LIS
Working Payer Series

No. 830

The Barycenter of the Distribution and Its Application to the Measurement of Inequality: The Balance of Inequality, the Gini Index, and the Lorenz Curve

Giorgio Di Maio

March 2022

Luxembourg Income Study (LIS), asbl

The barycenter of the distribution and its application to the measurement of inequality: The Balance of Inequality, the Gini index, and the Lorenz curve

Working paper version: 1.5

Date: March 10, 2022

Giorgio DI MAIO [*]

Università degli Studi dell'Insubria, Department of Economics,

Via Monte Generoso, 71 - 21100 Varese, Italy

giorgiodimaio@gmail.com

ORCiD: https://orcid.org/0000-0002-2461-0925

ResearchGate: https://www.researchgate.net/profile/Giorgio-Di-Maio

---

[*] Corresponding author.

This study is a new development of the "Balance of Inequality" research project, and in particular it is a development of some ideas in two previous working papers with Paolo Landoni (Di Maio & Landoni, 2015, 2017), which I want to thank for having supported the research in its early stages. Many people contributed comments to the research project. First of all, I wish to thank Michele Zenga, who gave me a lot of valuable advice, not just related to research, and Chiara Gigliarano and Michele Ciruzzi, which read and commented a previous version of this paper. I also wish to thank Alberto Bagnai, Giovanni Bono, Lisa Crosato, Lucia Dalla Pellegrina, Lorenzo Farci, Filippo Gaffuri Riva, Alessio Guandalini, Rocco Mosconi, Tommaso Rigon, and participants at the Seventh Meeting of the Society for the Study of Economic Inequality (ECINEQ) (New York, 2017), DEMS ReLunch Seminars (Milano, 2018), Second International Conference on Data Science and Social Research (Milano, 2019), International Conference on Distributions and Inequality Measures in Economics (DIME) (Milano, 2020), and LVI Scientific Meeting of the Italian Society of Economics, Demography and Statistics (SIEDS) (Firenze, 2021). Finally, I also wish to thank the LIS Support Team for the support provided for data analysis. The usual disclaimers apply.

The barycenter of the distribution and its application to the measurement of inequality: The Balance of Inequality, the Gini index, and the Lorenz curve

ABSTRACT

This paper introduces in statistics the notion of the barycenter of the distribution of a non-negative random variable Y with a positive finite mean μ_Y and the quantile function $Q(x)$. The barycenter is denoted by μ_X and defined as the expected value of the random variable X having the probability density function $f_X(x) = Q(x)/\mu_Y$.

For continuous populations, the Gini index is $2\mu_X - 1$, i.e., the normalization of the barycenter, which is in the range [0, 1/2], the concentration area is $\mu_X - 1/2$, and the Gini's mean difference is $4\mu_Y (\mu_X - 1/2)$. The same barycenter-based formulae hold for normalized discrete populations. The introduction of the barycenter allows for new economic, geometrical, physical, and statistical interpretations of these measures.

For income distributions, the barycenter represents the expected recipient of one unit of income, as if the stochastic process that leads to the distribution of the total income among the population was observable as it unfolds. The barycenter splits the population into two groups, which can be considered as "the winners" and "the losers" in the income distribution, or "the rich" and "the poor".

We provide examples of application to thirty theoretical distributions and an empirical application with the estimation of personal income inequality in Luxembourg Income Study Database's countries.

We conclude that the barycenter is a new measure of the location or central tendency of distributions, which may have wide applications in both economics and statistics.

KEYWORDS

Balance of Inequality, Balance of Inequality index, Barycenter, BOI index, Concentration, Concentration area, Concentration ratio, Gini index, Gini mean difference, Inequality, Income inequality, Lorenz curve, Pen parade, Quantile function.

JEL CLASSIFICATION

C10 – Econometric and Statistical Methods and Methodology: General

C18 – Econometric and Statistical Methods and Methodology: Methodological Issues: General

D31 – Distribution: Personal Income, Wealth, and Their Distributions

D63 – Welfare Economics: Equity, Justice, Inequality, and Other Normative Criteria and Measurement

CONTENTS

INTRODUCTION

Traditionally, the study of the distribution of a random variable of interest, Y [1], in a given population has been done, both theoretically and empirically, using two different but related approaches. The first approach, which we label as the Y-perspective, has been to study at first the frequency, i.e., the probability mass or density function, $f_Y(y)$, and then also the cumulative distribution function, $F_Y(y)$, of the variable of interest (Burr, 1942). The second approach, which we label as the X-perspective, has been to study what is now known as the quantile function, $Q(x) = F_Y^{-1}(x)$,[2] and the Lorenz curve (Lorenz, 1905). One of the most widely used applications of this approach, in particular concerning the distribution of income and wealth, is of course the Gini index (Gini, 1914, 2005).

In this article, we propose that these two approaches can be unified, enabling a greater understanding of the statistical analysis and many advancements, both theoretical and empirical, in particular concerning the measurement of inequality.

In more detail, assuming a non-negative variable of interest Y, e.g., income, with a positive finite mean, μ_Y, we propose to complete the X-perspective by introducing a second random variable, denoted by X, having the probability density function $f_X(x) = Q(x)/\mu_Y$, and the cumulative distribution function, $F_X(x)$, obtained integrating $f_X(x)$. With this addition, the distribution is associated with five characteristic functions: two, i.e., $f_Y(y)$ and $F_Y(y)$, related to the Y-perspective, and three, i.e., $Q(x)$, $f_X(x)$, and $F_X(x)$, related to the X-perspective. Furthermore, we show that the Lorenz curve corresponds with the portion of the graph of $F_X(\mathrm{x})$ in the range $0 \le x \le 1$.

Since two density functions are now associated with the same distribution, i.e., $f_Y(y)$ and $f_X(x)$, it is immediate to recognize the existence of two expectations. The first expected value is the mean of the distribution, $\mu_Y = E[Y]$, and we denote the second expectation as the *barycenter* of the distribution, $\mu_X = E[X]$.[3]

We show that there is no general relation between the mean and the barycenter, i.e., when there is a relation between them it is distribution-specific. Therefore, the barycenter is a new measure of central tendency or location, which adds to the mean, median, and mode.

---

[1] Note that we use Y for the variable of interest instead of the X commonly used in statistics. The reason for this choice is that the quantile function is the most important characteristic function in our paper, and it has the variable of interest on the y-axis. This same convention was used for example by Pietra (1915, 2014) and Nygård & Sandström (1981). The convention prevalent today in statistics, but not in econometrics, of denoting by X the variable of interest is probably due to the prevalence of the approach based on the probability mass or density function and cumulative distribution function which have the variable of interest on the x-axis.

[2] Indeed, whenever a scholar considered "*n* ordered quantities representing the value of a given variable observed on *n* units" (e.g., Pietra, 1915, 2014, p. 6), he was actually considering the (empirical) quantile function.

[3] Center of mass, center of gravity, and barycenter are synonyms (e.g., https://www.merriam-webster.com/dictionary/barycenter). We denote the expected value of X as barycenter to emphasize the physical analogy resulting from the fact that, by definition, the expectation of a random variable is the center of mass of its probability mass/density function (e. g., Bulmer, 1979, p. 52; Mood et al., 1974, p. 65). Unlike other languages in which it is widely used, the term barycenter is rare in the common use of the English language and can be easily associated with a new measure. It is a concise term of Greek derivation that recalls that the concept of center of mass was introduced by Archimedes (Archimedes, Eutocios, & Heiberg, 1881; Archimedes & Heath, 1897a) as "κέντρον τοῦ βάρεος" (kéntron tu báreos) (Archimedes & Heath, 1897b, p. 181), i.e., center of weight.

The introduction of the barycenter is particularly important for the measurement of inequality. For income distributions, i.e., when Y is income, whereas the mean is the expected income, assuming to extract it at random from the distribution, the barycenter is the expected income recipient, assuming to assign at random, according to the distribution, one unit of income to the individuals in the population sorted in non-decreasing order of income.

Thus, whereas the Y-perspective focuses on the outcome of the data generating process, e.g., the income earned over a year, the X-perspective focuses on the stochastic process itself, i.e., the distribution of income as it unfolds, *as if* it was observable.[4] These two perspectives, and the five functions they comprise, are different but complementary ways of looking at the same distribution. Depending on the problem to be addressed, one or the other function may be more useful to analyze the same stochastic process.

We consider both continuous and discrete populations and we identify the normalized x-scale as the plotting position that allows us to apply to discrete populations the same barycenter-based formulae valid for continuous populations.

For continuous and normalized discrete populations, the barycenter assumes values between 1/2, in the case of perfect equality, when it coincides with the location of the median of the distribution, and 1, in the case of perfect inequality. The importance of the barycenter is highlighted by the following results. First, the Gini index is $2\mu_X - 1$, i.e., it is the normalization of the barycenter. Second, the concentration area between the egalitarian line and the Lorenz curve is $\mu_X - 1/2$. Third, the Gini's mean difference is $4\mu_Y (\mu_X - 1/2)$. Thus, whereas the Gini index is a function of the barycenter of the distribution only, the Gini's mean difference is a function of both the barycenter and the mean of the distribution.[5]

By introducing the barycenter of the distribution, we are also able to associate the Gini index, the Gini's mean difference, and the concentration area with new graphical representations, which combine new geometrical and physical interpretations, in addition to the geometrical interpretation of the Gini index associated with the Lorenz curve. In particular, since it is based on the idea of weighing the income distribution, i.e., $Q(x)$, using the barycenter as the indicator of its inequality, the Balance of Inequality provides a physical interpretation and geometrical representation of the Gini index that can be more easily understood by the general public.

We illustrate the proposed methodology and the properties of the barycenter by providing examples of application to thirty theoretical distributions. In particular, for each distribution, we obtain $Q(x)$, $f_X(x)$, μ_X, and the Balance of Inequality (=Gini) index. We also provide an empirical application by estimating personal income inequality in the countries in the Luxembourg Income Study (LIS) Database.

[4] Champernowne (1953) assumed the distribution of incomes between an enumerable infinity of income ranges as developing by means of a stochastic process. Instead, we propose to consider the final distribution of income as the result of a stochastic process that assigns the total income to the individuals in the population one unit at a time. In the X-perspective, the focus is shifted from the incomes to the income recipients.

[5] This result also explains why the variance of Y and the Gini's mean difference are distribution-specific functions of each other, i.e., there is no general relation between them (Jasso, 1979; van der Vaart, 1968).

We conclude that the barycenter is a new measure of the location or central tendency of distributions, which may have wide applications in both economics and statistics.

The paper is organized as follows. Section 1 provides some preliminary definitions and concepts. We recall the Gini's concentration ratio R and mean difference, and their relation with the Lorenz curve, and the concept of center of mass and its relation with the expected value of a random variable. Section 2 introduces the notion of the barycenter of the distribution, for both continuous and discrete populations, the normalized x-scale, and the normalized Lorenz curve for discrete populations. Section 3 introduces the Balance of Inequality (=Gini) index. Section 4 contains the application of the proposed methodology to thirty theoretical distributions. Section 5 shows the relation between the barycenter and the Gini's mean difference. Section 6 proposes new interpretations for the Gini index and clarifies or motivates in a new way some of its properties based on the barycenter of the distribution. Section 7 provides two estimators of the population's barycenter and Balance of Inequality (=Gini) index, one for a random sample from the population and one for weighted observations. Section 8 contains an empirical application with the estimation of personal income inequality in LIS countries. Finally, Section 9 discusses the main results obtained and concludes.

1. PRELIMINARY DEFINITIONS AND NOTIONS

This section provides some preliminary definitions and concepts that we use throughout the paper. First, we define the variable of interest and the two cases of continuous and discrete populations. Second, we recall the Gini index and the Gini's mean difference, and their relationship with the Lorenz curve. Third, we recall the concept of the center of mass and its relationship with the expected value of a random variable.

1.1 CONTINUOS POPULATIONS

Let a population be a set of two or more individuals.[6] We assume that the variable of interest, Y, is a non-negative (discrete or) continuous random variable representing a quantitative characteristic of the individuals in a given population, e.g., income. [7] We also assume that the total value of Y for the population, y_T, is positive.

Let $f_Y(y)$ and $F_Y(y)$ denote the probability density function and the cumulative distribution function of Y, respectively.

Let $Q(x)$ denote the quantile function of Y, which can be defined directly (e.g., the Tukey Lambda distribution) or, in the case of continuous populations, obtained in closed form or numerically as the inverse of $F_Y(y)$. We also assume that $Q(x) = 0$ for $x \notin [0,1]$.

We define a population as continuous when it has an infinite number of people, and we assume that the population is normalized so that it is represented by the argument of $Q(x)$ in the range [0,1]. Because $Q(x)$ is a non-decreasing function, i.e., $Q(x+dx) \geq Q(x)$, the population is implicitly assumed to be sorted in the non-decreasing value of the variable of interest.

[6] An individual alone is not a population.

[7] The assumption that the random variable Y be continuous is not essential, and the analysis can simply be extended to the case of a discrete variable. The case of a non-positive variable is analogous to the case of a non-negative variable since it can be dealt with by changing the sign of the variable.

For income distributions, i.e., when Y is income, the value $x = F_Y(y)$ is the probability of observing an income below $y = Q(x)$ and represents the share of the population with an income less than or equal to y (Gastwirth, 1971, p. 1037). Therefore, $y = Q(x)$ is the value of income such that the probability of observing an income less than or equal to y is x, i.e., $F_Y(y) = \Pr(Y \leq y) = x$, and represents the income of an individual at the xth percentile of the distribution (Mehran, 1976, p. 805), i.e., the income of an individual at x. Thus, each value of x in the range [0,1] represents both the x fraction of the population with an income below $y = Q(x)$ and the position of an individual in the population sorted in non-decreasing order of income that receives that income y.[8]

1.2 DISCRETE POPULATIONS

Let's define a population as discrete when it has a finite number $n \geq 2$ of people. Let assume that the individuals in the population are sorted in non-decreasing order of size of the variable of interest Y, and let y_i denote the value of Y for an individual with rank i in the population, $i = 1,...,n$, so that y_1 denotes the minimum value and y_n the maximum value of Y in the population.

For discrete populations, the specification of the quantile function, i.e., $y_i = Q(x_i)$, is possible only after solving the problem of defining the position x_i to be assigned on the x-axis to an individual having rank i in the population.[9] We deal with this issue in Section 2.2.

1.3 THE GINI INDEX, THE GINI'S MEAN DIFFERENCE, AND THE LORENZ CURVE

The Gini index is widely used in statistics and social sciences. However, the same name, Gini index or Gini coefficient, is used in the literature with different meanings. In this article, by Gini index we always and only refer to the concentration ratio R introduced by Corrado Gini in 1914 (Gini, 1914, 2005)[10], two years after introducing the mean difference (Gini, 1912), Δ, which we refer to as the Gini's mean difference.

Let's consider the case of a discrete population. Let p_i denote the cumulative share of the population including the first i individuals, defined by

$$(1) \qquad p_i \equiv \frac{i}{n}$$

Let q_i denote the cumulative share of the total value of the variable of interest of the first i individuals, defined by

---

[8] For the inverse probability integral transform, given a continuous uniform variable X in [0,1] and an invertible cumulative distribution function F_Y, the random variable $Y = F_Y^{-1}(X)$ has distribution F_Y (Devroye, 1986). Note that we do not define the random variable X as a continuous uniform variable in [0,1].

[9] Note that we do not adopt the widely used definition of the quantile function proposed by Gastwirth (1971), i.e., $F_Y^{-1}(x) = \inf\{y \in R : F_Y(y) \geq x\}$, because for continuous populations it agrees with the usual definition of the inverse function and for discrete populations it imposes a particular specification of the quantile function that disagrees with the analysis conducted in this study.

[10] Unfortunately, this article remained largely unknown for ninety years - as evidenced by the disconcerting absence of citations to it (and to the other Gini's publications) even in very influential studies on the Gini index (e.g., Atkinson, 1970; Dasgupta, Sen, & Starrett, 1973; Rothschild & Stiglitz, 1973) - up to its meritorious translation into English (Giorgi, 2005). It could be said that the literature on the Gini index has developed behind a veil of ignorance (e.g., Giorgi, 2014).

(2) $$q_i \equiv \frac{\sum_{k=1}^{i} y_k}{\sum_{i=1}^{n} y_i} = \frac{\sum_{k=1}^{i} y_k}{y_T}$$

Gini defined the concentration ratio *R* (Gini, 1914, p. 1207, eq. 11, 2005, p. 6, eq. 11) by

(3) $$R \equiv \frac{\sum_{i=1}^{n-1} (p_i - q_i)}{\sum_{i=1}^{n-1} p_i}.$$

After some manipulations, he also gave two equivalent expressions for the concentration ratio *R*, the first (Gini, 1914, p. 1208, eq. 12, 2005, p. 7, eq. 12) is

(4) $$R = 1 - \frac{2}{(n-1)y_T} \sum_{i=1}^{n-1} q_i \ ,$$

and the second (Gini, 1914, p. 1208, eq. 12 bis, 2005, p. 7, eq. 13) is

(5) $$R = \frac{2\sum_{i=1}^{n} (i-1) y_i}{(n-1)y_T} - 1.$$

In the same article, Gini derived the following relationship between the concentration ratio *R* and the mean difference (Gini, 1914, p. 1239, 2005, p. 30)

(6) $$R = \frac{\Delta}{2\mu_Y},$$

where μ_Y is the arithmetic mean, and the Gini's mean difference is defined (Gini, 1912, p. 36, eq. 28) by

(7) $$\Delta = \frac{1}{n(n-1)} \sum_{k=1}^{n} \sum_{i=1}^{n} \left| y_i - y_k \right|$$

Finally, always in the same article, Gini also indicated the possible association between the concentration ratio *R* and the Lorenz curve, i.e., the curve connecting the points $\{(0,0),...,(p_i,q_i),...,(1,1)\}$. He noted, albeit in a contradictory way (Gini, 1914, pp. 1231–1233, 2005, pp. 24–25), that the concentration ratio *R* is the *limit* to which the ratio between the area limited by the Lorenz curve and the egalitarian line (concentration area) and the area of the triangle below the egalitarian line, which represents the concentration area in the case of maximum concentration, tends, "*when the number n of observed cases increases but their distribution remains the same*".[11]

Soon afterward, Gaetano Pietra (1915, 2014, p. 8, eq. 12) defined the Gini's concentration ratio *R* from the Lorenz curve in the case of continuous populations, i.e., "when the number of observations is very large" (Pietra, 2014, p. 8)

(8) $$R = \frac{\Delta}{2\mu_Y} = \frac{1/2 - \int_0^1 L(x)\,dx}{1/2} = 1 - 2\int_0^1 L(x)\,dx$$

[11] Our translation, which we believe to be more accurate than "when the number *n* of cases increases and the distribution of the character is unchanged" (Gini, 2005, p. 24), of the sentence in Italian "quando cresce il numero *n* dei casi osservati, mantenendosi però uguale la loro distribuzione" (Gini, 1914, p. 1231).

Thus, denoting by L the area under the Lorenz curve, and by C the concentration area, i.e., the area limited by the Lorenz curve and the egalitarian line, for continuous populations, the concentration ratio R is

(9) $$R = 1 - 2L = 2C$$

Note that for both continuous and discrete populations, the concentration ratio R is normalized, i.e., the minimum possible value of concentration is 0 and the maximum possible value is 1, i.e., $0 \leq R \leq 1$. This is not true for the pseudo-Gini index that is also widely used in the literature on income inequality, which, for discrete populations, is $G = R\,(n-1)/n$.[12]

In conclusion, it is important to note that Gini did not define his concentration ratio R based on the Lorenz curve and that, in the case of discrete populations, the association between the Gini index and the Lorenz curve is problematic. We propose in Section 2.4 that the solution to this problem is the use of the normalized x-scale.

1.4 THE CENTER OF MASS IN PHYSICS AND STATISTICS: THE EXPECTED VALUE

The center of mass (or center of gravity or barycenter) is one of the most fundamental concepts of physics and particularly of classical mechanics. It is also an important concept in statistics, as it is well-known that the expectation of a random variable is the center of mass of its probability mass or density function (e. g., Bulmer, 1979, p. 52; Mood, Graybill, & Boes, 1974, p. 65). However, because the denominator is unitary, the formulae of the expected value reported in statistics manuals are simplified, and for our purposes, it is convenient to make them explicit starting from the definition of the center of mass used in physics.

Consider a system with n particles, each with a given mass, $m_i \geq 0$, $i = 1, ..., n$, whose position vectors are, respectively, $\mathbf{r}_1$, $\mathbf{r}_2$,…, $\mathbf{r}_n$. The center of mass of the system is defined (Finzi, 1991, pp. 256–268; Fowles & Cassiday, 2005, pp. 275–276; MacMahon, 2007, pp. 73–78; Vivarelli, 1992, pp. 94–98) as the point whose position vector $\mathbf{r}_{CM}$ is

(10) $$\mathbf{r}_{CM} = \frac{\sum_{i=1}^{n} m_i \mathbf{r}_i}{\sum_{i=1}^{n} m_i}$$

On the x-axis of a Cartesian coordinate system, the position of the center of mass of the system, x_{CM}, is

(11) $$x_{CM} = \frac{\sum_{i=1}^{n} m_i x_i}{\sum_{i=1}^{n} m_i}$$

Turning to statistics, let X be a discrete random variable with distinct values x_j, $j = 1, ..., m$, and probability mass function $f_X(x_j) = P[X = x_j]$. Let $f_X(x_j)$ denote the mass associated with the mass point x_j (e.g., Mood et al., 1974, p. 58). Then, the center of mass of $f_X(x_j)$ is the expected value of X

[12] The origins of the pseudo-Gini index can be traced back to the very influential manual written by Kendall (1945, p. 43 eq. 2.25), who failed to mention that (6), i.e., the relation between the concentration ratio R and the mean difference, holds only when the simple mean difference, i.e., (7), is used. It is truly unfair to Gini that many criticisms of his index are actually directed at this pseudo-Gini index.

(12) $$x_{CM} = \frac{\sum_{j=1}^{m} f_X(x_j)x_j}{\sum_{j=1}^{m} f_X(x_j)} = \sum_{j=1}^{m} f_X(x_j)x_j = E\left[X\right]$$

Replacing the probability mass function with the relative frequency or probability of each distinct value, n_j/n, you can also show that the center of mass of $f_X(x_j)$ is the arithmetic mean of the values assumed by X

(13) $$x_{CM} = \frac{\sum_{j=1}^{m} f_X(x_j)x_j}{\sum_{j=1}^{m} f_X(x_j)} = \frac{\sum_{j=1}^{m} n_j x_j / n}{\sum_{j=1}^{m} n_j / n} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

In the case of a continuous random variable X with probability density function $f_X(x)$, you can show that the center of mass of $f_X(x)$ is the expected value of X by integrating the (mass) density

(14) $$x_{CM} = \frac{\int_{-\infty}^{\infty} f_X(x)xdx}{\int_{-\infty}^{\infty} f_X(x)dx} = \int_{-\infty}^{\infty} f_X(x)xdx = E\left[X\right]$$

Finally, the center of mass of a generic function $f(x)$ is

(15) $$x_{CM} = \frac{\int_{-\infty}^{\infty} f(x)xdx}{\int_{-\infty}^{\infty} f(x)dx}$$

To represent this physical analogy, in the figures we indicate a thick line under the function of interest as if this function was loading on a thin beam supported by a fulcrum. A triangle below the line represents the center of mass, i.e., the fulcrum in the point at which the beam balances.[13] For instance, this graphical representation is applied to indicate the mean of the distribution of Y, i.e., the center of mass of $f_Y(y)$, in Figure 2-4 Panel 1.

2. THE BARYCENTER OF THE DISTRIBUTION

This section introduces the random variable X associated with the non-negative random variable of interest Y and the notion of the barycenter of the distribution of Y, which is defined as the expected value of X. We also derive the relationship between the barycenter, the area of concentration, and the Gini index, and we propose new geometrical representations for the latter two measures and a physical interpretation for the barycenter. First, we consider the case of continuous populations and then the case of discrete populations, after having introduced the x-scales and in particular the normalized x-scale, which allows us to establish a correspondence between continuous and discrete populations. Finally, we also define in this section three new measures that are useful to characterize a distribution: the location of the mean, the barycentric value, and the barycentric share.

