Big(\frac{\overline{\theta}-\theta}{\sigma_{\overline{\theta}}} \le \frac{\overline{\theta}-\theta_L}{\sigma_{\overline{\theta}}}\Big) \\ &= P\Big(Z \le \frac{\overline{\theta}-\theta_L}{\sigma_{\overline{\theta}}}\Big) \;=\; P(Z \le z_{1-\alpha}), \end{aligned}$$

so the upper CI is

$$\big[\overline{\theta} - \sigma_{\overline{\theta}} \cdot z_{1-\alpha}, \infty\big) \;=\; \big[\overline{\theta} + \sigma_{\overline{\theta}} \cdot z_\alpha, \infty\big)\,.$$