2.1 CONTINUOUS POPULATIONS

In this section, we introduce the random variable X associated with the non-negative random variable of interest Y by defining its probability density function as the quantile function of Y divided by its mean. This

[13] Montgomery and Runger (2014, p. 74) used a similar graphical representation.

definition is the first key point of this section, and it allows us to define the cumulative distribution function, the expected value, and the variance of *X* by applying the standard definitions (e.g., Mood et al., 1974). The second key point of this section is the definition of the barycenter of the distribution of *Y* as the expected value of *X*. We also derive the relationship between the barycenter, the area of concentration, and the Gini index, and propose new geometrical representations for the latter two quantities. Finally, we define three new measures that are useful to characterize a distribution: the location of the mean, the barycentric value, and the barycentric share.

Before introducing the new definitions, we recall the standard definition of the mean of the distribution (e.g., Mood et al., 1974). The mean, μ_Y, or expectation of *Y* is:

(16) $$\mu_Y = E[Y] = \int_{-\infty}^{\infty} f_Y(y)ydy = \int_0^{\infty}\left[1 - F_Y(y)\right]dy - \int_{-\infty}^{0} F_Y(y)dy$$

Because we assumed *Y* to be non-negative, the mean is also given by the following simplified expressions

(17) $$\mu_Y = E[Y] = \int_0^{\infty} f_Y(y)ydy = \int_0^{\infty}\left[1 - F_Y(y)\right]dy$$

Note that the mean can also be obtained from the quantile function by

(18) $$\mu_Y = \int_0^1 Q(x)dx$$

Now, we introduce the random variable *X* by defining its probability density function.

Definition 1 – The random variable *X* and its probability density function, $f_X(x)$

Let *Y* be a non-negative continuous random variable having a positive finite mean, μ_Y, and the quantile function $Q(x)$. Then, *X* is defined as the random variable having the probability density function denoted by $f_X(x)$ and defined by

(19) $$f_X(x) \equiv \frac{Q(x)}{\int_0^1 Q(x)dx} = \frac{Q(x)}{\mu_Y} \quad \blacklozenge$$

Note that $f_X(x)$ fulfills the two necessary conditions for a probability density function (e.g., Mood et al., 1974), namely

(20) $$\begin{aligned} &i) \quad f_X(x) \geq 0, \forall x \in R \\ &ii) \quad \int_{-\infty}^{\infty} f_X(x)dx = 1 \end{aligned},$$

and therefore it is a legitimate probability density function. The first condition is fulfilled because *Y* is assumed to be a non-negative continuous random variable having a positive finite mean. Indeed, this assumption is necessary to define $f_X(x)$ from the quantile function. The second condition is fulfilled because $f_X(x)$ is obtained by dividing $Q(x)$ by its integral. Furthermore, $Q(x) = 0$ for $x \notin [0,1]$ implies $f_X(x) = 0$ for $x \notin [0,1]$.

The distribution of *X* is associated with the distribution of *Y* in such a way that the two random variables describe the same stochastic process from two different but related perspectives. For income distributions, *Y* is the income, assuming to extract it at random from its distribution, and *X* is the income recipient, assuming to assign at random, according to the distribution of *X*, a unit of income to the individuals in the population. Thus, whereas the *Y*-perspective focuses on the outcome of the data generating process, e.g., the income earned over a year, the X-perspective focuses on the stochastic process itself, i.e., the income distribution as it unfolds, *as if* it was observable.

Definition 2 – The cumulative distribution function of X, $F_X(x)$

The cumulative distribution function of X is denoted by $F_X(x)$ and defined by

$$(21) \qquad F_X(x) \equiv \int_{-\infty}^{x} f_X(t)dt \quad \blacklozenge$$

Because $f_X(x) = 0$ for $x \notin [0,1]$, $F_X(x) = 0$ for $x \leq 0$ and $F_X(x) = 1$ for $x \geq 1$.

For income distributions, $F_X(x) = \Pr(X \leq x)$ is the probability that a unit of income be assigned to an individual at a position in the range $[0, x]$ and it represents the share of total income by the xth fraction of the population with the lowest incomes. Thus, $F_X(x)$ also has the same interpretation of the Lorenz curve.

Proposition 1 – The Lorenz curve is the portion of the graph of $F_X(x)$ in the range $x \in [0,1]$. ♦

Proof. Applying the usual definition of the Lorenz curve in terms of the quantile function, i.e., as the inverse of the cumulative distribution function of Y (Gastwirth, 1971), Definition 1, and Definition 2, we obtain[14]

$$(22) \qquad L(x) = \frac{1}{\mu_Y}\int_0^x Q(t)dt = \int_0^x f_X(t)dt = F_X(x), \text{ for } 0 \leq x \leq 1 \quad \square$$

Now, using the standard definition of the expected value (e.g., Mood et al., 1974), we introduce the second key definition of this section.

Definition 3 – The expected value of X: The barycenter of the distribution of Y

The barycenter of the distribution of Y is defined as the expectation of X. It is denoted by μ_X and defined by

$$(23) \qquad \mu_X \equiv E[X] = \frac{1}{\mu_Y}\int_{-\infty}^{\infty} Q(x)xdx = \int_{-\infty}^{\infty} f_X(x)xdx = \int_0^{\infty}[1 - F_X(x)]dx - \int_{-\infty}^{0} F_X(x)dx \quad \blacklozenge$$

Because $Q(x) = 0$ for $x \notin [0,1]$ the barycenter is also given by the following simplified expressions

$$(24) \qquad \mu_X = \frac{1}{\mu_Y}\int_0^1 Q(x)xdx = \int_0^1 f_X(x)xdx = 1 - \int_0^1 F_X(x)dx$$

For income distributions, i.e., when Y is income, whereas the mean is the expected value of income, assuming to extract it at random from the distribution of Y, the barycenter is the expected income recipient, assuming to assign at random, according to the distribution of X, a unit of income to the individuals in the population sorted in the non-decreasing probability of receiving that income.

Proposition 2 – The barycenter of the distribution, μ_X, is the center of mass of the quantile function, $Q(x)$, and probability density function of X, $f_X(x)$. ♦

Proof. Implied by (15) and (24). □

In the figures, we represent this physical interpretation of the barycenter by indicating a thick line under the function of interest, i.e., $Q(x)$ or $f_X(x)$, as if this function was loading on a thin beam supported by a fulcrum. The triangle below the line represents the barycenter, i.e., the fulcrum in the point at which the beam balances (Figure 1 Panel 1-2, Figure 2-4 Panel 3-4).

[14] In a sense, we clarify and extend the intuition of Aaberge (2000, p. 641), which noted that "the Lorenz curve can be considered analogous to a cumulative distribution function".

Proposition 3 – The concentration area is equal to μ_X – 1/2. ♦

Proof. Denoting by L the area under the Lorenz curve, i.e., the area under the graph of $F_X(x)$ in the range $x \in [0,1]$, and by C the concentration area, i.e., the area limited by the Lorenz curve and the egalitarian line $f(x) = x$, and using (24), the barycenter is

(25) $\mu_X = 1 - L = 1/2 + C$,

whence

(26) $C = \mu_X - 1/2$. □

Using this relationship between the concentration area and the barycenter, we obtain a new geometrical representation of the concentration area, the extreme values of the barycenter, the relationship between the Gini index and the barycenter, and a new geometrical representation of the Gini index.

Proposition 4 – Geometrical representation of the concentration area.

The concentration area is equal to the area of the rectangle drawn in the plane of $F_X(x)$ with base μ_X – 1/2 and unit height (Figure 1 Panel 4). ♦

Proposition 5 – The barycenter of the distribution, μ_X, is in the range [1/2, 1].

(27) $x_{Median} = 1/2 \le \mu_X \le 1 = x_{Max}$. ♦

Proof. Using (25), whereas, in the case of perfect equality, the concentration area is null and μ_X = 0, in the case of perfect inequality, the concentration area is 1/2 and μ_X = 1. □

Proposition 6 – The Gini index is equal to 2μ_X – 1.

(28) $R = 2\mu_X - 1$. ♦

Proof. Implied by the relationship between the Gini index and the concentration area (8), i.e., R = 2C, and (26). □

Proposition 7 – Geometrical representation of the Gini index.

The Gini index is equal to the area of the rectangle drawn in the plane of $F_X(x)$ with base 2 (μ_X – 1/2) and unit height. ♦

After introducing the expected value of X, we now introduce its variance by applying the standard definitions (e.g., Mood et al., 1974).

Definition 4 – The variance of X

The variance of X is denoted by $Var[X]$ and defined by

(29) $Var[X] \equiv \int_{-\infty}^{\infty} (x - \mu_X)^2 f_X(x)dx = \int_{-\infty}^{\infty} f_X(x)x^2dx - \mu_X^2$ ♦

Because $Q(x)$ = 0 for $x \notin [0,1]$, the variance of X is also given by the following simplified expression

(30) $Var[X] = \int_0^1 f_X(x)x^2dx - \mu_X^2$

The variance of X can also be derived from the cumulative distribution function $F_X(x)$,

(31) $Var[X] = \int_0^{\infty} 2x[1 - F_X(x) + F_X(-x)]dx - \mu_X^2$,

however, this derivation is often much more complicated or even impossible analytically, as it is the use of the moment generating function, which is therefore not considered here.

We introduce now three new measures that are useful to characterize a distribution: the location of the mean, the barycentric value, and the barycentric share.

Definition 5 – The location of the mean, x_M

The location of the mean of the distribution of Y is denoted by x_M and defined as the value of the cumulative distribution function of Y, $F_Y(x)$, at the mean by

(32) $x_M \equiv F_Y(\mu_Y)$ ♦

Definition 6 – The barycentric value, y_B

The barycentric value of the distribution of Y is denoted by y_B and defined as the value of the quantile function, $Q(x)$, at the barycenter by

(33) $y_B \equiv Q(\mu_X)$ ♦

Definition 7 – The barycentric share, s_B

The barycentric share of the distribution of Y is denoted by s_B and defined as the value of the cumulative distribution of X, $F_X(x)$, at the barycenter by

(34) $s_B \equiv F_X(\mu_X)$ ♦

For income distributions, the location of the mean, x_M, represents the position on the x-axis of $Q(x)$, $f_X(x)$, and $F_X(x)$ of an individual in the population receiving the mean income and the barycentric value, y_B, represents the income of the expected income recipient. Note that in general, the barycentric income is different from the mean income (see Section 4). Finally, the barycentric share, s_B, represents the share of total income received by the μ_Xth fraction of the population with the lowest incomes, i.e., by individuals on the left of the barycenter.

These three measures, with the mean and the barycenter, identify four points of interest on the characteristic curves of a distribution: the point (μ_Y, x_M) on the graph of $F_Y(y)$, the points (x_M, μ_Y) and (μ_X, y_B) on the plot of $Q(x)$, and the point (μ_X, s_B) on the graph of $F_X(x)$.

Table 1 Column 2 summarizes the main formulae introduced in this section. Figure 2, Figure 3, and Figure 4 show the application of the methodology introduced in this section to the Uniform, Exponential, and Weibull distributions, respectively. Section 4 contains a broader application to 30 theoretical distributions.

Table 1, Figure 1, 2, 3, and 4.

2.2 *X*-SCALES FOR DISCRETE POPULATIONS: THE NORMALIZED *X*-SCALE

To apply to discrete populations the same approach introduced for continuous populations, it is necessary to solve a preliminary problem. Because this approach is based on the use of the quantile function, the problem is defining the position x_i to be assigned on the x-axis to an individual with rank i in the population so that the quantile function, i.e., $y_i = Q(x_i)$, can be specified.

Whereas this issue is usually overlooked in the literature on inequality measurement, i.e., a solution is implicitly or explicitly applied without discussion, it has been studied in statistics as the problem of defining the plotting position in quantile plots and probability papers, and many plotting positions have been proposed (e.g., Gumbel, 1954, pp. 13–15; Hyndman & Fan, 1996; Leon Harter, 1984). [15] The plotting position should be chosen according to the problem to be addressed, bearing in mind that it imposes a constraint on the definition of the quantile function (Hyndman & Fan, 1996).

We define an *x*-scale as a real-valued non-negative strictly increasing function that assigns to an individual with rank *i* in the population the position $x_i = x\text{-scale}(i)$ on the *x*-axis of the orthogonal plane $[x, f(x)]$.

We aim to study how the *x*-scales commonly used in the literature on income inequality affect the measures that we have introduced for continuous populations, in particular the barycenter, and to identify the *x*-scale that allows us to apply the same barycenter-based formulae to both continuous and discrete populations. Therefore, we consider only the *x*-scales that place individuals at constant distances and we define four *x*-axis scales that are used in the literature on income inequality and are interesting for our purposes. The first is the *natural x-scale*, $x_i = \{1,...,i,...,n\}$, used in the Pen Parade (Pen, 1971), in which the position of each individual corresponds with her rank. The second is the *Lorenz x-scale*, $x_i = \{1/n,...,i/n,...,1\}$, commonly used to draw the Lorenz curve. The third is the *normalized x-scale*, $x_i = \{0,...,(i-1)/(n-1),...1\}$, in which the positions are in the range $[0,1]$. Finally, the fourth *x*-scale is the *shifted natural x-scale*, in which the first individual is placed in the origin, $x_i = \{0,...,i-1,...,n-1\}$.[16]

Note that in the natural *x*-scale, whereas the median rank is (*n* + 1)/2 when *n* is even and it is *n*/2 when *n* is odd, the median location is $x_{Median} = (n + 1)/2$ for both even and odd *n*. Note also that in the Lorenz *x*-scale only, x_i denotes both the location of an individual with rank *i* and the fraction of the population up to this location. Furthermore, in the natural and Lorenz *x*-scales, because no individual is in the origin, the function of the individual position that one may wish to study, i.e., $y = f(x)$, must necessarily pass through the origin, i.e., it must be $f(0) = 0$. Except for the study of the Lorenz curve, which is commonly defined as including the point (0,0), this constraint can be a serious limitation.[17] The shifted natural *x*-scale solves this problem by adjusting the natural *x*-scale to include the origin. Finally, note that the normalized *x*-scale is the one implicitly used by Gini (1914, 2005) (see Section 2.4).

Table 2 shows the four *x*-scales defined, along with the median location and the constant distance between two adjacent individuals for each *x*-scale.

Table 2.

[15] For instance, Stigler (1974) considered the *x*-scale $x_i = \{1/(n+1),...,i/(n+1),...,n/(n+1)\}$, in which no individual is placed in the positions $x = 0$ and $x = 1$, a plotting position proposed by Weibull (1939) and Gumbel (1939).

[16] In statistics, the Lorenz *x*-scale is known as the "California method" (Leon Harter, 1984, p. 1616) and the normalized *x*-scale is known as the modal position (Gumbel, 1939).

[17] For instance, Milanovic (1997), studying a linear Pen Parade, was forced by the natural *x*-scale implicitly used to limit his analysis to the bundle of straight lines with a positive slope passing through the origin, i.e., he considered only the uniform income distributions having zero as lower bound.

The normalized *x*-scale has several properties that make it suitable to allow us to establish a correspondence between discrete and continuous populations.

Proposition 8 – Normalized *x*-scale for discrete populations and continuous populations

The normalized *x*-scale for discrete populations $x_i = \{0,...,(i-1)/(n-1),...1\}$ corresponds with the normalized scale of continuous populations, $x \in [0,1]$. ♦

Proof. In both cases, an individual with the smallest value of the variable of interest is in $x = 0$ and one with the maximum value is in $x = 1$. For instance, consider a Uniform distribution $U(a,b)$, whose quantile function is $Q(x) = (a+bx)I_{[0,1]}(x)$. Then $Q(0) = a$ and $Q1) = a+b$ for both discrete and continuous populations only when the normalized *x*-scale is used for discrete populations. □

Definition 8 – Normalized population

We define a population as normalized when it is continuous in the range $x \in [0,1]$ or it is discrete and the normalized *x*-scale $x_i = \left\{0,...,\left(i-1\right)/\left(n-1\right),...1\right\}$ is used. ♦

Proposition 9 – Median location and maximum location for normalized populations, x_{Median}, x_{Max}

For normalized populations, the median location, x_{Median}, is

(35) $x_{Median} = 1/2$,

and the maximum location, x_{Max}, i.e., the location of one maximum value of the variable of interest, is

(36) $x_{Max} = 1$ ♦

2.3 DISCRETE POPULATIONS

This section extends the definitions introduced for continuous populations to discrete populations.[18] We also define here the barycentric rank, which is a useful measure when discrete populations are considered.

In the case of discrete populations, *Y* is a non-negative random variable with *m* distinct values y_j, $j = 1,...,m$, each one with a relative frequency n_j / n. The probability mass function of *Y*, $f_Y(y_j)$, is

(37) $f_Y(y_j) = P[Y = y_j] = n_j / n,\ j = 1,...,m$,

and the cumulative distribution function of *Y*, $f_Y(y)$, is

(38) $F_Y(y) = \sum_{j:\, y_j \le y} f_Y(y_j)$

In the case of discrete populations, *X* is a discrete random variable with *n* distinct values $x_i = x\text{-scale}(i)$, $i = 1,...,n$. The quantile function, $Q(x_i)$, is

(39) $Q(x_i) = y_i$,

the probability mass function of *X*, $f_X(x_i)$, is

(40) $f_X(x_i) = P[X = x_i] = \dfrac{Q(x_i)}{\sum_{i=1}^{n} y_i} = \dfrac{y_i}{n\mu_Y}$,

and the cumulative distribution function of *X*, $F_X(x_i)$, is

[18] With the exception of the location of the mean.

(41) $F_X(x) = \sum_{i:\, x_i \le x} f_X(x_i)$

Note the important difference between (19) and (40). Whereas in the case of continuous populations, $f_X(x)$ is obtained by diving $Q(x)$ by μ_Y, in the case of discrete populations, $f_X(x_i)$ is obtained by diving $Q(x_i)$ by $n\mu_Y$. The two definitions are equivalent but whereas, in the case of continuous populations, we obtain a density distribution, in the case of discrete populations, we obtain a mass distribution.

Let $f_X(x_i)$ denote the mass associated with the mass point x_i. Then, the expected value of X, i.e., the barycenter of the distribution, x_B, is

(42) $$x_B = E[X] = \frac{\sum_{i=1}^{n} f_X(x_i)x_i}{\sum_{i=1}^{n} f_X(x_i)} = \frac{\sum_{i=1}^{n} y_i x_i}{\sum_{i=1}^{n} y_i}$$

For income distributions, $P[X = x_i]$ denotes the probability or relative frequency with which an individual with rank i and position x_i is selected for the allocation of one unit of the total income y_T. The barycenter is the position on the x-axis of the expected income recipient.

The barycenter is also the arithmetic mean of the y_T values assumed by X

(43) $$x_B = \frac{\sum_{i=1}^{n} f_X(x_i)x_i}{\sum_{i=1}^{n} f_X(x_i)} = \frac{1}{y_T}\sum_{j=1}^{y_T} x_j$$

Thus, for discrete populations, the barycenter's value depends on the x-scale.[19] However, the barycenter does not depend on the y-scale, as in the case of continuous populations.

Proposition 10 – The barycenter of the distribution, x_B, depends on the x-scale.

Proof. Implied by (42). □

Proposition 11 – The barycenter of the distribution, x_B, is y-scale independent.

Proof. Implied by (42). □

Note that an implication of Proposition 11, which holds for both continuous and discrete populations, is that distributions with different means can be compared graphically by using the plot of both f_X and F_X.

Table 2 shows the formulae for calculating the barycenter for each of the four x-scales introduced in Section 2.2.

Note that the normalized x-scale only allows us to establish a correspondence with the case of continuous populations. Indeed, the barycenter obtained for normalized discrete populations is the equivalent of the continuous populations' barycenter.

Definition 9 – Barycenter in the case of normalized discrete populations, μ_X

For normalized discrete populations, the barycenter is denoted by μ_X and defined by

(44) $$\mu_X \equiv \frac{\sum_{i=1}^{n}(i-1)y_i}{(n-1)\sum_{i=1}^{n} y_i}$$

Therefore, we can extend Proposition 5 to both continuous and normalized discrete populations.

[19] In the case of continuous populations, this is not a problem because the normalized scale is universally used.

Proposition 12 – For normalized populations, the barycenter, μ_X, is in the range [1/2, 1].

(45) $$x_{Median} = 1/2 \le \mu_X \le 1 = x_{Max} \quad \blacklozenge$$

Proof. Implied by Proposition 5 for continuous populations. For discrete populations and perfect equality, i.e., when $y_i = a$ for all i, a being a positive scalar, the barycenter is 1/2. In the case of perfect inequality, i.e., when $y_i = 0$ for $i = 1, ..., n–1$ and $y_n = a$, the barycenter is 1. □

Like the barycenter, also the variance of X depends on the x-scale used, and it is

(46) $$Var[X] = E[X^2] - E[X]^2 = \frac{\sum_{i=1}^{n} x_i^{\,2} y_i}{\sum_{i=1}^{n} y_i} - (x_B)^2$$

Table 2 shows the formulae for calculating the variance of X for each of the four x-scales introduced in Section 2.2. For a normalized discrete population, the variance of X, which we denote by σ_X^2 because it is the equivalent of the variance of X for continuous populations, is

(47) $$\sigma_X^2 = \frac{\sum_{i=1}^{n} (i-1)^2 y_i}{(n-1)^2 \sum_{i=1}^{n} y_i} - \mu_X^2 ,$$

Example. In the case of perfect equality, Y is a constant and X has a Uniform discrete distribution. When the natural x-scale is used, $E[X] = (n+1)/2$ and $Var[X] = (n+1)(n-1)/12$. When the normalized x-scale is used, $\mu_X = 1/2$ and $\sigma_X^2 = (n+1)/12(n-1)$.

The barycentric income, y_B, can be computed by the linear interpolation

(48) $$y_B = y_{i_L} + (i_B - i_L)\left(y_{i_R} - y_{i_L}\right),$$

where i_B is the barycentric rank, i.e., the rank corresponding to the barycenter, given by

(49) $$i_B = (n-1)\mu_X + 1 ,$$

i_L is the rank of an individual immediately to the left of the barycenter or at the barycenter, obtained as the integer part of i_B, $i_L = \lfloor i_B \rfloor$, and $i_R = i_L + 1$ is the rank of an individual immediately to the right of the barycenter.

Finally, the barycentric share can be computed by

(50) $$s_B = \sum_{i<i_B} y_i \Big/ \left(\sum_{i<i_B} y_i + \sum_{i>i_B} y_i\right),$$

where the individual at the barycenter, if any, is not considered so that the barycentric share is equal to 1/2 in the case of perfect equality, as in the continuous case.

Table 1 Column 2 summarizes the main formulae introduced in this section for the case of normalized discrete populations. Figure 5 illustrates these formulae with a numerical example.

Figure 5.

2.4 THE NORMALIZED LORENZ CURVE

In this section, we show that using the normalized *x*-scale for discrete populations, we obtain a normalized Lorenz curve (Figure 5 Panel 5) that is consistent with the Lorenz curve for continuous populations.

The egalitarian line, $e(x)$, connects the points $\{(x_i, p_i)\}$. When the normalized *x*-scale is used, $e(0) = 1/n$ and $e(1) = 1$, and the egalitarian line is represented by the function

(51) $$e(x) = \frac{1}{n} + \frac{n-1}{n}x$$

Note that the intercept of the egalitarian line with the *x*-axis is $x = -1/(n-1)$ and it moves towards the origin as n increases. The area under the egalitarian line to be considered is equal to the sum of the areas of the $(n-1)$ rectangles having base $1/(n-1)$ and height $p_i = 1/n$, which is equal to 1/2

(52) $$E = \frac{1}{n-1}\sum_{i=1}^{n-1} p_i = \frac{1}{n-1}\sum_{i=1}^{n-1}\frac{i}{n} = 1/2$$

The concentration area C, i.e., the area between the egalitarian line and the graph of $F_X(x)$ in the range [0,1], which we label as the normalized Lorenz curve, is equal to the sum of the areas of the $(n-1)$ rectangles having base $1/(n-1)$ and height $(p_i - q_i)$, and it is equal to $\mu_X - 1/2$,

(53) $$C = \frac{1}{n-1}\sum_{i=1}^{n-1}\left(p_i - q_i\right) = \mu_X - 1/2$$

Thus, by using the normalized *x*-scale, we obtain the normalized Lorenz curve such that also for discrete populations, as for continuous populations, i) the ratio between the concentration area and the area under the egalitarian area is equal to the concentration ratio R, as defined by Gini for the case of discrete populations, as shown by (3), and ii) R is equal to $2\mu_X - 1$,

(54) $$R = \frac{C}{E} = 2C = 2\mu_X - 1.$$

Note that instead, when the Lorenz *x*-scale is used, i.e., when the Lorenz curve is plotted, as usual, by considering the points $\{(0,0),(p_1,q_1),(p_2,q_2),...,(1,1)\}$, the area under the egalitarian line is equal to the sum of the areas of the $(n-1)$ rectangles with base $1/n$ and height $p_i = 1/n$, and therefore, it is $E_L = (n-1)/2n$, and the concentration area is $C_L = C\,(n-1)/n$. Thus, when the Lorenz *x*-scale is used, the correspondence between discrete and continuous populations is lost, because $R = C_L/E_L = 2C \neq 2C_L$. On the other hand, considering the area of the triangle under the egalitarian line, $E_T = 1/2$, would not only amount to comparing a continuous set of points with a discrete one, but it would also make $R \neq C_L/E_T = 2C_L$, so that the geometrical interpretation of the Gini index based on the Lorenz curve would also be lost. Note, however, that one would obtain the pseudo-Gini index, i.e., $G = 2C_L = R\,(n-1)/n$, discussed in Section 1.3.

Therefore, we conclude that the normalized *x*-scale is the only *x*-scale for discrete populations that is consistent with i) the definition of the concentration ratio R given by Gini, ii) the geometrical interpretation of R as the ratio between the concentration area and the area under the egalitarian line, $R = C/E$, and iii) the case of continuous populations, for which $R = 2C$.

3. THE BARYCENTER AND THE BALANCE OF INEQUALITY (=GINI) INDEX

In this section, we define an inequality index that is based on the notion of using the center of mass of $Q(x)$ or $f_X(x)$, i.e., the barycenter introduced in Section 2, to measure the concentration of distributions.[20] We label this approach to the measurement of inequality as the Balance of Inequality approach. First, we introduce this inequality index, which we label as the Balance of Inequality (*BOI*) index. Then we show that this index is equal to the Gini's concentration ratio *R*.[21]

Definition 10 – The Balance of Inequality (*BOI*) index

The Balance of Inequality index is denoted by *BOI* and defined as the normalization of the barycenter of the distribution, x_B, by

(55) $$BOI \equiv \frac{x_B - x_{Median}}{x_{Max} - x_{Median}} \quad \blacklozenge$$

The physical intuition behind the Balance of Inequality is simple. Let's consider an income distribution and assume that individuals in the population are sorted in non-decreasing order of income. In the case of perfect equality, i.e., when all individuals have the same income, the barycenter is in the median position. In the case of perfect inequality, i.e., when one individual receives all the income, the barycenter is in the position of this individual, at the right end. For any other distribution, the barycenter is between these two extremes and it indicates the income concentration. The further the barycenter is from the position of the median, the greater the income inequality. Thus, the Balance of Inequality "weighs" the distribution and the barycenter is a measure of inequality. By normalizing the barycenter, we obtain an inequality index that assumes values in the range [0,1]. The Balance of Inequality combines this physical intuition with a graphical representation of income inequality that shows at the same time the entire income distribution, the barycenter, i.e., the expected income recipient, and the *BOI* index on a scale between 0, i.e., the case of perfect equality, and 1, i.e., the case of perfect inequality (Figure 1 Panel 2, Figure 2-5 Panel 3).

Proposition 13 – The Balance of Inequality index is *x*-scale and *y*-scale invariant

Proof. Implied by Proposition 11 and Definition 10. □

Proposition 14 – For normalized populations, the Balance of Inequality index is $2\mu_X - 1$

(56) $$BOI = 2\mu_X - 1 \quad \blacklozenge$$

Proof. Implied by Proposition 12. □

---

[20] The more concentrated is a distribution and the more it is unequal. As noted by Lorenz (1905, p. 215), the aim is to study whether income and wealth are concentrated or diffused among the population. As noted by Gini, the same method is "applicable not only to incomes and wealth, but to all other quantitative characteristics […] to obtain a rough estimate of the various degrees of inequality which the distribution of these characteristics presents" (Gini, 1921, p. 124).

[21] The idea of using the center of mass of the distribution to measure its concentration is not entirely new. We acknowledge that the same idea was proposed by Fernando Giaccardi (1950a, 1950b). Studying discrete populations, he showed that not only the Gini index but also the Bonferroni and De Vergottini indices, among others, are obtainable from a more general expression based on this "mechanical" approach, which he labelled as the concentration index *K*. The Gini index is obtained when individuals are placed at constant distances and the other indices are obtained by varying the positions in which individuals are considered. Giaccardi contribution has been neglected by the literature on the measurement of inequality and only by chance we discovered it on October 31, 2021.

Proposition 15 – The Balance of Inequality index is equal to the Gini index

(57) $BOI = R$ ♦

Proof for continuous populations. Implied by Proposition 6 and Proposition 14.

Proof for discrete populations. Using (5), (44), and (56),[22] we obtain that the Balance of Inequality index is the Gini's concentration ratio R

(58) $$BOI = 2\mu_X - 1 = \frac{2\sum_{i=1}^{n}(i-1)\,y_i}{(n-1)\sum_{i=1}^{n} y_i} - 1 = R \quad \square$$

Because the Balance of Inequality index is based on a different approach than the Gini index, i.e., it is based on the use of the barycenter of the distribution, and it is accompanied by new economic, geometrical, physical, and statistical interpretations, in the following we use the denomination Balance of Inequality (=Gini) index and the symbol *BOI*.

Note that Definition 10 applies for both continuous and discrete populations, and to any x-scale. Table 2 shows the Balance of Inequality index formulae for each of the four x-scales introduced in Section 2.2. Note that, for Proposition 13, all the formulae give the same result, which can also be expressed as

(59) $$BOI = \frac{2}{(n-1)} \frac{\sum_{i=1}^{n} i y_i}{\sum_{i=1}^{n} y_i} - \frac{n+1}{n-1},$$

which is a well-known expression for the Gini index (Jasso, 1979).

4. EXAMPLES: THEORETICAL DISTRIBUTIONS

In this section, we show the results of the application of the methodology introduced in the previous sections to thirty theoretical distributions (Table 3). We selected these distributions among those usually applied in the study of income inequality and those that could show the wider utility of the proposed methodology, choosing the distributions that allowed the derivation in closed form of $Q(x)$ and μ_X. For each distribution, starting from $f_Y(y)$ or $F_Y(y)$, we derived i) $Q(x)$, μ_Y, $f_X(x)$, and $F_X(x)$; ii) the barycenter, μ_X, the Balance of inequality (=Gini) index, and its range; iii) x_M, y_B, and s_B.

For each distribution, Table 4 shows the parameters space, $F_Y(y)$ and $f_X(x)$[23]; Table 5 shows the barycenter, μ_X, the *BOI* index, and its range; and Table 6 shows μ_X, x_M, and the sign or value of their difference.[24] Only closed-form expressions are reported, and excessively long expressions are omitted. Overall, however, the omissions are very few.

Table 3, 4, 5, and 6.

[22] The same result can be shown using (4), (55), and the expression of the barycenter for the natural x-scale (Table 2).

[23] For each distribution, Table A2.1 in Appendix 2 shows the five characteristic functions, i.e., $f_Y(y)$, $F_Y(y)$, $Q(x)$, $f_X(x)$, and $F_X(x)$.

[24] For each distribution, Table A2.2 in Appendix 2 shows the median, μ_Y, y_B, x_M, μ_X and the sign or value of $\mu_X - x_M$.

We double-checked the expressions obtained analytically by a numerical example and a graphical representation.[25] Figure 2, Figure 3, and Figure 4 are examples of these graphical representations, for the Uniform, Exponential, and Weibull distributions, respectively.[26]

This application to thirty theoretical distributions shows that whereas in some cases the location of the mean only is fixed (e.g., Gumbel, Uniform, and U-Quadratic distribution), in others also the barycenter is fixed (e.g., Exponential, Half Logistic, Half Normal, and Triangular distribution) and in these cases the difference $\mu_X - x_M$ is positive and constant, i.e., the point (μ_X, y_B) is on the right of the point (x_M, μ_Y) on the curve $Q(x)$ at a constant distance irrespective of the values of the parameters of the distribution. In other cases, the difference $\mu_X - x_M$ is negative and variable, and in still others, it can be positive or negative based on the value of the parameters. Therefore, we can advance the following two propositions.

Proposition 16 – There is no general relation between the barycenter and the mean, i.e., when there is a relation between them it is distribution-specific. ♦

Proof. The Exponential distribution, $F_Y(y;\lambda>0)=(1-e^{-\lambda y})I_{[0,\infty)}(y)$, is a counterexample showing the absence of a general relation between the mean and the barycenter. Whereas the mean, $\mu_Y = 1/\lambda$, depends on the distribution parameter λ, the barycenter is fixed, μ_Y = ¾ (Figure 3). □

Proposition 17 – The barycenter is a new measure of the location or central tendency of a distribution.

Because there is no general relation between the barycenter and the mean, the barycenter is a new measure of central tendency or location, which adds to the mean, median, and mode.[27] . ♦

We conclude by noting that these examples also show that the expressions of $Q(x)$ and $f_X(x)$ are much simpler than that of $F_X(x)$ (Table A2.1), and therefore the derivation of the Balance of inequality (=Gini) index from the normalization of the barycenter is much simpler than its derivation from the Lorenz curve. Note also that the ranges of the *BOI* index (Table 5 Column 4) may be a guide for the selection of the appropriate theoretical distribution to be applied to a data set. For instance, a *BOI* index equal to 0.3 is consistent with a Pareto I distribution but not with a Pareto II distribution.

5. THE BARYCENTER AND THE GINI'S MEAN DIFFERENCE

Using the barycenter, we can express the Gini's mean difference in terms of the mean and the barycenter of the distribution, i.e., as a function of just these two measures of central tendency, and we can give a new geometrical representation for the Gini's mean difference.

Proposition 18 – The Gini's mean difference is equal to four times the product of the mean by the distance of the barycenter from the median location.

[25] We used Wolfram Mathematica v. 11.2 and an ad-hoc *R* code (Di Maio, 2022).

[26] Appendix 3 shows similar graphical representations for almost all theoretical distribution listed in Table 3.

[27] Winkler (2009, p. 153) observed that "In the early days of statistics the number of measures of location or central tendency proliferated. Of those only the arithmetic mean survived, with the median and mode taking distant second and third places." However, as far as we have been able to ascertain, the barycenter of the distribution is a new measure of central tendency.

(60) $$\Delta = 4\mu_Y \left(\mu_X - \frac{1}{2} \right) \quad ♦$$

Proof. Implied by (6) and Proposition 6. □

Proposition 19 – Geometrical representation of the Gini's mean difference.

The Gini's mean difference is equal to the area of the rectangle built on the quantile function's plot having as base twice the distance of the barycenter from the median location and as height twice the mean of the distribution (Figure 1 Panel 3). ♦

An application of (60) is that the limiting value of the correlation between variates-values and ranks in samples of *n* as the sample size *n* tends to infinity, introduced by Stuart (1954), can be expressed in terms of the mean, the barycenter, and the standard deviation, σ_Y, of the distribution by

(61) $$C = \frac{\sqrt{3}}{2}\frac{\Delta}{\sigma_Y} = 2\sqrt{3}\frac{\mu_Y}{\sigma_Y}\left(\mu_X - \frac{1}{2} \right)$$

This application suggests that the Gini's mean difference is not a measure of dispersion, like the standard deviation, but a measure of central tendency that summarizes the mean and the barycenter and that the barycenter of the distribution may have wide applications in statistics.

6. BARYCENTER-BASED NEW INSIGHTS FOR THE GINI INDEX

Since the Balance of Inequality index is equal to the concentration index *R*, i.e., the Gini index, it has also the same well-known properties. However, using the barycenter of the distribution, we can propose new interpretations for the Gini index and clarify or motivate in a new way some of its properties. First of all, we summarize the results obtained in the previous sections by the following proposition.

Proposition 20 – The Balance of Inequality (=Gini) index is the normalization of the barycenter of the distribution.

Proof. Implied by Definition 10 and Proposition 15. □

The barycenter and the Gini index of a distribution convey the same information and the geometrical, economic, physical, and statistical interpretations of the barycenter presented in the previous sections can be applied also to the Gini index.

Let now consider an income distribution in a discrete population, and let ε be the amount of a positive income transfer from an individual with rank *i* to one with rank *j*. Let's define this transfer as progressive (regressive) when it occurs from an individual with a greater (smaller) income to one with a smaller (greater) income, and as rank-preserving when it leaves unchanged the ranks of the two individuals. The principle of progressive transfers (e.g., Ebert, 1988) has the following barycenter-based physical interpretation.

Proposition 21 – Rank-preserving income transfers and the Balance of Inequality (=Gini) index

A rank-preserving progressive (regressive) transfer shifts the barycenter to the left (right) and reduces (increases) the *BOI* index. The movement of the barycenter and the variation in inequality depend exclusively on the product of the amount of the transfer times the distance between the two individuals. ♦

Proof. Using (44) and (56), the movement of the barycenter due to the income transfer is

$$(62) \qquad \partial\mu_X = \frac{(i-j)}{(n-1)}\frac{\varepsilon}{\sum_{i=1}^{n} y_i}$$

and the *BOI* index variation is $\partial BOI = 2\,\partial\mu_X$. □

This property accords with common sense. Indeed, it means that a given transfer to a poor recipient decreases inequality the more the richer the donor is and that, for a given transfer, the maximum decrease in inequality occurs when the recipient is the poorest person and the donor is the richest person in the population.

Proposition 22 – The Balance of Inequality (=Gini) index attaches the same weight to all rank-preserving transfers of the same amount that occur between two individuals separated by the same distance, regardless of their incomes.

Proof. Implied by (62). □

A greater progressive transfer between two high–income individuals, in the right tail of the distribution, may decrease the Gini index more than a smaller one between two individuals belonging to the middle-income class. Thus, as also shown by Aaberge (2000), it is in general not true that "the Gini coefficient attaches more weight to transfers affecting middle income classes" (Atkinson, 1970, pp. 256–257) or that it is "more sensitive to transfers at the center of the distribution than at the tails" (Alvaredo, 2011, p. 274).

Proposition 23 – The barycenter and the Balance of Inequality (=Gini) index attach the same weight to all incomes. ♦

Proof. Using (49), (44) can be written as

$$(63) \qquad \sum_{i<i_B}\left[\mu_X - \frac{(i-1)}{(n-1)}\right]y_i = \sum_{i>i_B}\left[\frac{(i-1)}{(n-1)} - \mu_X\right]y_i$$

From a physical point of view, this equation conveys the second condition for the static equilibrium of the system composed of the masses y_i placed in the normalized positions $x_i = \{0,\ldots,(i-1)/(n-1),\ldots 1\}$ on a beam supported by a fulcrum placed in the center of mass, i.e., the summation of all moments acting on the system is zero. From an economic point of view, it implies that, because the distance between adjacent individuals is constant, the barycenter and the *BOI* index attach the same weight to all incomes. □

Proposition 24 – Rank-preserving additions to incomes and the Balance of Inequality (=Gini) index

When the receiver position is on the right (left) of the barycenter, a rank-preserving positive addition to the income of one individual moves the barycenter to the right (left), and increases (decreases) inequality, as measured by the *BOI* index. ♦

Proof. Implied by (63). □

Note that this proposition implies that it does not matter whether the recipient's income is greater or smaller than the mean income of the population.

Proposition 25 – The Balance of Inequality (=Gini) index depends on the barycenter of the income distribution and does not depend on the mean income.

Proof. Implied by Proposition 16. □

For instance, an equal proportionate increase (or decrease) of all incomes does not affect the Gini index because the barycenter is unaffected, not because the Gini index is defined relative to the mean, as claimed by some authors (e.g., Atkinson, 1970, p. 253). On the contrary, an equal addition to all incomes decreases the Gini index because the barycenter decreases, moving to the left toward the median location, not because the mean income increases.

The barycenter of the income distribution divides the population into two groups, which can be considered as both "the winners" and "the losers" in the income distribution and "the rich" and "the poor".

Proposition 26 – The higher the Balance of Inequality (=Gini) index, the more are "the losers".

"The winners" ("the losers") in the income distribution are the income recipients on the right (left) of the barycenter: they have a greater (smaller) probability of receiving one unit of income than the expected income recipient.

Proposition 27 – The higher the Balance of Inequality (=Gini) index, the more are "the poor"

"The rich" ("the poor") are the income recipients on the right (left) of the barycenter that have an income greater (smaller) than the barycentric income: a small positive addition to their income increases (decreases) income inequality.

7. THE BARYCENTER AND INFERENCE: POINT ESTIMATES

In this section, we provide two estimators of the population's barycenter and *BOI* index, one for a random sample from the population and one for weighted observations.

7.1 RANDOM SAMPLE

Let $y_1 \le y_2 \le ... y_i \le ... \le y_n$ be the order statistics, arranged in non-decreasing order of magnitude, of a sequence of *n* independent and identically distributed non-negative random variables with the same probability density function $f_Y(y)$ and such that their sum is positive. The rank order statistics are the set of positive integers from 1 to the sample size *n*. Using (40), the empirical probability mass function of *X* is

(64) $$\hat{f}_X(x_i) = \frac{y_i}{\sum_{i=1}^{n} y_i}$$

where $x_i = x\text{-scale}(i)$. Applying the normalized *x*-scale, i.e., using (44), the plugin estimator of the population's barycenter is

(65) $$\hat{x}_{B.N} = \frac{\sum_{i=1}^{n}(i-1)\,y_i}{(n-1)\sum_{i=1}^{n} y_i} ,$$

and the estimation of the *BOI* index is obtained by

(66) $$B\hat{O}I = 2\hat{x}_{B.N} - 1$$

Note that the difference with the commonly used estimator of the Gini index based on (59) (e.g., Langel & Tillé, 2013 eq. 2) is that we do not estimate directly the Gini index but we obtain the estimation of the Gini index from the estimation of the barycenter. In the next section, this method allows us to use the additive property of the center of mass to obtain a new estimator of the Gini index for weighted observations that is not based on the Lorenz curve.

7.2 WEIGHTED OBSERVATIONS

The Gini index is generally estimated through a random sample obtained by survey sampling, in which each observation has an associated sampling weight that is an estimate of how many individuals in the population of interest each observation represents. Sampling weights are needed to make the random sample representative of the overall population and are used to correct for sampling bias, unit or item non-response bias, and data collection bias (e.g., Langel & Tillé, 2013).

Let $y_1,...,y_k,...,y_m$ denote m observations, i.e., incomes, in a random sample, sorted in non-decreasing order, and $w_1,...,w_k,...,w_m$ the associated sampling weights. The quantile function of this random sample is a discrete step function, with one step for each weighted observation. Assuming that sampling weights are integers, i.e., each observation represents an integer number of people, and using the normalized x-scale, we can estimate the center of mass of each step, $\hat{x}_{B.N,k}$, by

(67) $$\hat{x}_{B.N,k} = \frac{1}{n-1}\left\{\left[\left(\sum_{j=1}^{k-1} w_j\right) + \frac{w_k+1}{2}\right] - 1\right\},$$

and the share of the total income of each weighted observation, $\hat{s}_k$, by

(68) $$\hat{s}_k = \frac{y_k w_k}{\sum_{k=1}^{m} y_k w_k}$$

Thus, we obtain a plugin estimator of the barycenter for weighted observations, by

(69) $$\hat{x}_{B.N} = \sum_{k=1}^{m} \hat{s}_k \hat{x}_{B.N,k},$$

and the estimate of the Balance of Inequality (= Gini) index for weighted observations, by

(70) $$B\hat{O}I = 2\hat{x}_{B.N} - 1,$$

Note that you would obtain the same estimations by duplicating each weighted observation y_k in the random sample ($w_k - 1$) times, and then applying the plugin estimators (65) and (66). However, this procedure may not be feasible for very large datasets with the available computing capacity.

Finally, note that for non-integers sampling weights, which imply the assumption of continuous populations,[28] the estimator of the center of mass of each step is

(71) $$\hat{x}_{B.C,k} = \left(\sum_{j=1}^{m} w_j\right)^{-1}\left[\left(\sum_{j=1}^{k-1} w_j\right) + \frac{w_k}{2}\right]$$

Thus, the plugin estimator of the barycenter for weighted observations is

(72) $$\hat{x}_{B.C} = \sum_{k=1}^{m} \hat{s}_k \hat{x}_{B.C,k},$$

[28] Even if actual populations can often be considered virtually infinite (Monti, 1991), in many applications the population considered is relatively small and in these cases (67) is preferable.

and the estimate of the Balance of Inequality (= Gini) index for weighted observations is

(73) $$B\hat{O}I = 2\hat{x}_{B.C} - 1$$

8. EMPIRICAL APPLICATION: PERSONAL INCOME INEQUALITY IN LIS COUNTRIES

In this section, we apply the plugin estimators of the barycenter (69) and *BOI* index (70) to estimate personal income inequality in some countries. We use the harmonized data sets in the Luxembourg Income Study (LIS) Database provided by the LIS Cross-National Data Center (2021), and we consider the year 2016 (Wave X), or the previous year closest to 2016 when data for 2016 are not available. We also add a handful of historical data to give some depth to the analysis.

We consider the total personal income variable (*pitotal*) that is "the sum of cash and non-cash income from labor (including wage income, self-employment income, and fringe benefits, but excluding own consumption), income from pensions (including both public and private pensions) and non-pension public social benefits whose eligibility is based on individual rather than household characteristics (namely wage replacement benefits, such as maternity and parental leave benefits, unemployment benefits, sickness and work injury benefits, and disability benefits), as well as private scholarships" (LIS Cross-National Data Center, 2019). We include in the analysis all the countries for which personal data are available.

We keep all observations, i.e., we do not bottom and top code data to avoid introducing a bias in the estimation. However, we make negative and missing incomes zero, if any. We also round to the nearest integer the personal weight (*ppopwgt*) provided by LIS, which is the population individual cross-sectional weight variable that inflates the result to reflect the total individual population covered by the dataset (LIS Cross-National Data Center, 2019).

Table A4.1 in Appendix 4 reports the details of this analysis, showing the great heterogeneity between the different countries as regards the use of the personal weights and the number of observations per inhabitant. Whereas for some countries, observations may represent even less than a dozen people, for other countries, they represent no less than a few thousand people. The number of negative and missing values is low for all countries and zero for many. The rounding of personal weights does not affect the estimated population sizes and results significantly.[29]

Figure 6 shows the two main results: i) the estimated barycenter of the income distribution, i.e., the expected income recipient, and ii) the estimated Balance of Inequality (= Gini) index for each country. These results confirm some stylized facts. Income inequality is lower in European countries, and in particular in the countries of Central and Northern Europe. For these countries, the expected income recipient is between the 74th (Hungary) and 83rd (Ireland) percentile. The United States of America, China, Japan, and Taiwan are just above this level. Income inequality is higher in Central and South America, e.g., in Mexico, the expected income recipient is at the 89th percentile, and it is very high in the low-income countries of Africa and Asia

[29] The difference between the results obtained by applying (69) and (70) to rounded personal weights and those obtained by applying (72) and (73) to non-rounded personal weights is negligible.

for which data are available. The maximum value of income inequality is found in India, where the expected income recipient is at the 94.5th percentile, and therefore the *BOI* index is around 0.89.

Figure 6.

9. DISCUSSION AND CONCLUSIONS

Whenever the object of study is the distribution of a quantitative characteristic of the individuals in a population, e.g., income, represented by the non-negative random variable Y with a positive finite mean, μ_Y, and the quantile function $Q(x)$, this distribution can also be studied by introducing a second random variable X having the probability density function $f_X(x) = Q(x)/\mu_Y$.

The barycenter of the distribution, i.e., the expected value of this second random variable, $\mu_X = E[X]$, is a new measure of the location or central tendency of distributions, which adds to the mean, median, and mode.

The barycenter of the distribution is particularly important for the measurement of inequality. Indeed, the same barycenter-based formulae can be applied to both continuous and discrete populations to derive some of the most used inequality measures. Indeed (see also Table 1 Panel C), using the barycenter obtained by applying (24) for continuous populations and (44) for discrete populations, you derive:

- the Balance of inequality (=Gini) index, $BOI = 2\mu_X - 1$,
- the concentration area between the egalitarian line and the Lorenz curve, $C = \mu_X - 1/2$,
- the area under the Lorenz curve, $L = 1 - \mu_X$,
- the Gini's mean difference, $\Delta = 4\mu_Y (\mu_X - 1/2)$.

The introduction of the barycenter allows for new economic, geometrical, physical, and statistical interpretations and graphical representations of these measures (Figure 1-5).

In particular, this study shows that the Gini index, i.e., the concentration ratio R (Gini, 1914, 2005), is the normalization of the barycenter, and therefore, it provides a new statistical foundation for this inequality index, which is the most used but also the most criticized. Indeed, considering continuous populations, many authors define the Gini index from the Lorenz curve. However, the association between the Lorenz curve and the Gini index is problematic for discrete populations, and the normalized Lorenz curve introduced in this paper is a solution to this problem. Furthermore, a direct definition of the Gini index as “the ratio to the mean of half the average over all pairs of the absolute deviations between people” (Deaton, 1997, p. 139) was also deemed unsatisfactory by many, even if it corresponds with one of the ways you can write the concentration ratio R. Finally, the relation between the mean and the Gini index was not at all clear, and this study makes it clear that there is no relation because the Gini index depends on the barycenter and does not depend on the mean of the distribution.

We propose that, for income distributions, the barycenter represents the expected recipient of one unit of income, as if the stochastic process that leads to the distribution of the total income among the population was observable as it unfolds.

The Balance of Inequality (Figure 1 Panel 2, Figure 2-5 Panel 3) provides a physical interpretation and geometrical representation of the Gini index. This graphical representation can be easily understood by the general public because it provides a way of illustrating inequality similar to the commonly used one based on a double-pan balance. However, whereas the double-pan balance can only show that a few richer ones have a wealth that weighs a lot more than the wealth of many poorer ones, the Balance of Inequality shows the entire distribution of wealth, the position of its center of mass, i.e., the barycenter, and the value of the inequality index in the same graphical representation. The interpretation is straightforward: the more to the right the barycenter, the greater the inequality measured by the Balance of Inequality (=Gini) index.

The application to thirty theoretical distributions shows that the Balance of Inequality (=Gini) index can be obtained much more simply from the quantile function than from the Lorenz curve. This indication could be useful to approach the problem of ranking intersecting Lorenz curves in a new way.

The empirical application, with the estimation of personal income inequality in Luxembourg Income Study Database's countries, illustrates the interpretation of the Balance of Inequality (=Gini) index as the normalization of the position of the expected income recipient. The barycenter also splits the population into two groups, which can be considered as "the winners" and "the losers" in the income distribution, or "the rich" and "the poor".

The application of the barycenter-based expression of the Gini's mean difference to the limiting value of the correlation between variates-values and ranks in samples (Stuart, 1954) suggests that the barycenter, as a new measure of the location or central tendency of the distribution, may have wide applications in both economics and statistics.

10. REFERENCES

Aaberge, R. (2000). Characterizations of Lorenz curves and income distributions. *Social Choice and Welfare*, *17*(4), 639–653. https://doi.org/10.1007/s003550000046

Alvaredo, F. (2011). A note on the relationship between top income shares and the Gini coefficient. *Economics Letters*, *110*(3), 274–277.

Archimedes, Eutocios, & Heiberg, J. L. (1881). Ἐπιπέδων ἰσορροπιῶν ἢ κέντρα βαρῶν ἐπιπέδων α' - De planorum equilibris siue de centris grauitatis planorum I. In *Archimedis opera omnia cum commentariis Eutocii. E codice Florentinum recensuit, latine uertit notisque illustrauit Volumen II* (pp. 142–187). Lipsia: B. G. Teubneri. Retrieved from https://archive.org/details/archimedisopera00heibgoog

Archimedes, & Heath, T. L. (1897a). On the equilibrium of planes or the centres of gravity of planes. Book I. In *The works of Archimedes edited in modern notations with introductory chapters* (pp. 189–202). Cambridge: University Press. Retrieved from https://archive.org/details/worksofarchimede00arch

Archimedes, & Heath, T. L. (1897b). *The works of Archimedes edited in modern notations with introductory chapters*. Cambridge: University Press. Retrieved from https://archive.org/details/worksofarchimede00arch

Atkinson, A. B. (1970). On the measurement of inequality. *Journal of Economic Theory*, *2*(3), 244–263. https://doi.org/10.1016/0022-0531(70)90039-6

Beirlant, J., Vynckier, P., & Teugels, J. L. (1996). Excess Functions and Estimation of the Extreme-Value Index. *Bernoulli*, *2*(4), 293. https://doi.org/10.2307/3318416

Bulmer, M. G. (1979). *Principles of statistics*. New York, NY: Dover Publications.

Burr, I. W. (1942). Cumulative Frequency Functions. *The Annals of Mathematical Statistics*, *13*(2), 215–232. https://doi.org/10.1214/aoms/1177731607

Champernowne, D. G. (1953). A Model of Income Distribution. *The Economic Journal*, *63*(250), 318. https://doi.org/10.2307/2227127

Dagum, C. (1990). Generation And Properties Of Income Distribution Functions. In C. Dagum & M. Zenga (Eds.), *Income and Wealth Distribution, Inequality and Poverty* (pp. 1–17). https://doi.org/10.1007/978-3-642-84250-4_1

Dasgupta, P., Sen, A., & Starrett, D. (1973). Notes on the measurement of inequality. *Journal of Economic Theory*, *6*(2), 180–187. https://doi.org/10.1016/0022-0531(73)90033-1

Deaton, A. (1997). *The Analyisis of Household Surveys: A Microeconometric Approach to Development Policy*. Baltimore, MD: The Johns Hopkins University Press.

Devroye, L. (1986). *Non-Uniform Random Variate Generation*. New York, NY: Springer Science+Business Media New York. https://doi.org/10.1007/978-1-4613-8643-8

Di Maio, G. (2022). Package “BOI”: The Balance of Inequality. *Mimeo*.

Di Maio, G., & Landoni, P. (2015). Beyond the Gini Index: Measuring Inequality with the Balance of Inequality Index. *A/Simmetrie Working Papers Series (ISSN 2421-7131)*, *06*. Retrieved from https://ideas.repec.org/p/ais/wpaper/1506.html

Di Maio, G., & Landoni, P. (2017). The Balance of Inequality: A rediscovery of the Gini’s R concentration ratio and a new inequality decomposition by population subgroups based on a physical rationale. *Seventh Meeting of the Society for the Study of Economic Inequality (ECINEQ), Graduate Center of the City University of New York, New York City, July 17-19, 2017*. Retrieved from http://www.ecineq.org/ecineq_nyc17/FILESx2017/CR2/p195.pdf

Ebert, U. (1988). Measurement of inequality: An attempt at unification and generalization. *Social Choice and Welfare*, *5*(2–3), 147–169. https://doi.org/10.1007/BF00735758

Finzi, B. (1991). *Meccanica razionale. Volume primo. Teorie introduttive. Cinematica - Statica*. Bologna: Zanichelli.

Fowles, G. R., & Cassiday, G. L. (2005). *Analytical Mechanics: Seventh Edition*. Belmont, CA: Thomson Brooks/Cole.

Gastwirth, J. L. (1971). A General Definition of the Lorenz Curve. *Econometrica*, *39*(6), 1037–1039. https://doi.org/10.2307/1909675

Giaccardi, F. (1950a). Indici di concentrazione. *Giornale Degli Economisti e Annali Di Economia*, *9*(5/6), 282–290.

Giaccardi, F. (1950b). Un criterio per la costruzione di indici di concentrazione. *Rivista Italiana Di Economia Demografia e Statistica*, *4*(3/4), 527–538.

Gini, C. (1912). *Variabilità e mutabilità: contributo allo studio delle distribuzioni e delle relazioni statistiche. Fascicolo 1° Introduzione - Indici di variabilità - Indici di mutabilità*. Bologna: Tipografia di Paolo Cuppini. Retrieved from http://www.digibess.it/fedora/repository/object_download/openbess:TO043-00408/PDF/openbess_TO043-00408.pdf

Gini, C. (1914). Sulla misura della concentrazione e della variabilità dei caratteri. *Atti Del Reale Istituto Veneto Di Scienze, Lettere Ed Arti*, *LXXVIII*(Parte Seconda), 1203–1248.

Gini, C. (1921). Measurement of Inequality of Incomes. *The Economic Journal*, *31*(121), 124–126.

Gini, C. (2005). On the measurement of concentration and variability of characters. *METRON - International Journal of Statistics*, *63*(1), 3–38.

Giorgi, G. M. (2005). Reasons for a translation. *METRON - International Journal of Statistics*, *63*(1), 1–2.

Giorgi, G. M. (2014). A couple of good reasons to translate papers of the Italian statistical tradition. *METRON*, *72*(1), 1–3. https://doi.org/10.1007/s40300-013-0032-x

Giorgi, G. M., & Nadarajah, S. (2010). Bonferroni and Gini indices for various parametric families of distributions. *METRON*, *68*(1), 23–46. https://doi.org/10.1007/BF03263522

Gumbel, E. J. (1939). La Probabilité des Hypothéses. *Comptes Rendus de l'Académie Des Sciences (Paris)*, *209*, 645–647.

Gumbel, E. J. (1954). *Statistical Theory of Extreme Values and Some Practical Applications*. Washington, DC: U.S. National Bureau of Standards.

Hankin, R. K. S., & Lee, A. (2006). A new family of non-negative distributions. *Australian and New Zealand Journal of Statistics*, *48*(1), 67–78. https://doi.org/10.1111/j.1467-842X.2006.00426.x

Hyndman, R. J., & Fan, Y. (1996). Sample Quantiles in Statistical Packages. *The American Statistician*, *50*(4), 361–365. https://doi.org/10.1080/00031305.1996.10473566

Jasso, G. (1979). On Gini's Mean Difference and Gini's Index of Concentration. *American Sociological Review*, *44*(5), 867. https://doi.org/10.2307/2094535

Kendall, M. G. (1945). *The advance theory of statistics. Volume I* (2nd ed.). Charles Griffin and company.

Kleiber, C., & Kotz, S. (2003). *Statistical Size Distributions in Economics and Actuarial Sciences*. Hoboken, NJ: John Wiley & Sons. https://doi.org/10.1002/0471457175

Kumaraswamy, P. (1980). A generalized probability density function for double-bounded random processes. *Journal of Hydrology*, *46*(1–2), 79–88. https://doi.org/10.1016/0022-1694(80)90036-0

Langel, M., & Tillé, Y. (2013). Variance estimation of the Gini index: Revisiting a result several times published. *Journal of the Royal Statistical Society. Series A: Statistics in Society*, *176*(2), 521–540. https://doi.org/10.1111/j.1467-985X.2012.01048.x

Leon Harter, H. (1984). Another look at plotting positions. *Communications in Statistics - Theory and Methods*, *13*(13), 1613–1633. https://doi.org/10.1080/03610928408828781

LIS Cross-National Data Center. (2019). *Generic Codebook for LIS Database*. Luxembourg: LIS Cross-National Data Center.

LIS Cross-National Data Center. (2021). *Luxembourg Income Study (LIS) Database, http://www.lisdatacenter.org (multiple countries; 2-10/11/2021)*. Luxembourg: LIS.

Lorenz, M. O. (1905). Methods of measuring the concentration of wealth. *Publications of the American Statistical Association*, *9*(70), 209–219. https://doi.org/10.1080/15225437.1905.10503443

MacMahon, D. (2007). *Statics and dynamics demistified: A self-teaching guide*. New York, NY: The McGraw-Hill Companies.

Mehran, F. (1976). Linear Measures of Income Inequality. *Econometrica*, *44*(4), 805. https://doi.org/10.2307/1913446

Milanovic, B. (1997). A simple way to calculate the Gini coefficient, and some implications. *Economics Letters*, *56*(1), 45–49. https://doi.org/10.1016/S0165-1765(97)00101-8

Montgomery, D. C., & Runger, G. C. (2014). *Applied Statistics and Probability for Engineers* (6th ed.). Hoboken, NJ: John Wiley & Sons.

Monti, A. C. (1991). The study of the Gini concentration ratio by means of the influence function. *Statistica*, *51*(4), 561–580. Retrieved from https://rivista-statistica.unibo.it/article/view/885

Mood, A. M., Graybill, F. A., & Boes, D. C. (1974). *Introduction to the theory of statistics* (3rd ed.). New York, NY: McGraw-Hill.

Nygård, F., & Sandström, A. (1981). *Measuring income inequality*. Stockholm: Aimqvist & Wicksell International.

Pen, J. (1971). *Income Distribution*. London: Allen Lane, The Penguin Press.

Pietra, G. (1915). Delle relazioni tra gli indici di variabilità (Nota I). *Atti Del Reale Istituto Veneto Di Scienze, Lettere e Arti*, *LXXIV*(Parte Prima), 775–792.

Pietra, G. (2014). On the relationships between variability indices (Note I). *METRON*, *72*(1), 5–16.

Ramberg, J. S., & Schmeiser, B. W. (1972). An approximate method for generating symmetric random variables. *Communications of the ACM*, *15*(11), 987–990. https://doi.org/10.1145/360827.360840

Ramberg, J. S., & Schmeiser, B. W. (1974). An approximate method for generating asymmetric random variables. *Communications of the ACM*, *17*(2), 78–82. https://doi.org/10.1145/360827.360840

Rothschild, M., & Stiglitz, J. E. (1973). Some further results on the measurement of inequality. *Journal of Economic Theory*, *6*(2), 188–204. https://doi.org/10.1016/0022-0531(73)90034-3

Stigler, S. M. (1974). Linear Functions of Order Statistics with Smooth Weight Functions. *The Annals of Statistics*, *2*(4), 676–693. https://doi.org/10.1214/aos/1176342756

Stoppa, G. (1990). A New Model For Income Size Distributions. In C. Dagum & M. Zenga (Eds.), *Income and Wealth Distribution, Inequality and Poverty* (pp. 33–41). Berlin: Springer-Verlag. https://doi.org/10.1007/978-3-642-84250-4_3

Stuart, A. (1954). The correlation between variate-values and ranks in samples from a continuous distribution. *British Journal of Statistical Psychology*, *7*(1), 37–44. https://doi.org/10.1111/j.2044-8317.1954.tb00138.x

Topp, C. W., & Leone, F. C. (1955). A Family of J-Shaped Frequency Functions. *Journal of the American Statistical Association*, *50*(269), 209. https://doi.org/10.2307/2281107

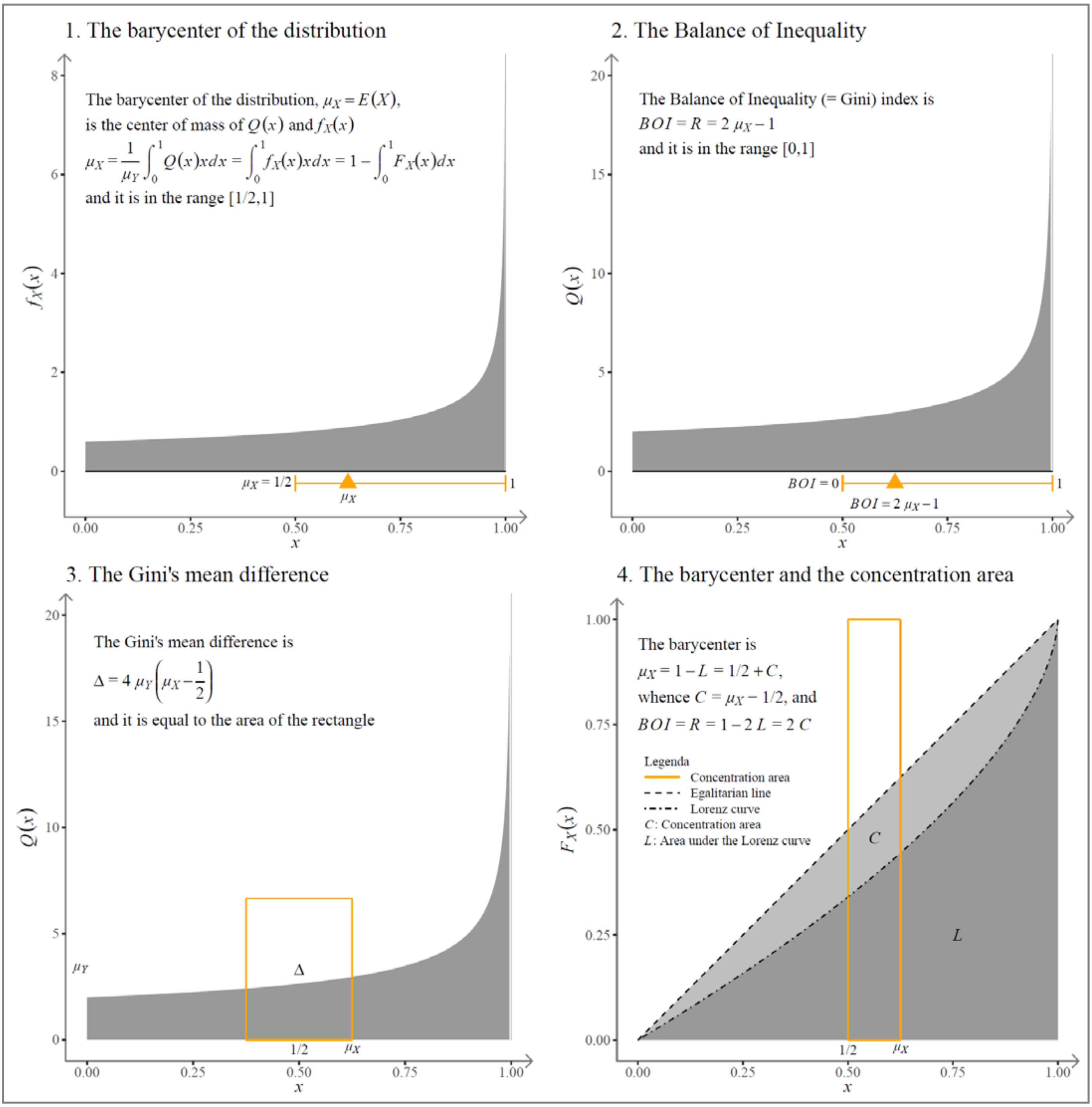
van der Vaart, H. R. (1968). Variances, statistical study of. In *International Encyclopedia of the Social Sciences, Vol. 15* (pp. 292–300). New York, NY: Macmillan.

Vivarelli, M. D. (1992). *Appunti di meccanica razionale*. Milano: Masson.

Weibull, W. (1939). *The Phenomenon of Rupture in Solids*. Stockholm: Generalstabens litografiska anstalts förlag.

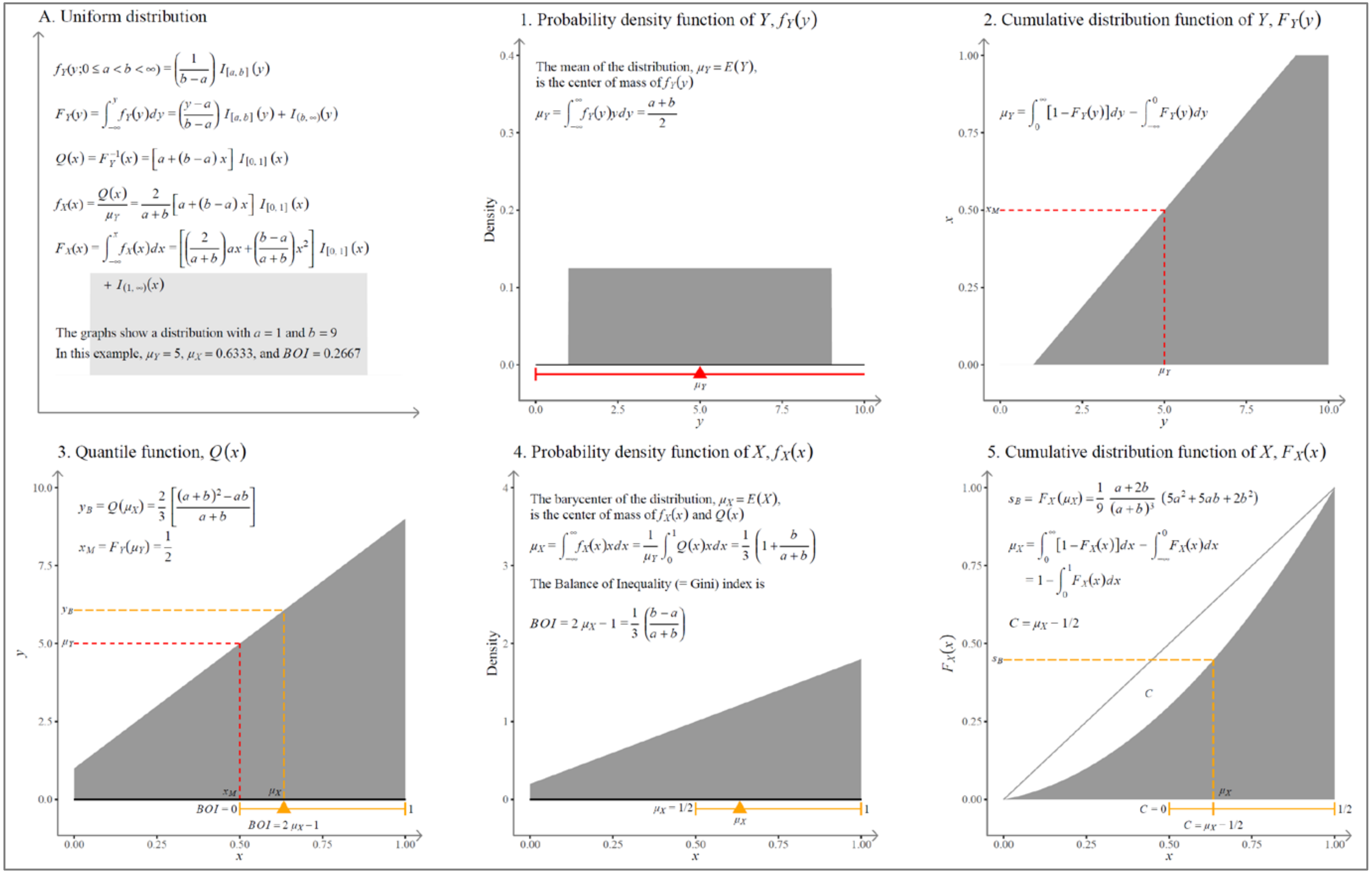
Winkler, O. W. (2009). *Interpreting Economic and Social Data*. Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-68721-4

Figure 1 – Physical interpretation of the barycenter and Balance of inequality (= Gini) index, and geometrical representation of the Gini's mean difference and concentration area



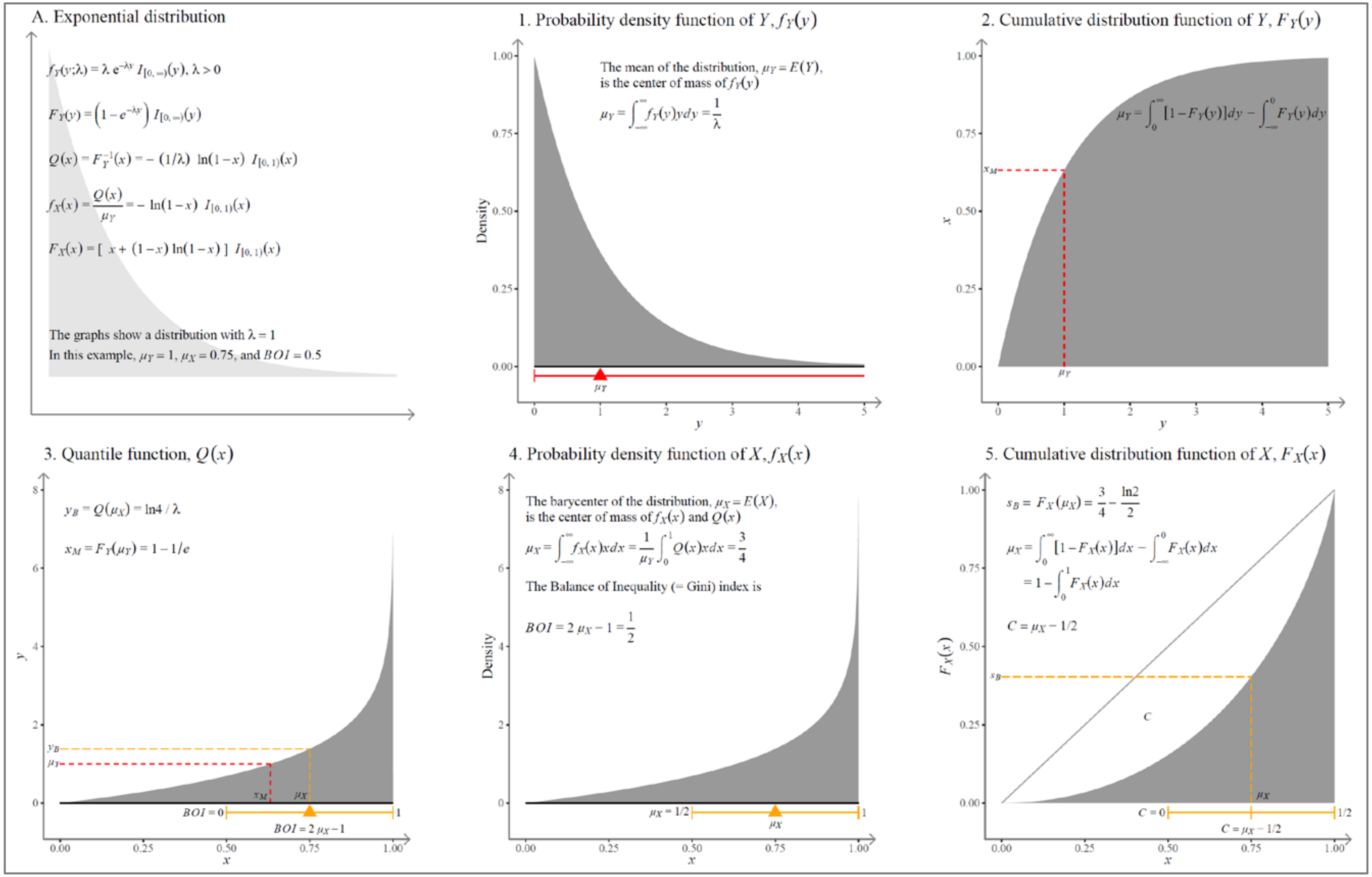
Notes. Using an example Pareto distribution, this figure shows the physical interpretations and geometrical representations introduced in Section 2, 3, and 5. Panel 1 and 2 show the physical interpretation of the barycenter and Balance of Inequality (=Gini) index as the center of mass and normalization of the center of mass, respectively, of the quantile function of Y and probability density function of X. Panel 3 and 4 show the geometrical representation of the Gini's mean difference in the plane of the quantile function of Y and concentration area in the plane of the cumulative distribution function of X, respectively.

Figure 2 – The *Y*-perspective and *X*-perspective, and the barycenter and Balance of Inequality (= Gini) index for a Uniform distribution



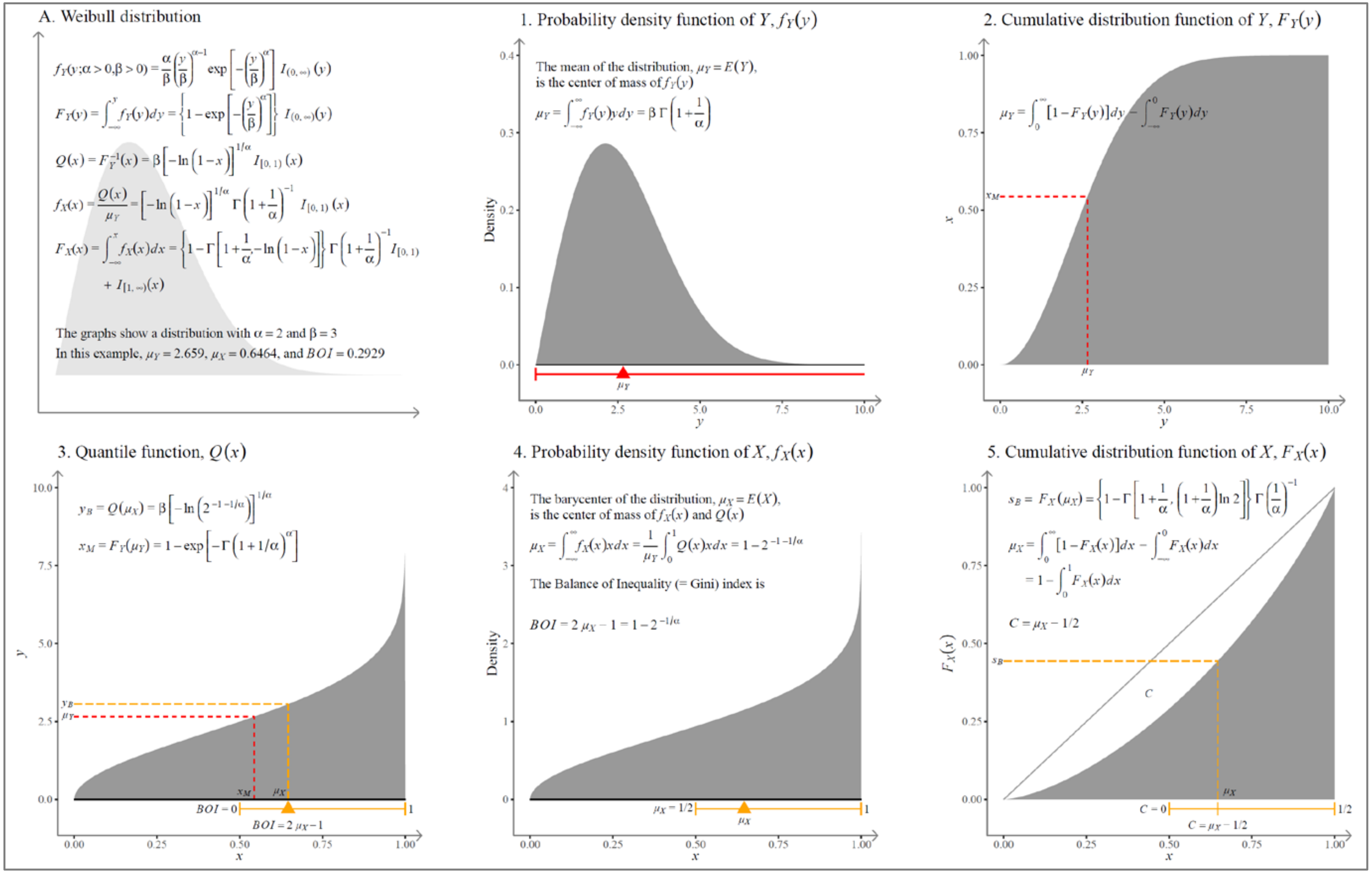
Notes. This figure shows the five characteristic functions of a Uniform distribution (Panel A), the expressions of the mean μ_Y (Panel 2), barycenter μ_X and Balance of Inequality (= Gini) index (Panel 4), location of the mean x_M and barycentric income y_B (Panel 3), and barycentric share s_B (Panel 5). Panel 1 and 4 also show the physical interpretation of the mean and barycenter as the center of mass of $f_Y(y)$ and $f_X(x)$, respectively. Panel 3 shows the physical interpretation of the Balance of Inequality (= Gini) index as the normalization of the barycenter, which is also the center of mass of $Q(x)$. Panel 5 also shows the egalitarian line and the concentration area C. Similar figures for other distributions are shown in Appendix 3.

Figure 3 – The *Y*-perspective and *X*-perspective, and the barycenter and Balance of Inequality (= Gini) index for an Exponential distribution



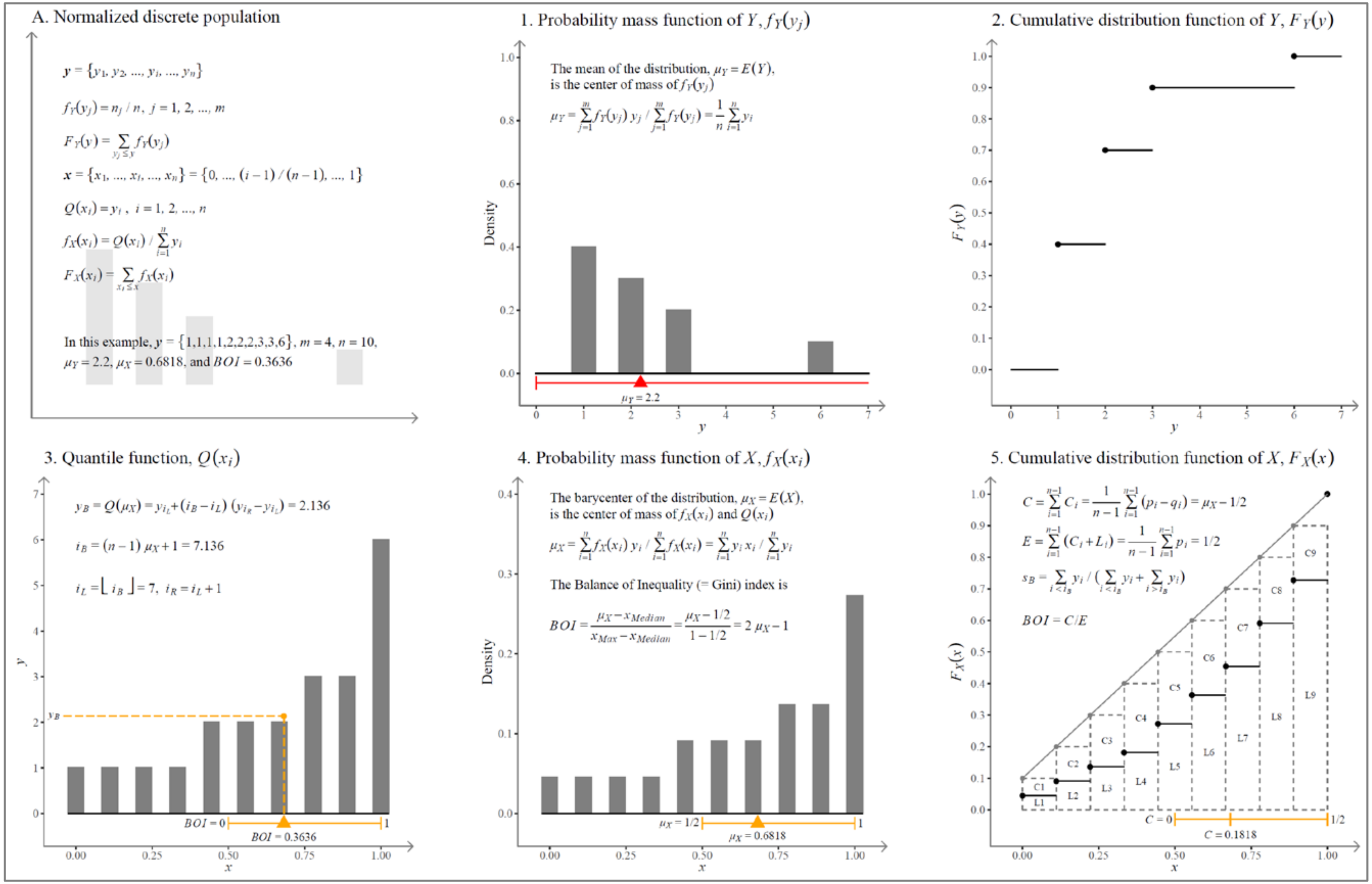
Notes. This figure shows the five characteristic functions of an Exponential distribution (Panel A), the expressions of the mean μ_Y (Panel 2), barycenter μ_X and Balance of Inequality (= Gini) index (Panel 4), location of the mean x_M and barycentric income y_B (Panel 3), and barycentric share s_B (Panel 5). Panel 1 and 4 also show the physical interpretation of the mean and barycenter as the center of mass of $f_Y(y)$ and $f_X(x)$, respectively. Panel 3 shows the physical interpretation of the Balance of Inequality (= Gini) index as the normalization of the barycenter, which is also the center of mass of $Q(x)$. Panel 5 also shows the egalitarian line and the concentration area C. Similar figures for other distributions are shown in Appendix 3.

Figure 4 – The Y-perspective and X-perspective, the barycenter and Balance of Inequality (= Gini) index for a Weibull distribution



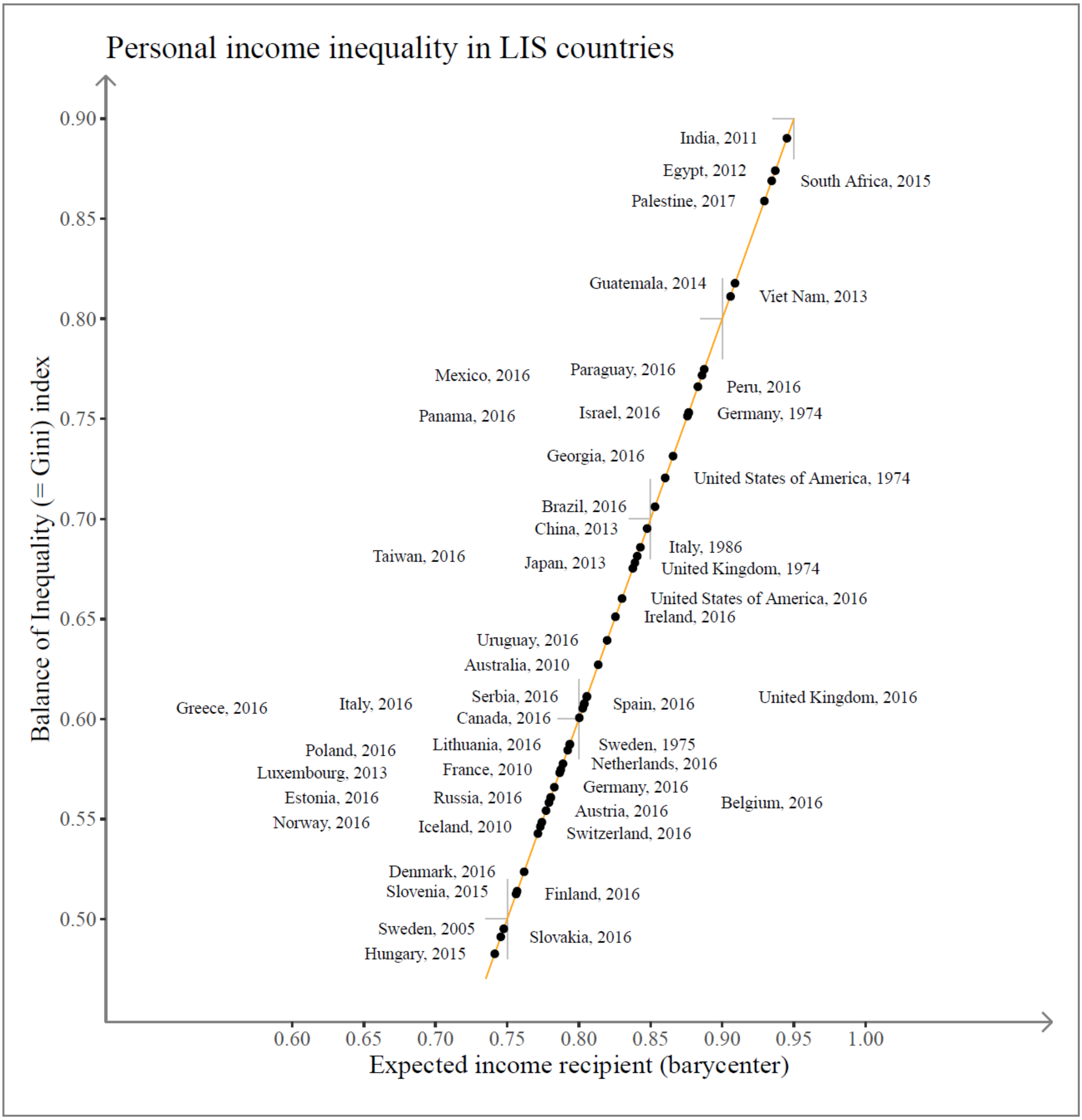
Notes. This figure shows the five characteristic functions of a Weibull distribution (Panel A), the expressions of the mean μ_Y (Panel 2), barycenter μ_X and Balance of Inequality (= Gini) index (Panel 4), location of the mean x_M and barycentric income y_B (Panel 3), and barycentric share s_B (Panel 5). Panel 1 and 4 also show the physical interpretation of the mean and barycenter as the center of mass of $f_Y(y)$ and $f_X(x)$, respectively. Panel 3 shows the physical interpretation of the Balance of Inequality (= Gini) index as the normalization of the barycenter, which is also the center of mass of $Q(x)$. Panel 5 also shows the egalitarian line and the concentration area C. Similar figures for other distributions are shown in Appendix 3.

Figure 5 – The Y-perspective and X-perspective, the barycenter and Balance of Inequality (= Gini) index for a normalized discrete population



Notes. This figure shows the five characteristic functions for a normalized discrete population (Panel A), the expressions of the mean μ_Y (Panel 2), barycenter μ_X and Balance of Inequality (= Gini) index (Panel 4), barycentric income y_B (Panel 3), and barycentric share s_B (Panel 5). Panel 1 and 4 also show the physical interpretation of the mean and barycenter as the center of mass of $f_Y(y_j)$ and $f_X(x_i)$, respectively. Panel 3 shows the physical interpretation of the Balance of Inequality (= Gini) index as the normalization of the barycenter, which is also the center of mass of $Q(x_i)$. Panel 5 also shows the egalitarian line and the concentration area C.

Figure 6 – Empirical application: Personal income inequality in Luxembourg Income Study Database's countries



Notes. This figure shows the results of the empirical application of the methodology for the estimation of the distributions' barycenter and Balance of Inequality (=Gini) index by using weighted observations introduced in Section 7. On the basis of the estimated total individual income inequality in each country, the countries considered are aligned along the line $BOI = 2\mu_X - 1$, which represents the inequality of the income distribution as a function of the barycenter of the distribution, i.e. the expected income recipient. For instance, the maximum value of income inequality is found in India, where the expected income recipient is at the 94.5th percentile and therefore the Balance of Inequality (= Gini) index is around 0.89. The estimates are made by using the total personal income variable (*pitotal*) in the Luxembourg Income Study (LIS) Database provided by the LIS Cross-National Data Center.

Table 1 – The *Y*-perspective and the *X*-perspective, the barycenter and its application to inequality measurement

Function/Measure	**Continuous population**	**Normalized discrete population**
(1)	(2)	(3) $x_i = \{0,...,(i-1)/(n-1),...,1\}$
PANEL A		
The *Y*-perspective and the *X*-perspective		
Probability density/mass function of *Y*	$f_Y(y) = \frac{d}{dy}F_Y(y)$	$f_Y(y_j) = \frac{n_j}{n}$, $j = 1,...,m$,
Cumulative distribution function of *Y*	$F_Y(y) = \int_0^y f(t)dt = Q^{-1}(y)$	$F_Y(y) = \sum_{y_j \le y} f_Y(y_j)$
Quantile function	$Q(x) = F_Y^{-1}(x)$	$Q(x_i) = y_i$, $i = 1,...,n$
Probability density/mass function of *X*	$f_X(x) = Q(x)/\mu_Y$	$f_X(x_i) = y_i / \sum_{i=1}^n y_i$, $i = 1,...,n$
Cumulative distribution function of *X*	$F_X(x) = \int_0^x f(t)dt$	$F_X(x) = \sum_{x_i \le x} f_X(x_i)$
PANEL B		
The mean, the barycenter, and other measures		
Mean	$\mu_Y = E[Y] = \int_0^1 Q(x)dx$ $= \int_0^1 f_Y(y)ydy = 1 - \int_0^1 F_Y(y)dy$	$\mu_Y = \frac{\sum_{j=1}^m f_Y(y_j)y_j}{\sum_{j=1}^m f_Y(y_j)} = \frac{1}{n}\sum_{i=1}^n y_i$
Barycenter	$\mu_X = E[X] = \frac{1}{\mu_Y}\int_0^1 Q(x)xdx$ $= \int_0^1 f_X(x)xdx = 1 - \int_0^1 F_X(x)dx$	$\mu_X = \frac{\sum_{i=1}^n y_i x_i}{\sum_{i=1}^n y_i} = \frac{\sum_{i=1}^n (i-1) y_i}{(n-1)\sum_{i=1}^n y_i}$
Variance of *X*	$\sigma_X^2 = \int_0^1 f_X(x)x^2dx - \mu_X^2$	$\sigma_X^2 = \frac{\sum_{i=1}^n (i-1)^2 y_i}{(n-1)^2 \sum_{i=1}^n y_i} - \mu_X^2$
Location of the mean	$x_M = F_Y(\mu_Y)$	
Barycentric value	$y_B = Q(\mu_X)$	$y_B = y_{i_L} + [i_B - i_L](y_{i_R} - y_{i_L})$
Barycentric rank		$i_B = (n-1)\mu_X + 1$ $i_L = \lfloor i_B \rfloor$, $i_R = i_L + 1$
Barycentric share	$s_B = F_X(\mu_X)$	$s_B = \sum_{i<i_B} y_i / \left(\sum_{i<i_B} y_i + \sum_{i>i_B} y_i\right)$
PANEL C	**For both populations**	**Range**
Barycenter-based formulae	(4)	(5)
Median location	$x_{Median} = 1/2$	
Maximum location	$x_{Max} = 1$	
Balance of inequality (=Gini) index	$BOI = \frac{\mu_X - x_{Median}}{x_{Max} - x_{Median}} = 2\mu_X - 1$	$0 \le BOI \le 1$
Area under the egalitarian line	$E = 1/2$	
Concentration area	$C = \mu_X - 1/2$	$0 \le C \le 1/2$
Area under the Lorenz curve	$L = 1 - \mu_X$	$0 \le L \le 1/2$
Gini's mean difference	$\Delta = 4\mu_Y(\mu_X - 1/2)$	$0 \le \Delta \le 2\mu_Y$

Notes. This table summarizes Section 2, 3, and 5 by showing the main formulae for continuous and normalized discrete populations. Panel A shows the five characteristic functions of distributions, the first two belonging to the *Y*-perspective and the last three to the *X*-perspective. Panel B shows the formulae for calculating the mean, the barycenter, and the other measures introduced in Section 2. Panel C shows the barycenter-based formulae applicable to both continuous and normalized discrete populations for the computation of the Balance of inequality (=Gini) index, the concentration area, the area under the Lorenz curve, and the Gini's mean difference.

Table 2 – The median location, the barycenter, the variance of *X*, and the Balance of Inequality (=Gini) index for four *x*-scales

***x*-scale**	**Distance between two adjacent individuals**	**Median location**	**Barycenter**	**Normalized barycenter**	**Variance of *X***	**Balance of Inequality (=Gini) index**
(1)	(2) $x_{i+1}-x_i$	(3) x_{Median}	(4) $x_B=E[X]=\frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n y_i}$	(5) $\mu_X\in\left[\frac{1}{2},1\right]$	(6) $Var[X]=\frac{\sum_{i=1}^n x_i^2 y_i}{\sum_{i=1}^n y_i}-x_B^2$	$BOI=\frac{x_B-x_{Median}}{x_{Max}-x_{Median}}$
Natural scale (Pen Parade) $x_i=\{1,\ldots,i,\ldots,n\}$	1	$\frac{n+1}{2}$	$x_B=\frac{\sum_{i=1}^n i y_i}{\sum_{i=1}^n y_i}=i_B$	$\mu_X=\frac{x_B-1}{n-1}$	$Var[X]=\frac{\sum_{i=1}^n i^2 y_i}{\sum_{i=1}^n y_i}-x_B^2$	$BOI=\frac{x_B-(n+1)/2}{n-(n+1)/2}$
Lorenz scale $x_i=\left\{\frac{1}{n},\ldots,\frac{i}{n},\ldots,1\right\}$	$\frac{1}{n}$	$\frac{n+1}{2n}$	$x_B=\frac{\sum_{i=1}^n i y_i}{n\sum_{i=1}^n y_i}$	$\mu_X=\frac{nx_B-1}{n-1}$	$Var[X]=\frac{\sum_{i=1}^n i^2 y_i}{n^2\sum_{i=1}^n y_i}-x_B^2$	$BOI=\frac{x_B-(n+1)/2n}{1-(n+1)/2n}$
Normalized scale $x_i=\left\{0,\ldots,\frac{i-1}{n-1},\ldots,1\right\}$	$\frac{1}{n-1}$	$\frac{1}{2}$	$x_B=\frac{\sum_{i=1}^n (i-1) y_i}{(n-1)\sum_{i=1}^n y_i}$	$\mu_X=x_B=\frac{i_B-1}{n-1}$	$\sigma_X^2=\frac{\sum_{i=1}^n (i-1)^2 y_i}{(n-1)^2\sum_{i=1}^n y_i}-\mu_X^2$	$BOI=\frac{\mu_X-1/2}{1-1/2}=2\mu_X-1$
Shifted natural scale $x_i=\{0,\ldots,i-1,\ldots,n-1\}$	1	$\frac{n-1}{2}$	$x_B=\frac{\sum_{i=1}^n (i-1) y_i}{\sum_{i=1}^n y_i}$	$\mu_X=\frac{x_B}{n-1}$	$Var[X]=\frac{\sum_{i=1}^n (i-1)^2 y_i}{\sum_{i=1}^n y_i}-x_B^2$	$BOI=\frac{x_B-(n-1)/2}{(n-1)-(n-1)/2}$

Notes. This table shows in Column 1 the four *x*-scales introduced in Section 2. For each *x*-scale, Column 2 shows the (constant) distance between two adjacent individuals, Column 3 shows the median location, Column 4 and 5 show the expression of the barycenter and the normalized barycenter, respectively, Column 6 shows the expression of the variance of *X*, and Column 7 shows the expression of the Balance of Inequality (=Gini) index.

Table 3 – Examples: Theoretical distributions

	Distribution (1)	Reference (2)	Figure (3)
1	Champernowne-Fisk	(Adapted from Dagum, 1990, p. 10)	Figure A3.1
2	Davies	(Hankin & Lee, 2006, p. 67)	Figure A3.2
3	Exponential	(Mood et al., 1974, p. 112)	Figure A3.3
4	Exponential, Exponentiated	(Adapted from Giorgi & Nadarajah, 2010, p. 40)	Figure A3.4
5	Extreme-value	(Adapted from Beirlant, Vynckier, & Teugels, 1996, p. 293)	Figure A3.5
6	Frechet I	(Adapted from Giorgi & Nadarajah, 2010, p. 31)	Figure A3.6
7	Frechet II	(Adapted from Giorgi & Nadarajah, 2010, p. 31)	Figure A3.7
8	Gumbel	(Adapted from Mood et al., 1974, p. 118)	Figure A3.8
9	Half Logistic	(Adapted from Giorgi & Nadarajah, 2010, p. 41)	Figure A3.9
10	Half Normal	(Adapted from Mood et al., 1974, p. 111)	Figure A3.10
11	Kumaraswamy	(Adapted from Kumaraswamy, 1980)	Figure A3.11
12	Log-Gompertz	(Adapted from Dagum, 1990, p. 10)	Figure A3.12
13	Normal, symmetrically truncated	(Adapted from Mood et al., 1974, p. 111)	Figure A3.13
14	Normal, symmetrically truncated 95%	(Adapted from Mood et al., 1974, p. 111)	Figure A3.14
15	Pareto I	(Kleiber & Kotz, 2003, p. 59)	Figure A3.15
16	Pareto II	(Kleiber & Kotz, 2003, p. 60)	Figure A3.16
17	Pareto, Generalized	(Adapted from Giorgi & Nadarajah, 2010, p. 40)	Figure A3.17
18	Power function I	(Adapted from Giorgi & Nadarajah, 2010, p. 33)	Figure A3.18
19	Power function II	(Adapted from Giorgi & Nadarajah, 2010, p. 34)	Figure A3.19
20	Rayleigh	(Devroye, 1986, p. 29)	Figure A3.20
21	Stoppa	(Stoppa, 1990)	Figure A3.21
22	Topp-Leone	(Adapted from Topp & Leone, 1955, p. 212)	Figure A3.22
23	Triangular	(Devroye, 1986, p. 29)	Figure A3.23
24	Tukey Lambda I	(Adapted from Ramberg & Schmeiser, 1972, p. 988)	Figure A3.24
25	Tukey Lambda III	(Ramberg & Schmeiser, 1972, p. 988)	Figure A3.25
26	Tukey Lambda IV	(Ramberg & Schmeiser, 1974, p. 78)	-
27	Tukey Lambda, Generalized	(Hankin & Lee, 2006, p. 71)	Figure A3.26
28	Uniform	(Mood et al., 1974, p. 118)	Figure A3.27
29	U-Quadratic	(Adapted from Giorgi & Nadarajah, 2010, p. 38)	Figure A3.28
30	Weibull	(Kleiber & Kotz, 2003, p. 174)	Figure A3.29

Notes. This table shows the 30 theoretical distributions selected to illustrate the methodology introduced in Section 2 and 3. For each of the distributions listed alphabetically in Column 1, Column 2 shows the reference from which the distribution was taken, with any adaptation required to obtain the distribution of a non-negative random variable, and Column 3 indicates the summary figure in Appendix 3 for that distribution, similar to the figures in the main text for the Uniform (Figure 2), Exponential (Figure 3), and Weibull (Figure 4) distributions.

Table 4 – Examples: Parameters, cumulative distribution function of *Y*, and probability density function of *X*

	Distribution	Parameters	Cumulative distribution function of *Y*	Probability density function of *X*
	(1)	(2)	(3) $F_Y(y)$	(4) $f_X(x)=Q(x)/\mu_Y\left[=F_Y^{-1}(x)/\mu_Y\right]$
1	Champernowne-Fisk	$(\alpha>1,\lambda>0)$	$\left(1+\lambda y^{-\alpha}\right)^{-1} I_{(0,\infty)}(y)$	$\frac{\alpha}{\pi}\sin\left(\frac{\pi}{\alpha}\right)\lambda^{1/\alpha}\left(\frac{1}{x}-1\right)^{-1/\alpha}$ $I_{(0,1)}(y)$
2	Davies	$(c>0,\lambda_1>0,$ $0<\lambda_2\le 1)$	-	$\left[\mathrm{B}\left(1+\lambda_1,1+\lambda_2\right)\right]^{-1}\frac{x^{\lambda_1}}{(1-x)^{\lambda_2}} I_{[0,1)}(x)$
3	Exponential	$(\lambda>0)$	$\left(1-e^{-\lambda y}\right)I_{[0,\infty)}(y)$	$-\ln\left(1-x\right)I_{[0,1)}(x)$
4	Exponential, Exponentiated	$(\lambda>0)$	$\left[1-\exp\left(-\lambda y\right)\right]^{\alpha} I_{[0,\infty)}(y)$	$-\frac{1}{H_\alpha}\ln\left(1-x^{1/\alpha}\right)I_{[0,1)}(x)$
5	Extreme-value	$(0<\alpha<1,$ $\beta>0)$	$\exp\left[-\left(\frac{\alpha y}{\beta}\right)^{-1/\alpha}\right]I_{(0,\infty)}(y)$	$\frac{\left[-\ln(x)\right]^{-\alpha}}{\Gamma\left(1-\alpha\right)}I_{(0,1)}(x)$
6	Frechet I	$(\alpha>1,\beta>0)$	$\exp\left[-\left(\frac{y}{\beta}\right)^{-\alpha}\right]I_{(0,\infty)}(y)$	$\frac{\left[-\ln\left(x\right)\right]^{-1/\alpha}}{\Gamma\left(1-1/\alpha\right)}I_{(0,1)}(x)$
7	Frechet II	$(y_0\ge 0,a>1,\beta>0)$	$\exp\left[-\left(\frac{y-y_0}{\beta}\right)^{-\alpha}\right]I_{(0,\infty)}(y)$	$\frac{y_0+\beta\left[-\ln(x)\right]^{-1/\alpha}}{y_0+\beta\,\Gamma\left(1-1/\alpha\right)}I_{(0,1)}(x)$
8	Gumbel	$(\alpha/\beta>\ln(\ln 10^{21}),$ $\beta>0)$	$\exp\left[-\exp\left(-\frac{y-\alpha}{\beta}\right)\right]I_{(0,\infty)}(y)$	$\frac{\alpha-\beta\ln\left[-\ln(x)\right]}{\alpha+\beta\gamma}I_{(0,1)}(y)$
9	Half Logistic	$(\lambda>0)$	$1-\exp\left(-2\lambda y\right)I_{[0,\infty)}(y)$	$-\ln(1-x)I_{[0,1)}(y)$
10	Half Normal	$(\sigma>0)$	$\mathrm{erf}^{-1}\left(\frac{y}{\sigma\sqrt{2}}\right)I_{[0,\infty)}(y)$	$\sqrt{\pi}\,\mathrm{erf}^{-1}(x)I_{[0,1]}(x)$
11	Kumaraswamy	$(a>0,b>0)$	$\left[1-\left(1-y^{a}\right)^{b}\right]I_{[0,1]}(y)$	$aby^{a-1}\left(1-y^{a}\right)^{b-1}I_{[0,1]}(y)$
12	Log-Gompertz	$(\alpha>1,\lambda>0)$	$\exp\left(-\lambda y^{-\alpha}\right)I_{(0,\infty]}(y)$	$\frac{\left[-\ln(x)\right]^{-1/\alpha}}{\Gamma\left(1-1/\alpha\right)}I_{(0,1)}(y)$
13	Normal, symmetrically truncated	$(\mu>0,\sigma>0,$ $z\le\mu/\sigma)$	$\left\{\frac{1}{2}\left[1+\mathrm{erf}\left(\frac{y-\mu}{\sigma\sqrt{2}}\right)\right]-t\right\}\left(1-2t\right)^{-1}$ $I_{[\mu-z\sigma,\mu+z\sigma]}(y)+I_{(\mu+z\sigma,\infty)}(y)$ $t=\Phi_{\mu,\sigma^2}(a)=1-\Phi_{\mu,\sigma^2}(b)$	$\left\{1+\frac{\sigma}{\mu}\sqrt{2}\,\mathrm{erf}^{-1}\left[\left(1-2t\right)\left(2x-1\right)\right]\right\}$ $I_{[0,1]}(x)$
14	Normal, symmetrically truncated 95%	$(\mu>0,\sigma>0,$ $\mu/\sigma\ge 1.96)$	$\frac{1}{0.95}\left\{\frac{1}{2}\left[1+\mathrm{erf}\left(\frac{y-\mu}{\sigma\sqrt{2}}\right)\right]-0.025\right\}$ $I_{[\mu-1.96\sigma,\mu+1.96\sigma]}(y)+I_{(\mu+1.96\sigma,\infty)}(y)$	$\left\{1+\frac{\sigma}{\mu}\sqrt{2}\,\mathrm{erf}^{-1}\left[0.95\left(2x-1\right)\right]\right\}$ $I_{[0,1]}(x)$
15	Pareto I	$(y_0>0,\alpha>1)$	$\left[1-\left(\frac{y_0}{y}\right)^{\alpha}\right]I_{[y_0,\infty)}(y)$	$\left(\frac{\alpha-1}{\alpha}\right)(1-x)^{-1/\alpha}I_{[0,1)}(x)$
16	Pareto II	$(\alpha>1,\beta>0)$	$\left[1-\left(1+\frac{y}{\beta}\right)^{-\alpha}\right]I_{[0,\infty)}(y)$	$(\alpha-1)\left[(1-x)^{-1/\alpha}-1\right]I_{[0,1)}(x)$
17	Pareto, Generalized	$(\alpha>0,\beta>0)$	$\left[1-\left(1-\frac{\alpha}{\beta}y\right)^{1/\alpha}\right]I_{[0,\beta/\alpha)}(y)$ $+I_{[\beta/\alpha,\infty)}(y)$	$\left(1+\frac{1}{\alpha}\right)\left[1-\left(1-x\right)^{\alpha}\right]I_{[0,1]}(x)$
18	Power function I	$(a>0)$	$y^{a-1}I_{[0,1)}(y)+I_{[1,\infty)}(y)$	$\left(\frac{1+a}{a}\right)x^{1/a}I_{[0,1]}(x)$

	Distribution	Parameters	Cumulative distribution function of *Y*	Probability density function of *X*
	(1)	(2)	(3) $F_Y(y)$	(4) $f_X(x) = Q(x)/\mu_Y \left[= F_Y^{-1}(x)/\mu_Y\right]$
19	Power function II	$(b>0)$	$\left[1-(1-y)^b\right] I_{[0,1)}(y) + I_{[1,\infty)}(y)$	$(1+b)\left[1-(1-x)^{1/b}\right] I_{[0,1]}(x)$
20	Rayleigh	$(\sigma>0)$	$\left[1-\exp\left(-\frac{y^2}{2\sigma^2}\right)\right] I_{[0,\infty)}(y)$	$\frac{2}{\sqrt{\pi}}\left[-\ln(1-x)\right]^{1/2} I_{[0,1)}(x)$
21	Stoppa	$(y_0>0,$ $\alpha>1, \theta>0)$	$\left[1-\left(\frac{y_0}{y}\right)^{\alpha}\right]^{\theta} I_{(y_0,\infty)}(y)$	$\frac{\left(1-x^{1/\theta}\right)^{-1/\alpha}}{\theta\,\mathrm{B}(1-1/\alpha,\theta)} I_{[0,1)}(y)$
22	Topp-Leone	$(0<\alpha<1)$	$\left(2y-y^2\right)^{\alpha} I_{[0,1]}(y) + I_{(1,\infty)}(y)$	$\left[1-\mathrm{B}(1/2, 1+\alpha)/2\right]^{-1}$ $\left[1-\left(1-x^{1/a}\right)^{1/2}\right] I_{[0,1]}(y)$
23	Triangular	$(a>0)$	$\left(\frac{2}{a}y-\frac{1}{a}y^2\right) I_{[0,a]}(y)$ $+I_{(a,\infty)}(y)$	$3\left[1-(1-x)^{1/2}\right] I_{[0,1]}(y)$
24	Tukey Lambda I	$(\lambda>0)$	-	$\left[1+x^{\lambda}-(1-x)^{\lambda}\right] I_{[0,1]}(x)$
25	Tukey Lambda III	$(\lambda_1 \geq \lambda_2^{-1},$ $\lambda_2>0, \lambda_3>0)$	-	$\left[1+\frac{x^{\lambda_3}-(1-x)^{\lambda_3}}{\lambda_1\lambda_2}\right] I_{[0,1]}(x)$
26	Tukey Lambda IV	$(\lambda_1 \geq \lambda_2^{-1},$ $\lambda_2, \lambda_3, \lambda_4>0)$	-	$\left[\frac{\lambda_1\lambda_2+x^{\lambda_3}-(1-x)^{\lambda_4}}{\lambda_1\lambda_2+(1+\lambda_3)^{-1}+(1+\lambda_4)^{-1}}\right] I_{[0,1]}(x)$
27	Tukey Lambda, Generalized	$(\lambda>0,$ $\lambda_1>0, \lambda_2>0)$	-	$\frac{(1+\lambda_1)(1+\lambda_2)}{1+(2+\lambda_1)\lambda_2}$ $\left[1+x^{\lambda_1}-(1-x)^{\lambda_2}\right] I_{[0,1]}(x)$
28	Uniform	$(0 \leq a < b < \infty)$	$\left(\frac{y-a}{b-a}\right) I_{[a,b]}(y) + I_{(b,\infty)}(y)$	$\frac{2}{a+b}\left[a+(b-a)x\right] I_{[0,1]}(x)$
29	U-Quadratic	$(0 \leq a < b < \infty)$	$\frac{1}{2}\left[1-\left(\frac{a+b-2y}{b-a}\right)^3\right] I_{[a,b]}(y)$ $+I_{(b,\infty)}(y)$	$\left[1+\left(\frac{b-a}{a+b}\right)(2y-1)^{1/3}\right] I_{[0,1]}(y)$
30	Weibull	$(\alpha>0, \beta>0)$	$\left\{1-\exp\left[-\left(\frac{y}{\beta}\right)^{\alpha}\right]\right\} I_{(0,\infty)}(y)$	$\frac{\left[-\ln(1-x)\right]^{1/\alpha}}{\Gamma(1+1/\alpha)} I_{[0,1)}(y)$

Notes. In this table, for each of the distributions listed in Table 2, Column 2 shows the values of the distribution's parameters such that the random variable *Y* is non-negative, and Column 3 and 4 show the cumulative distribution function of *Y* and probability density function of *X*, respectively.

Table 5 – Examples: The barycenter of the distribution, the Balance of Inequality (=Gini) index and its range

	Distribution (1)	**Barycenter** (2) $\mu_X = E[X]$	**BOI index (= Gini index)** (3) $BOI = R = 2\mu_X - 1$	(4) Range	(5) R as in
1	Champernowne-Fisk	$\frac{1}{2}\left(1+\frac{1}{\alpha}\right)$	$\frac{1}{\alpha}$	$(0,1)$	
2	Davies	$\frac{1+\lambda_1}{2+\lambda_1-\lambda_2}$	$\frac{\lambda_1+\lambda_2}{2+\lambda_1-\lambda_2}$	$(0,1)$	(Giorgi & Nadarajah, 2010)
3	Exponential	$\frac{3}{4}$	$\frac{1}{2}$	$\frac{1}{2}$	(Giorgi & Nadarajah, 2010)
4	Exponential, Exponentiated	$\frac{H_{2\alpha}}{2H_\alpha}$	$\frac{H_{2\alpha}}{H_\alpha}-1$	$(0,1)$	*
5	Extreme-value	$2^{\alpha-1}$	$2^{\alpha}-1$	$(0,1)$	
6	Frechet I	$2^{-1+1/\alpha}$	$2^{1/\alpha}-1$	$(0,1)$	(Giorgi & Nadarajah, 2010)
7	Frechet II	$\frac{y_0+2^{1/\alpha}\beta\,\Gamma(1-1/\alpha)}{2\left[y_0-2\beta\,\Gamma(1-1/\alpha)\right]}$	$\frac{\beta\,\Gamma(1-1/\alpha)}{y_0+\beta\,\Gamma(1-1/\alpha)}\left(2^{1/\alpha}-1\right)$	$(0,1)$	
8	Gumbel	$\frac{1}{2}\left(1+\frac{\beta\ln 2}{\alpha+\beta\gamma}\right)$	$\frac{\beta\ln 2}{\alpha+\beta\gamma}$	$(0,1)$	
9	Half Logistic	$\frac{3}{4}$	$\frac{1}{2}$	$\frac{1}{2}$	*
10	Half Normal	$\frac{1}{\sqrt{2}}$	$\sqrt{2}-1$	0.4142	(Giorgi & Nadarajah, 2010)
11	Kumaraswamy	$1-\frac{\mathrm{B}(1+1/a,2b)}{\mathrm{B}(1+1/a,b)}$	$1-\frac{2\mathrm{B}(1+1/a,2b)}{\mathrm{B}(1+1/a,b)}$	$(0,1)$	(Giorgi & Nadarajah, 2010)
12	Log-Gompertz	$2^{1/\alpha-1}$	$2^{1/\alpha}-1$	$(0,1)$	
13	Normal, symmetrically truncated	$\frac{\sigma}{\mu\sqrt{\pi}}\left\{\frac{\mathrm{erf}\left[\sqrt{2}\,\mathrm{erf}^{-1}(1-2t)\right]}{(1-2t)^2}-\frac{\sqrt{2}\exp\left[-\mathrm{erf}^{-1}(1-2t)^2\right]}{(1-2t)}\right\}$	$\frac{\sigma}{\mu\sqrt{\pi}}\left\{\frac{\mathrm{erf}\left[\sqrt{2}\,\mathrm{erf}^{-1}(1-2t)\right]}{(1-2t)^2}-\frac{\sqrt{2}\exp\left[-\mathrm{erf}^{-1}(1-2t)^2\right]}{(1-2t)}\right\}$	$(0,1)$	
14	Normal, symmetrically truncated 95%	$\frac{1}{2}+\frac{90}{361}\frac{\sigma}{\mu}$	$\frac{180}{361}\frac{\sigma}{\mu}$	$(0,1)$	
15	Pareto I	$\frac{\alpha}{2\alpha-1}$	$\frac{1}{2\alpha-1}$	$(0,1)$	(Giorgi & Nadarajah, 2010)
16	Pareto II	$\frac{1}{2}\left(\frac{3\alpha-1}{2\alpha-1}\right)$	$\frac{\alpha}{2\alpha-1}$	$\left(\frac{1}{2},1\right)$	*
17	Pareto, Generalized	$\frac{3+\alpha}{4+2\alpha}$	$\frac{1}{2+\alpha}$	$\left(0,\frac{1}{2}\right)$	*
18	Power function I	$\frac{1+a}{1+2a}$	$\frac{1}{1+2a}$	$(0,1)$	(Giorgi & Nadarajah, 2010)
19	Power function II	$\frac{1}{2}\left(\frac{1+3b}{1+2b}\right)$	$\frac{b}{1+2b}$	$\left(0,\frac{1}{2}\right)$	(Giorgi & Nadarajah, 2010)
20	Rayleigh	$1-\frac{1}{2\sqrt{2}}$	$1-\frac{1}{\sqrt{2}}$	0.2929	

	Distribution (1)	Barycenter (2) $\mu_X = E[X]$	BOI index (= Gini index) (3) $BOI = R = 2\mu_X - 1$	(4) Range	(5) R as in
21	Stoppa	$\frac{\mathrm{B}(1-1/\alpha, 2\theta)}{\mathrm{B}(1-1/\alpha, \theta)}$	$\frac{2\mathrm{B}(1-1/\alpha, 2\theta)}{\mathrm{B}(1-1/\alpha, \theta)} - 1$	$(0,1)$	(Stoppa, 1990)
22	Topp-Leone	$\frac{1}{2}\left[\frac{\mathrm{B}(1/2, 1+2\alpha)-2}{\mathrm{B}(1/2, 1+\alpha)-2}\right]$	$\frac{\mathrm{B}(1/2, 1+2\alpha)-2}{\mathrm{B}(1/2, 1+\alpha)-2} - 1$	$(0,1)$	*
23	Triangular	$\frac{7}{10}$	$\frac{2}{5}$	$\frac{2}{5}$	
24	Tukey Lambda I	$\frac{1}{2} - \frac{1}{1+\lambda} + \frac{2}{2+\lambda}$	$\frac{2\lambda}{2+3\lambda+\lambda^2}$	$\left(0, \frac{35}{102}\right)$	*
25	Tukey Lambda III	$\frac{1}{2} + \frac{\lambda_3}{\lambda_1\lambda_2\left(2+3\lambda_3+\lambda_3^2\right)}$	$\frac{2\lambda_3}{\lambda_1\lambda_2\left(2+3\lambda_3+\lambda_3^2\right)}$	$(0,1)$	
26	Tukey Lambda IV	$\left[\lambda_1\lambda_2/2 + (2+\lambda_3)^{-1} - \left(2+3\lambda_4+\lambda_4^2\right)^{-1}\right]\left[\lambda_1\lambda_2 + (1+\lambda_3)^{-1} - (1+\lambda_4)^{-1}\right]^{-1}$	$\left\{2\lambda_4 + \lambda_3\left[2+\lambda_4(6+\lambda_3+\lambda_4)\right]\right\}\left\{(2+\lambda_3)(2+\lambda_4)\left[\lambda_4-\lambda_3 + \lambda_1\lambda_2(1+\lambda_3)(1+\lambda_4)\right]\right\}^{-1}$	$(0,1)$	*
27	Tukey Lambda, Generalized	$\left\{(1+\lambda_1)\left[4+\lambda_2(4+\lambda_1)(3+\lambda_2)\right]\right\}\left\{2(2+\lambda_1)(2+\lambda_2)\left[1+(2+\lambda_1)\lambda_2\right]\right\}^{-1}$	$\frac{2\lambda_2+\lambda_1\left[2+\lambda_2(6+\lambda_1+\lambda_2)\right]}{(2+\lambda_1)(2+\lambda_2)\left[1+(2+\lambda_1)\lambda_2\right]}$	$(0,1)$	*
28	Uniform	$\frac{1}{3}\left(1+\frac{b}{a+b}\right)$	$\frac{1}{3}\left(\frac{b-a}{a+b}\right)$	$\left(0, \frac{1}{3}\right)$	(Giorgi & Nadarajah, 2010)
29	U-Quadratic	$\frac{5}{7} - \frac{3}{7}\frac{a}{(a+b)}$	$\frac{3}{7} - \frac{6}{7}\frac{a}{(a+b)}$	$\left(0, \frac{3}{7}\right)$	*
30	Weibull	$1-2^{-1-1/\alpha}$	$1-2^{-1/\alpha}$	$(0,1)$	(Giorgi & Nadarajah, 2010)

Notes. In this table, for each of the distributions listed in Table 2, Column 2 shows the expression of the barycenter of the distribution, and Column 3 and 4 show the expression of the Balance of Inequality (=Gini) index and its range, respectively. Finally, Column 5 shows the reference in which the same expression of the Gini index can be found, if any. An asterisk in Column 5 indicates that the expression we have obtained for the Gini index by using the barycenter of the distribution is different and simpler than that reported by Giorgi and Nadarajah (2010), which used the Lorenz curve. Column 5 is unreferenced when we were unable to find previous Gini index expressions for that distribution. However, we cannot exclude that they have already been published. Symbols and functions used are listed in Appendix 1.

Table 6 – Examples: The *x*-distance between the barycenter and the location of the mean

	Distribution (1)	Barycenter (2) $\mu_X = E[X]$	Mean's location (3) $x_M = F_Y(\mu_Y)$	*x*-distance (4) $\mu_X - x_M$		Distribution (1)	Barycenter (2) $\mu_X = E[X]$	Mean's location (3) $x_M = F_Y(\mu_Y)$	*x*-distance (4) $\mu_X - x_M$
1	Champernowne-Fisk	$\frac{1}{2}\left(1+\frac{1}{\alpha}\right)$	$\left[1+\left(\frac{\alpha}{\pi}\right)^{\alpha}\csc\left(\frac{\pi}{\alpha}\right)^{-\alpha}\right]^{-1}$	$\mu_X - x_M > 0$	15	Pareto I	$\frac{\alpha}{2\alpha-1}$	$1-\left(\frac{\alpha-1}{\alpha}\right)^{\alpha}$	$\mu_X - x_M < 0$
3	Exponential	$\frac{3}{4}$	$1-\frac{1}{e}$	0.1179	16	Pareto II	$\frac{1}{2}\left(\frac{3\alpha-1}{2\alpha-1}\right)$	$1-\left(\frac{\alpha-1}{\alpha}\right)^{\alpha}$	$\mu_X - x_M > 0$
4	Exponential, Exponentiated	$\frac{H_{2\alpha}}{2H_{\alpha}}$	$\left(1-\cosh H_{\alpha}+\sinh H_{\alpha}\right)^{\alpha}$	$\mu_X - x_M <> 0$	17	Pareto, Generalized	$\frac{3+\alpha}{4+2\alpha}$	$1-\left(\frac{1}{1+\alpha}\right)^{1/\alpha}$	$\mu_X - x_M > 0$
5	Extreme-value	$2^{\alpha-1}$	$\exp\left[-\Gamma(1-\alpha)^{-1/\alpha}\right]$	$\mu_X - x_M <> 0$	18	Power function I	$\frac{1+a}{1+2a}$	$\left(\frac{a}{1+a}\right)^{a}$	$\mu_X - x_M > 0$
6	Frechet I	$2^{-1+1/\alpha}$	$\exp\left[-\Gamma\left(1-\frac{1}{\alpha}\right)\right]^{-\alpha}$	$\mu_X - x_M <> 0$	19	Power function II	$\frac{1}{2}\left(\frac{1+3b}{1+2b}\right)$	$1-\left(\frac{b}{1+b}\right)^{b}$	$\mu_X - x_M > 0$
7	Frechet II	$\frac{y_0+2^{1/\alpha}\beta\,\Gamma(1-1/\alpha)}{2\left[y_0-2\beta\,\Gamma(1-1/\alpha)\right]}$	$\exp\left[-\Gamma\left(1-\frac{1}{\alpha}\right)\right]^{-\alpha}$	$\mu_X - x_M <> 0$	20	Rayleigh	$1-\frac{1}{2\sqrt{2}}$	$1-\exp\left(-\frac{\pi}{4}\right)$	0.1024
8	Gumbel	$\frac{1}{2}\left(1+\frac{\beta\ln 2}{\alpha+\beta\gamma}\right)$	$\exp\left(-\cosh\gamma+\sinh\gamma\right)$	$\mu_X - x_M > 0$	21	Stoppa	$\frac{\mathrm{B}(1-1/\alpha,2\theta)}{\mathrm{B}(1-1/\alpha,\theta)}$	$\left[1-\frac{\mathrm{B}(1-1/\alpha,\theta)^{-\alpha}}{\theta^{\alpha}}\right]^{\theta}$	$\mu_X - x_M < 0$
9	Half Logistic	$\frac{3}{4}$	$1-\frac{1}{e}$	0.1179	22	Topp-Leone	$\frac{1}{2}\left[\frac{\mathrm{B}(1/2,1+2\alpha)-2}{\mathrm{B}(1/2,1+\alpha)-2}\right]$	$\left[1-\frac{1}{4}\mathrm{B}\left(\frac{1}{2},1+\alpha\right)\right]^{\alpha}$	$\left(0,\frac{13}{90}\right)$
10	Half Normal	$\frac{1}{\sqrt{2}}$	$\mathrm{erf}\left(\frac{1}{\sqrt{\pi}}\right)$	0.1320	23	Triangular	$\frac{7}{10}$	$\frac{5}{9}$	0.1444
11	Kumaraswamy	$1-\frac{\mathrm{B}(1+1/a,2b)}{\mathrm{B}(1+1/a,b)}$	$1-\left[1-b^{a}\mathrm{B}(1+1/a,b)^{a}\right]^{b}$	$\mu_X - x_M > 0$	28	Uniform	$\frac{1}{3}\left(1+\frac{b}{a+b}\right)$	$\frac{1}{2}$	$\frac{1}{6}\left(\frac{b-a}{a+b}\right)>0$
12	Log-Gompertz	$2^{1/\alpha-1}$	$\exp\left[-\Gamma(1-1/\alpha)^{-\alpha}\right]$	$\mu_X - x_M <> 0$	29	U-Quadratic	$\frac{5}{7}-\frac{3}{7}\frac{a}{(a+b)}$	$\frac{1}{2}$	$\frac{3}{14}\frac{(b-a)}{(a+b)}>0$
14	Normal, symmetrically truncated 95%	$\frac{1}{2}+\frac{90}{361}\frac{\sigma}{\mu}$	$\frac{1}{2}$	$\mu_X - x_M > 0$	30	Weibull	$1-2^{-1-1/a}$	$1-\exp\left[-\Gamma\left(1+\frac{1}{a}\right)^{a}\right]$	$\mu_X - x_M > 0$

Notes. In this table, for some of the distributions listed in Table 2, Column 2 shows the expression of the barycenter of the distribution, and Column 3 the expression of the location of the mean. Column 4 shows the value or the sign of their difference. When $\mu_X - x_M > 0$, the point (μ_X, y_B) is on the right of the point (x_M, μ_Y) on the curve of the quantile function, and the barycentric value is greater than the mean. $\mu_X - x_M <> 0$ indicates that the sign of the difference depends on the value of the distribution's parameters. Symbols and functions used are listed in Appendix 1.

APPENDIX

(Supplemental material intended for publication online)

APPENDIX 1 SYMBOLS AND FUNCTIONS

This annex describes the symbols and functions used in the article.

Table A1.1 reports the list of these symbols and functions, and some useful identities.

Table A1.1 – Description of the variables

SYMBOLS	
Euler-Mascheroni constant	$\gamma \approx 0.577$
Closed form omitted due to its length	[...]
No closed form	$[ncf]$
INDICATOR FUNCTION	
Indicator function	$I_{\{(0,1]\}} = I_{(0,1]} = \begin{cases} 1 & \text{if } 0 < x \le 1 \\ 0 & \text{otherwise} \end{cases}$
GAMMA FUNCTION	
Euler gamma function	$\Gamma(z>0) = \int_0^{\infty} t^{z-1} e^{-t} dt$
	Identities:
	$\Gamma(1+a) = a\Gamma(a)$
	$\Gamma(1) = 1$
	$\Gamma(1/2) = 2\Gamma(3/2) = \sqrt{\pi}$
	$\Gamma(n) = 1 \cdot 2 \cdot 3 \cdot \ldots \cdot (n-1) = (n-1)!$
	$\Gamma(n+1) = n!$
	$\Gamma(1+z) = z\Gamma(z)$
	$\Gamma(1-z) = -z\Gamma(-z)$
	$\Gamma(x) = (x-1)\Gamma(x-1)$
	$\Gamma(x-1) = \dfrac{\Gamma(x)}{(x-1)}$
	$\Gamma(x)\Gamma(-x) = -\dfrac{\pi}{x\sin(\pi x)}$
	$\Gamma(x)\Gamma(1-x) = -\dfrac{\pi}{\sin(\pi x)}$
Incomplete gamma function	$\Gamma(a>0, z>0) = \int_z^{\infty} y^{a-1} e^{-t} dt$
Generalized incomplete gamma function	$\Gamma(a>0, z_0>0, z_1>0) = \Gamma(a, z_0) - \Gamma(a, z_1)$

(…)

BETA FUNCTION	
Euler beta function	$B(a>0,b>0)=\int_0^1 y^{a-1}(1-y)^{b-1}\,dy$
	Identities:
	$B(a,b)=B(b,a)$
	$B(a,b)=\dfrac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$
	$\dfrac{\Gamma(a)}{\Gamma(a+b)}=\dfrac{B(a,b)}{\Gamma(b)}$
	$\dfrac{\Gamma(b)\Gamma(a+c)}{\Gamma(c)\Gamma(a+b)}=\dfrac{B(a,b)}{\Gamma(a)}\dfrac{\Gamma(a)}{B(a,c)}=\dfrac{B(a,b)}{B(a,c)}$
Incomplete beta function	$\mathrm{B}_z(a>0,b>0)=\int_0^z y^{a-1}(1-y)^{b-1}\,dy$
Regularized incomplete beta function	$I_z(a>0,b>0)=\dfrac{\mathrm{B}_z(a,b)}{\mathrm{B}(a,b)}$

ERROR FUNCTION	
Error function	$\mathrm{erf}(z)=\dfrac{2}{\pi}\int_0^z \exp\left(-t^2\right)dt$
	Identities:
	$\mathrm{erf}(-z)=-\mathrm{erf}(z)$
Complementary error function	$\mathrm{erfc}(z)=1-\mathrm{erf}(z)$
Generalized error function	$\mathrm{erf}(z_0,z_1)=\mathrm{erf}(z_1)-\mathrm{erf}(z_0)$ $=\dfrac{2}{\pi}\int_{z_0}^{z_1}\exp\left(-t^2\right)dt$
Inverse error function	$\mathrm{erf}^{-1}(s)=z,$ $s=\mathrm{erf}(z)$
Imaginary error function	$\mathrm{erfi}(z)=\mathrm{erf}(iz)/i$

HARMONIC NUMBERS	
Harmonic numbers	$H_n=\sum_{i=1}^{n}\dfrac{1}{i}$

TRIGONOMETRIC FUNCTIONS	
Sine of *z*	$\sin(z)$
Cosecant of *z*	$\csc(z)=1/\sin(z)$

HYPERBOLIC FUNCTIONS	
Hyperbolic sine of *z*	$\sinh(z)$
Hyperbolic cosecant of *z*	$\mathrm{csch}(z)=1/\sinh(z)$

FACTORIAL	
Factorial	$n!=1\cdot2\cdot3\cdot\ldots\cdot n=\Gamma(n+1)$ $\Gamma(n)=(n-1)!$

APPENDIX 2 THEORETICAL DISTRIBUTIONS: TABLES

With reference to Section 4, this annex contains the tables representing the application of the methodology introduced in Section 2 and 3 to thirty theoretical distributions of non-negative continuous random variables.

Table A2.1 – Examples: The five characteristic functions: $f_Y(y)$, $F_Y(y)$, $Q(x)$, $f_X(x)$, and $F_X(x)$

	Distribution (1)	Probability density function of *Y* (2) $f_Y(y)$	Cumulative distribution function of *Y* (3) $F_Y(y)$	Quantile function (4) $Q(x)\left[=F_Y^{-1}(y)\right]$	Probability density function of *X* (5) $f_X(x)=Q(x)/\mu_Y$	Cumulative distribution function of *X* (6) $F_X(x)$
1	Champernowne-Fisk	$f_Y(y;\alpha>1,\lambda>0)$ $=\alpha\lambda y^{\alpha-1}\left(\lambda+y^{-\alpha}\right)^{-2}$ $I_{(0,\infty)}(y)$	$\left(1+\lambda y^{-\alpha}\right)^{-1} I_{(0,\infty)}(y)$	$\lambda^{1/\alpha}\left(\frac{1}{x}-1\right)^{-1/\alpha} I_{(0,1)}(y)$	$\frac{\alpha}{\pi}\sin\left(\frac{\pi}{\alpha}\right)\lambda^{1/\alpha}\left(\frac{1}{x}-1\right)^{-1/\alpha}$ $I_{(0,1)}(y)$	$\frac{1}{\pi}\sin\left(\frac{\pi}{\alpha}\right)\left[\alpha-\left(\frac{\alpha}{1+\alpha}\right)\right.$ $\left.{}_2F_1\left(1,1;2+1/\alpha;x\right)\right]$ $\left(\frac{1}{x}-1\right)^{-1/\alpha} x\, I_{(0,1)}(y)$
2	Davies	-	-	$cx^{\lambda_1}(1-x)^{-\lambda_2} I_{[0,1)}(x)$	$\left[\mathrm{B}\left(1+\lambda_1,1+\lambda_2\right)\right]^{-1}$ $\frac{x^{\lambda_1}}{(1-x)^{\lambda_2}} I_{[0,1)}(x)$	$I_x\left(1+\lambda_1,1-\lambda_2\right)$ $I_{[0,1)}(x)+I_{[1,\infty)}(x)$
3	Exponential	$f_Y(y;\lambda>0)$ $=\lambda e^{-\lambda y} I_{[0,\infty)}(y)$	$\left(1-e^{-\lambda y}\right) I_{[0,\infty)}(y)$	$-\frac{1}{\lambda}\ln(1-x) I_{[0,1)}(x)$	$-\ln(1-x) I_{[0,1)}(x)$	$\left[x+(1-x)\ln(1-x)\right]$ $I_{[0,1)}(y)+I_{[1,\infty)}(y)$
4	Exponential, Exponentiated	$f_Y(y;\alpha>0,\lambda>0)$ $=\alpha\lambda\exp(-\lambda y)$ $\left[1-\exp(-\lambda y)\right]^{\alpha-1} I_{[0,\infty)}(y)$	$\left[1-\exp(-\lambda y)\right]^{\alpha} I_{[0,\infty)}(y)$	$-\frac{1}{\lambda}\ln\left(1-x^{1/\alpha}\right) I_{[0,1)}(x)$	$-\frac{1}{H_\alpha}\ln\left(1-x^{1/\alpha}\right) I_{[0,1)}(x)$	$[ncf]$
5	Extreme-value	$f_Y(y;0<\alpha<1,\beta>0)$ $=\frac{1}{\beta}\left(\frac{\alpha y}{\beta}\right)^{-1-1/\alpha}$ $\exp\left[-\left(\frac{\alpha y}{\beta}\right)^{-1/\alpha}\right] I_{(0,\infty)}(y)$	$\exp\left[-\left(\frac{\alpha y}{\beta}\right)^{-1/\alpha}\right] I_{(0,\infty)}(y)$	$\frac{\beta}{\alpha}\left[-\ln(x)\right]^{-\alpha} I_{(0,1)}(x)$	$\frac{\left[-\ln(x)\right]^{-\alpha}}{\Gamma(1-\alpha)}$ $I_{(0,1)}(x)$	$\frac{\Gamma\left[1-\alpha,-\ln(x)\right]}{\Gamma(1-\alpha)} I_{(0,1)}(x)$
6	Frechet I	$f_Y(y;\alpha>1,\beta>0)$ $=\frac{\alpha}{\beta}\left(\frac{y}{\beta}\right)^{-1-\alpha}\exp\left\{-\left(\frac{y}{\beta}\right)^{-\alpha}\right\}$ $I_{(0,\infty)}(y)$	$\exp\left[-\left(\frac{y}{\beta}\right)^{-\alpha}\right] I_{(0,\infty)}(y)$	$\beta\left[-\ln(x)\right]^{-1/\alpha} I_{(0,1)}(x)$	$\frac{\left[-\ln(x)\right]^{-1/\alpha}}{\Gamma(1-1/\alpha)} I_{(0,1)}(x)$	$\frac{\Gamma\left[1-1/\alpha,-\ln(x)\right]}{\Gamma(1-1/\alpha)}$ $I_{(0,1)}(x)+I_{[1,\infty)}(x)$

	Distribution (1)	Probability density function of *Y* (2) $f_Y(y)$	Cumulative distribution function of *Y* (3) $F_Y(y)$	Quantile function (4) $Q(x)\left[=F_Y^{-1}(y)\right]$	Probability density function of *X* (5) $f_X(x)=Q(x)/\mu_Y$	Cumulative distribution function of *X* (6) $F_X(x)$
7	Frechet II	$f_Y(y;y_0\geq 0,\alpha>1,\beta>0)$ $=\frac{\alpha}{\beta}\left(\frac{y-y_0}{\beta}\right)^{-1-\alpha}$ $\exp\left\{-\left(\frac{y-y_0}{\beta}\right)^{-\alpha}\right\}I_{(0,\infty)}(y)$	$\exp\left[-\left(\frac{y-y_0}{\beta}\right)^{-\alpha}\right]I_{(0,\infty)}(y)$	$\left\{y_0+\beta\left[-\ln(x)\right]^{-1/\alpha}\right\}I_{(0,1)}(x)$	$\frac{y_0+\beta\left[-\ln(x)\right]^{-1/\alpha}}{y_0+\beta\,\Gamma\left(1-1/\alpha\right)}I_{(0,1)}(x)$	$\frac{y_0x+\beta\,\Gamma\left[1-1/\alpha,-\ln\left(x\right)\right]}{y_0+\beta\,\Gamma\left(1-1/\alpha\right)}$ $I_{(0,1)}(x)+I_{[1,\infty)}(x)$
8	Gumbel	$f_Y(y;\frac{\alpha}{\beta}>\ln(\ln 10^{21}),\beta>0)=$ $\frac{1}{\beta}\exp\left[\frac{\alpha-y}{\beta}-\exp\left(\frac{\alpha-y}{\beta}\right)\right]$ $I_{(0,\infty)}(y)$	$\exp\left[-\exp\left(-\frac{y-\alpha}{\beta}\right)\right]I_{(0,\infty)}(y)$	$\left\{\alpha-\beta\ln\left[-\ln(x)\right]\right\}I_{(0,1)}(y)$	$\frac{\alpha-\beta\ln\left[-\ln(x)\right]}{\alpha+\beta\gamma}$ $I_{(0,1)}(y)$	$\frac{\alpha x-\beta x\ln\left[-\ln(x)\right]+\beta\,\mathrm{li}(x)}{\alpha+\beta\gamma}$ $I_{(0,1)}(y)$
9	Half Logistic	$f_Y(y;\lambda>0)$ $=2\lambda\exp\left(-2\lambda y\right)$ $I_{[0,\infty)}(y)$	$1-\exp\left(-2\lambda y\right)I_{[0,\infty)}(y)$	$-\frac{1}{2\lambda}\ln(1-x)I_{[0,1)}(y)$	$-\ln(1-x)$ $I_{[0,1)}(y)$	$\left[x+\left(1-x\right)\ln\left(1-x\right)\right]$ $I_{[0,1)}(y)$
10	Half Normal	$f_Y(y;\sigma>0)$ $=2\phi_{0,\sigma^2}(y)I_{[0,\infty)}(y)$ $=\frac{1}{\sigma}\sqrt{\frac{2}{\pi}}\exp\left[-\frac{1}{2}\left(\frac{y}{\sigma}\right)^2\right]$ $I_{[0,\infty)}(y)$	$\mathrm{erf}^{-1}\left(\frac{y}{\sigma\sqrt{2}}\right)I_{[0,\infty)}(y)$	$\sigma\sqrt{2}\,\mathrm{erf}^{-1}(x)I_{[0,1]}(x)$	$\sqrt{\pi}\,\mathrm{erf}^{-1}(x)$ $I_{[0,1]}(x)$	$\left\{1-\exp\left[-\mathrm{erf}^{-1}(x)^2\right]\right\}$ $I_{[0,1]}(x)$
11	Kumaraswamy	$f_Y(y;a>0,b>0)$ $=aby^{a-1}\left(1-y^a\right)^{b-1}I_{[0,1]}(y)$	$\left[1-\left(1-y^a\right)^b\right]I_{[0,1]}(y)$	$\left[1-\left(1-x^{1/b}\right)\right]^{1/a}I_{[0,1]}(x)$	$\frac{\left[1-\left(1-x^{1/b}\right)\right]^{1/a}}{bB\left(1+1/a,b\right)}I_{[0,1]}(x)$	$[ncf]$
12	Log-Gompertz	$f_Y(y;\alpha>1,\lambda>0)$ $=\alpha\lambda y^{-1-\alpha}\exp\left(-\lambda y^{-\alpha}\right)$ $I_{(0,\infty]}(y)$	$\exp\left(-\lambda y^{-\alpha}\right)I_{(0,\infty]}(y)$	$\left[-\frac{1}{\lambda}\ln(x)\right]^{-1/\alpha}I_{(0,1)}(y)$	$\frac{\left[-\ln(x)\right]^{-1/\alpha}}{\Gamma\left(1-1/\alpha\right)}$ $I_{(0,1)}(y)$	$\frac{\Gamma\left[1-1/\alpha,-\ln(x)\right]}{\Gamma\left(1-1/\alpha\right)}$ $I_{(0,1)}(y)+I_{[1,\infty)}(y)$

	Distribution (1)	Probability density function of *Y* (2) $f_Y(y)$	Cumulative distribution function of *Y* (3) $F_Y(y)$	Quantile function (4) $Q(x)\left[=F_Y^{-1}(y)\right]$	Probability density function of *X* (5) $f_X(x)=Q(x)/\mu_Y$	Cumulative distribution function of *X* (6) $F_X(x)$
13	Normal, symmetrically truncated	$f_Y(y;\mu>0,\sigma>0,z\le\mu/\sigma)$ $=\frac{\phi_{\mu,\sigma^2}(y)}{(1-2t)}I_{[\mu-z\sigma,\mu+z\sigma]}(y)$ $t=\Phi_{\mu,\sigma^2}(\mu-z\sigma)$ $=1-\Phi_{\mu,\sigma^2}(\mu+z\sigma)$	$\left\{\frac{1}{2}\left[1+\operatorname{erf}\left(\frac{y-\mu}{\sigma\sqrt{2}}\right)\right]-t\right\}$ $(1-2t)^{-1}$ $I_{[\mu-z\sigma,\mu+z\sigma]}(y)+I_{(\mu+z\sigma,\infty)}(y)$ $t=\Phi_{\mu,\sigma^2}(a)=1-\Phi_{\mu,\sigma^2}(b)$	$\left\{\mu+\sigma\sqrt{2}\operatorname{erf}^{-1}\left[(1-2t)\right.\right.$ $\left.\left.(2x-1)\right]\right\}I_{[0,1]}(x)$	$\left\{1+\frac{\sigma}{\mu}\sqrt{2}\operatorname{erf}^{-1}\left[(1-2t)\right.\right.$ $\left.\left.(2x-1)\right]\right\}I_{[0,1]}(x)$	$x+\frac{\sigma}{\mu\sqrt{2\pi}(1-2t)}$ $\left\{\exp\left[-\operatorname{erf}^{-1}(1-2t)^2\right]\right.$ $\left.-\exp\left\{-\operatorname{erf}^{-1}\left[(1-2t)(2x-1)\right]^2\right\}\right\}$
14	Normal, symmetrically truncated 95%	$f_Y(y;\mu>0,\sigma>0,\mu/\sigma\ge1.96)$ $=\frac{\phi_{\mu,\sigma^2}(y)}{0.95}I_{[\mu-1.96\sigma,\mu+1.96\sigma]}(y)$ $\Phi_{\mu,\sigma^2}(\mu-1.96\sigma)=0.025$ $=1-\Phi_{\mu,\sigma^2}(\mu+1.96\sigma)$	$\frac{1}{0.95}\left\{\frac{1}{2}\left[1+\operatorname{erf}\left(\frac{y-\mu}{\sigma\sqrt{2}}\right)\right]\right.$ $\left.-0.025\right\}I_{[\mu-1.96\sigma,\mu+1.96\sigma]}(y)$ $+I_{(\mu+1.96\sigma,\infty)}(y)$	$\left\{\mu+\sigma\sqrt{2}\operatorname{erf}^{-1}\left[0.95\right.\right.$ $\left.\left.(2x-1)\right]\right\}I_{[0,1]}(x)$	$\left\{1+\frac{\sigma}{\mu}\sqrt{2}\operatorname{erf}^{-1}\left[0.95\right.\right.$ $\left.\left.(2x-1)\right]\right\}I_{[0,1]}(x)$	$\left\{x-\frac{\sigma}{\mu}\left\{\frac{139}{331}\right.\right.$ $\exp\left\{-\operatorname{erf}^{-1}\left[0.95(2x-1)\right]^2\right\}$ $\left.\left.-\frac{55}{894}\right\}I_{[0,1]}(x)$
15	Pareto I	$f_Y(y;y_0>0,\alpha>1)$ $=\frac{\alpha}{y_0}\left(\frac{y_0}{y}\right)^{\alpha}I_{[y_0,\infty)}(y)$	$\left[1-\left(\frac{y_0}{y}\right)^{\alpha}\right]I_{[y_0,\infty)}(y)$	$y_0(1-x)^{-1/\alpha}I_{[0,1)}(x)$	$\left(\frac{\alpha-1}{\alpha}\right)(1-x)^{-1/\alpha}I_{[0,1)}(x)$	$1-(1-x)^{(\alpha-1)/\alpha}I_{[0,1)}(y)$ $+I_{[1,\infty)}(y)$
16	Pareto II	$f_Y(y;y_0>0,\alpha>1)$ $=\frac{\alpha}{y_0}\left(1+\frac{y}{y_0}\right)^{-(1+\alpha)}I_{[0,\infty)}(y)$	$\left[1-\left(1+\frac{y}{\beta}\right)^{-\alpha}\right]I_{[0,\infty)}(y)$	$\beta\left[(1-x)^{-1/\alpha}-1\right]I_{[0,1)}(x)$	$(\alpha-1)\left[(1-x)^{-1/\alpha}-1\right]I_{[0,1)}(x)$	$\left\{\alpha\left[1-(1-x)^{(\alpha-1)/\alpha}\right]-(\alpha-1)x\right\}$ $I_{[0,1)}(y)+I_{[1,\infty)}(y)$
17	Pareto, Generalized	$f_Y(y;\alpha>0,\beta>0)$ $=\frac{1}{\beta}\left(1-\frac{\alpha}{\beta}y\right)^{1/\alpha-1}I_{[0,\beta/\alpha)}(y)$	$\left[1-\left(1-\frac{\alpha}{\beta}y\right)^{1/\alpha}\right]$ $I_{[0,\beta/\alpha)}(y)+I_{[\beta/\alpha,\infty)}(y)$	$\frac{\beta}{\alpha}\left[(1-x)^{\alpha}-1\right]$ $I_{[0,1]}(x)$	$\left(1+\frac{1}{\alpha}\right)\left[1-(1-x)^{\alpha}\right]$ $I_{[0,1]}(x)$	$\frac{1}{\alpha}\left[(1+\alpha)x+(1-x)^{1+\alpha}-1\right]$ $I_{[0,1]}(x)$
18	Power function I	$f_Y(y;a>0)$ $=ay^{a-1}I_{(0,1)}(y)$	$y^{a-1}I_{[0,1)}(y)+I_{[1,\infty)}(y)$	$x^{1/a}I_{[0,1]}(x)$	$\left(\frac{1+a}{a}\right)x^{1/a}I_{[0,1]}(x)$	$x^{(a+1)/a}I_{[0,1]}(y)+I_{(1,\infty)}(y)$
19	Power function II	$f_Y(y;b>0)$ $=\left[b(1-y)^{b-1}\right]I_{(0,1)}(y)$	$\left[1-(1-y)^{b}\right]I_{[0,1)}(y)+I_{[1,\infty)}(y)$	$\left[1-(1-x)^{1/b}\right]I_{[0,1]}(x)$	$(1+b)\left[1-(1-x)^{1/b}\right]I_{[0,1]}(x)$	$\left\{b(1-x)\left[(1-x)^{1/b}-1\right]+x\right\}$ $I_{[0,1]}(y)+I_{(1,\infty)}(y)$

	Distribution (1)	Probability density function of *Y* (2) $f_Y(y)$	Cumulative distribution function of *Y* (3) $F_Y(y)$	Quantile function (4) $Q(x)\left[=F_Y^{-1}(y)\right]$	Probability density function of *X* (5) $f_X(x)=Q(x)/\mu_Y$	Cumulative distribution function of *X* (6) $F_X(x)$
20	Rayleigh	$f_Y(y;\sigma>0)$ $=\frac{y}{\sigma^2}\exp\left\{-\frac{y^2}{\sigma^2}\right\}I_{[0,\infty)}(y)$	$\left[1-\exp\left(-\frac{y^2}{2\sigma^2}\right)\right]I_{[0,\infty)}(y)$	$\sqrt{2}\sigma\left[-\ln(1-x)\right]^{1/2}I_{[0,1)}(x)$	$\frac{2}{\sqrt{\pi}}\left[-\ln(1-x)\right]^{1/2}I_{[0,1)}(x)$	$\left\{\frac{\text{erfi}\left[\ln(1-x)^{1/2}\right]}{\ln(1-x)^{1/2}}-\frac{2}{\pi}(1-x)\right\}$ $\left[-\ln(1-x)\right]^{1/2}I_{[0,1)}(y)+I_{[1,\infty)}(y)$
21	Stoppa	$f_Y(y;y_0>0,\alpha>1,\theta>0)$ $=\frac{\alpha\theta}{y}\left(\frac{y_0}{y}\right)^{\alpha}\left[1-\left(\frac{y_0}{y}\right)^{\alpha}\right]^{\theta-1}$ $I_{(y_0,\infty)}(y)$	$\left[1-\left(\frac{y_0}{y}\right)^{\alpha}\right]^{\theta}I_{(y_0,\infty)}(y)$	$y_0\left(1-x^{1/\theta}\right)^{-1/\alpha}I_{[0,1)}(y)$	$\frac{\left(1-x^{1/\theta}\right)^{-1/\alpha}}{\theta\,\mathrm{B}(1-1/\alpha,\theta)}$ $I_{[0,1)}(y)$	$[ncf]$
22	Topp-Leone	$f_Y(y;0<\alpha<1)$ $=2\alpha(1-y)\left(2y-y^2\right)^{\alpha-1}$ $I_{[0,1]}(y)$	$\left(2y-y^2\right)^{\alpha}I_{[0,1]}(y)+I_{(1,\infty)}(y)$	$\left[1-\left(1-x^{1/a}\right)^{1/2}\right]I_{[0,1]}(y)$	$\left[1-\mathrm{B}(1/2,1+\alpha)/2\right]^{-1}$ $\left[1-\left(1-x^{1/a}\right)^{1/2}\right]I_{[0,1]}(y)$	[...]
23	Triangular	$f_Y(y;a>0)$ $=\frac{2}{a}\left(1-\frac{y}{a}\right)I_{[0,a]}(y)$	$\left(\frac{2}{a}y-\frac{1}{a}y^2\right)I_{[0,a]}(y)$ $+I_{(a,\infty)}(y)$	$a\left[1-(1-x)^{1/2}\right]I_{[0,1]}(y)$	$3\left[1-(1-x)^{1/2}\right]I_{[0,1]}(y)$	$\left\{3x-2\left[1-(1-x)^{3/2}\right]\right\}$ $I_{[0,1]}(y)+I_{(1,\infty)}(y)$
24	Tukey Lambda I	-	-	$\frac{1}{\lambda}+\frac{x^{\lambda}-(1-x)^{\lambda}}{\lambda}I_{[0,1]}(x)$	$\left[1+x^{\lambda}-(1-x)^{\lambda}\right]$ $I_{[0,1]}(x)$	$\left\{\left[1+\lambda-(1-x)^{\lambda}+x^{\lambda}\right]x+(1-x)^{\lambda}\right.$ $\left.-1\right\}(1+\lambda)^{-1}I_{[0,1)}(y)+I_{[1,\infty)}(y)$
25	Tukey Lambda III	-	-	$\left[\lambda_1+\frac{x^{\lambda_3}-(1-x)^{\lambda_3}}{\lambda_2}\right]I_{[0,1]}(x)$	$\left[1+\frac{x^{\lambda_3}-(1-x)^{\lambda_3}}{\lambda_1\lambda_2}\right]$ $I_{[0,1]}(x)$	$\left\{\left[\lambda_1\lambda_2(1+\lambda_3)-(1-x)^{\lambda_3}+x^{\lambda_3}\right]x\right.$ $\left.+(1-x)^{\lambda_3}-1\right\}\left[\lambda_1\lambda_2(1+\lambda_3)\right]^{-1}$ $I_{[0,1)}(y)+I_{[1,\infty)}(y)$
26	Tukey Lambda IV	-	-	$\left[\lambda_1+\frac{x^{\lambda_3}-(1-x)^{\lambda_4}}{\lambda_2}\right]I_{[0,1]}(x)$	$\left[\frac{\lambda_1\lambda_2+x^{\lambda_3}-(1-x)^{\lambda_4}}{\lambda_1\lambda_2+(1+\lambda_3)^{-1}+(1+\lambda_4)^{-1}}\right]$ $I_{[0,1]}(x)$	$\left\{(1+\lambda_4)x^{1+\lambda_3}+(1+\lambda_3)\right.$ $\left.\left[(1-x)^{1+\lambda_4}+\lambda_1\lambda_2(1+\lambda_4)x-1\right]\right\}$ $\left[\lambda_4-\lambda_3+\lambda_1\lambda_2(1+\lambda_3)(1+\lambda_4)\right]^{-1}$

	Distribution	Probability density function of ***Y***	Cumulative distribution function of ***Y***	Quantile function	Probability density function of ***X***	Cumulative distribution function of ***X***
	(1)	(2) $f_Y(y)$	(3) $F_Y(y)$	(4) $Q(x)\left[=F_Y^{-1}(y)\right]$	(5) $f_X(x)=Q(x)/\mu_Y$	(6) $F_X(x)$
27	Tukey Lambda, Generalized	-	-	$\lambda\left[1+x^{\lambda_1}-(1-x)^{\lambda_2}\right]I_{[0,1]}(x)$	$\frac{(1+\lambda_1)(1+\lambda_2)}{1+(2+\lambda_1)\lambda_2}\left[1+x^{\lambda_1}-(1-x)^{\lambda_2}\right]I_{[0,1]}(x)$	$\frac{1}{1+(2+\lambda_1)\lambda_2}\left\{(1+\lambda_2)x^{1+\lambda_1}+(1+\lambda_1)\left[(1-x)^{1+\lambda_2}+(1+\lambda_2)x-1\right]\right\}I_{[0,1)}(y)+I_{[1,\infty)}(y)$
28	Uniform	$f_Y(y;0\le a<b<\infty)=\left(\frac{1}{b-a}\right)I_{[a,b]}(y)$	$\left(\frac{y-a}{b-a}\right)I_{[a,b]}(y)+I_{(b,\infty)}(y)$	$\left[a+(b-a)x\right]I_{[0,1]}(x)$	$\frac{2}{a+b}\left[a+(b-a)x\right]I_{[0,1]}(x)$	$\left[\left(\frac{2a}{a+b}\right)x+\left(\frac{b-a}{a+b}\right)x^2\right]I_{[0,1]}(y)+I_{(1,\infty)}(y)$
29	U-Quadratic	$f_Y(y;0\le a<b<\infty)=\frac{12}{(b-a)^3}\left(y-\frac{a+b}{2}\right)^2 I_{[a,b]}(y)=\frac{3(a+b-2y)^2}{(b-a)^3}I_{[a,b]}(y)$	$\frac{1}{2}\left[1-\left(\frac{a+b-2y}{b-a}\right)^3\right]I_{[a,b]}(y)+I_{(b,\infty)}(y)$	$\left[\frac{a+b}{2}+\left(\frac{b-a}{2}\right)(2y-1)^{1/3}\right]I_{[0,1]}(y)$	$\left[1+\left(\frac{b-a}{a+b}\right)(2y-1)^{1/3}\right]I_{[0,1]}(y)$	$\left\{x+\frac{3}{8}\frac{(a-b)}{(a+b)}\left[1-(2x-1)^{4/3}\right]\right\}I_{[0,1)}(y)+I_{[1,\infty)}(y)$
30	Weibull	$f_Y(y;a>0,\beta>0)=\frac{a}{\beta}\left(\frac{y}{\beta}\right)^{a-1}e^{-(y/\beta)^a}I_{(0,\infty)}(y)$	$\left\{1-\exp\left[-\left(\frac{y}{\beta}\right)^{\alpha}\right]\right\}I_{(0,\infty)}(y)$	$\beta\left[-\ln(1-x)\right]^{1/\alpha}I_{[0,1)}(x)$	$\frac{\left[-\ln(1-x)\right]^{1/\alpha}}{\Gamma(1+1/\alpha)}I_{[0,1)}(y)$	$\left\{1\text{-}\frac{\Gamma\left[(1+1/a),-\ln(1-x)\right]}{\Gamma(1+1/a)}\right\}I_{[0,1)}(y)+I_{[1,\infty)}(y)$

Notes. For each of the distributions listed in Table 2, this table shows its five characteristic functions, i.e., $f_Y(y)$ (Column 2), $F_Y(y)$ (Column.3), $Q(x)$ (Column.4), $f_X(x)$ (Column.5), and $F_X(x)$ (Column.6). Symbols and functions used are listed in Appendix 1.

Table A2.2 – Examples: The median, the mean and its location, the barycenter and the barycentric value, and the *x*-distance between the barycenter and the mean

	Distribution (1)	**Median** (2) $Q(0.5)$	**Mean** (3) $\mu_Y = E[Y]$	**Barycentric value** (4) $y_B = Q(\mu_X)$	**Mean's location** (5) $x_M = F_Y(\mu_Y)$	**Barycenter** (6) $\mu_X = E[X]$	***x*-distance** (7) $\mu_X - x_M$
1	Champernowne-Fisk	$\lambda^{1/\alpha}$	$\frac{\pi}{\alpha}\csc\left(\frac{\pi}{\alpha}\right)\lambda^{1/\alpha}$	$\left(\frac{\alpha-1}{\alpha+1}\right)^{-1/\alpha}\lambda^{1/\alpha}$	$\left[1+\left(\frac{\alpha}{\pi}\right)^{\alpha}\csc\left(\frac{\pi}{\alpha}\right)^{-\alpha}\right]^{-1}$	$\frac{1}{2}\left(1+\frac{1}{\alpha}\right)$	$\mu_X - x_M > 0$
2	Davies	$c2^{\lambda_2-\lambda_1}$	$c\,\mathrm{B}\left(1+\lambda_1, 1-\lambda_2\right)$	$c\left(\frac{1+\lambda_1}{2+\lambda_1-\lambda_2}\right)^{\lambda_1}\left(\frac{1-\lambda_2}{2+\lambda_1-\lambda_2}\right)^{-\lambda_2}$	-	$\frac{1+\lambda_1}{2+\lambda_1-\lambda_2}$	-
3	Exponential	$\frac{\ln 2}{\lambda}$	$\frac{1}{\lambda}$	$\frac{\ln 4}{\lambda}$	$1-\frac{1}{e}$	$\frac{3}{4}$	$\frac{1}{e}-\frac{1}{4}\approx 0.1179$
4	Exponential, Exponentiated	$-\frac{1}{\lambda}\ln\left(1-2^{-1/\alpha}\right)$	$\frac{H_\alpha}{\lambda}$	$-\frac{1}{\lambda}\ln\left[1-\left(\frac{H_{2\alpha}}{2H_\alpha}\right)^{1/\alpha}\right]$	$\left(1-\cosh H_\alpha+\sinh H_\alpha\right)^{\alpha}$	$\frac{H_{2\alpha}}{2H_\alpha}$	$\mu_X - x_M <> 0$
5	Extreme-value	$\frac{\beta}{\alpha}\ln 2^{-\alpha}$	$-\beta\,\Gamma(-\alpha)$	$\frac{\beta}{\alpha}(\alpha-1)^{-\alpha}(-\ln 2)^{-\alpha}$	$\exp\left[-\Gamma(1-\alpha)^{-1/\alpha}\right]$	$2^{\alpha-1}$	$\mu_X - x_M <> 0$
6	Frechet I	$\beta\left[\ln(2)\right]^{-1/\alpha}$	$\beta\,\Gamma\left(1-\frac{1}{\alpha}\right)$	$\beta\left[-\ln\left(2^{-1+1/\alpha}\right)\right]^{-1/\alpha}$	$\exp\left[-\Gamma\left(1-\frac{1}{\alpha}\right)\right]^{-\alpha}$	$2^{-1+1/\alpha}$	$\mu_X - x_M <> 0$
7	Frechet II	$y_0+\beta(\ln 2)^{-1/\alpha}$	$y_0+\beta\,\Gamma\left(1-\frac{1}{\alpha}\right)$	$y_0+\beta\left\{\ln 2 - \ln\left[\frac{y_0+2^{1/\alpha}\beta\,\Gamma(1-1/\alpha)}{y_0+\beta\,\Gamma(1-1/\alpha)}\right]\right\}^{-1/\alpha}$	$\exp\left[-\Gamma\left(1-\frac{1}{\alpha}\right)\right]^{-\alpha}$	$\frac{y_0+2^{1/\alpha}\beta\,\Gamma(1-1/\alpha)}{2\left[y_0-2\beta\,\Gamma(1-1/\alpha)\right]}$	$\mu_X - x_M <> 0$
8	Gumbel	$\alpha-\beta\ln[\ln 2]$	$\alpha+\beta\gamma$	$\alpha-\beta\ln\left[\ln 2-\ln\left(1+\frac{\beta\ln 2}{\alpha+\beta\gamma}\right)\right]$	$\exp(-\cosh\gamma+\sinh\gamma)$	$\frac{1}{2}\left(1+\frac{\beta\ln 2}{\alpha+\beta\gamma}\right)$	$\mu_X - x_M > 0$
9	Half Logistic	$\frac{\ln 2}{2\lambda}$	$\frac{1}{2\lambda}$	$\frac{\ln 2}{\lambda}$	$1-\frac{1}{e}$	$\frac{3}{4}$	0.1179
10	Half Normal	$\sigma\sqrt{2}\,\mathrm{erf}^{-1}\left(\frac{1}{2}\right)$	$\sigma\sqrt{\frac{2}{\pi}}$	$\sigma\sqrt{2}\,\mathrm{erf}^{-1}\left(\frac{1}{\sqrt{2}}\right)$	$\mathrm{erf}\left(\frac{1}{\sqrt{\pi}}\right)$	$\frac{1}{\sqrt{2}}$	0.1320
11	Kumaraswamy	$\left[1-\left(1-2^{-1/b}\right)\right]^{1/a}$	$bB(1+1/a, b)$	$\left\{1-\left[\frac{B(1+1/a,2b)}{B(1+1/a,b)}\right]^{1/b}\right\}^{1/a}$	$1-\left[1-b^a B(1+1/a,b)^a\right]^b$	$1-\frac{B(1+1/a,2b)}{B(1+1/a,b)}$	$\mu_X - x_M > 0$
12	Log-Gompertz	$\left[\frac{\ln 2}{\lambda}\right]^{-1/\alpha}$	$\lambda^{1/\alpha}\,\Gamma(1-1/\alpha)$	$\left[\frac{\lambda}{(1-1/\alpha)\ln 2}\right]^{1/\alpha}$	$\exp\left[-\Gamma(1-1/\alpha)^{-\alpha}\right]$	$2^{1/\alpha-1}$	$\mu_X - x_M <> 0$

	Distribution (1)	Median (2) $Q(0.5)$	Mean (3) $\mu_Y = E[Y]$	Barycentric value (4) $y_B = Q(\mu_X)$	Mean's location (5) $x_M = F_Y(\mu_Y)$	Barycenter (6) $\mu_X = E[X]$	*x*-distance (7) $\mu_X - x_M$
13	Normal, symmetrically truncated	μ	μ	$\mu+\sigma\sqrt{2}\,\mathrm{erf}^{-1}\left\{\frac{\sigma}{\mu\sqrt{\pi}}\right.$ $\left\{-\sqrt{2}\exp\left[-\mathrm{erf}^{-1}(1-2t)^2\right]\right.$ $\left.\frac{\mathrm{erf}\left[\sqrt{2}\,\mathrm{erf}^{-1}(1-2t)\right]}{(1-2t)}\right\}$	$\frac{1}{2}$	$\frac{\sigma}{\mu\sqrt{\pi}}\left\{\frac{\mathrm{erf}\left[\sqrt{2}\,\mathrm{erf}^{-1}(1-2t)\right]}{(1-2t)^2} - \frac{\sqrt{2}\exp\left[-\mathrm{erf}^{-1}(1-2t)^2\right]}{(1-2t)}\right\}$	$\mu_X - x_M > 0$
14	Normal, symmetrically truncated 95%	μ	μ	$\mu+\sigma\sqrt{2}\,\mathrm{erf}^{-1}\left(\frac{9}{19}\frac{\sigma}{\mu}\right)$	$\frac{1}{2}$	$\frac{1}{2}+\frac{90}{361}\frac{\sigma}{\mu}$	$\mu_X - x_M > 0$
15	Pareto I	$y_0 2^{1/\alpha}$	$y_0\left(\frac{\alpha}{\alpha-1}\right)$	$y_0\left(\frac{\alpha-1}{2\alpha-1}\right)^{1/\alpha}$	$1-\left(\frac{\alpha-1}{\alpha}\right)^{\alpha}$	$\frac{\alpha}{2\alpha-1}$	$\mu_X - x_M < 0$
16	Pareto II	$y_0\left(2^{1/\alpha}-1\right)$	$y_0\left(\frac{1}{\alpha-1}\right)$	$\beta\left[2^{1/\alpha}\left(\frac{2\alpha-1}{\alpha-1}\right)^{1/\alpha}-1\right]$	$1-\left(\frac{\alpha-1}{\alpha}\right)^{\alpha}$	$\frac{1}{2}\left(\frac{3\alpha-1}{2\alpha-1}\right)$	$\mu_X - x_M > 0$
17	Pareto, Generalized	$\frac{\beta}{\alpha}\left(2^{-\alpha}-1\right)$	$\frac{\beta}{1+\alpha}$	$\frac{\beta}{\alpha}\left[1-\left(\frac{1+\alpha}{4+2\alpha}\right)^{\alpha}\right]$	$1-\left(\frac{1}{1+\alpha}\right)^{1/\alpha}$	$\frac{3+\alpha}{4+2\alpha}$	$\mu_X - x_M > 0$
18	Power function I	$2^{-1/a}$	$\frac{a}{1+a}$	$\left(\frac{1+a}{1+2a}\right)^{1/a}$	$\left(\frac{a}{1+a}\right)^{a}$	$\frac{1+a}{1+2a}$	$\mu_X - x_M > 0$
19	Power function II	$1-2^{-1/b}$	$\frac{1}{1+b}$	$1-\left(\frac{1+b}{2+4b}\right)^{1/b}$	$1-\left(\frac{b}{1+b}\right)^{b}$	$\frac{1}{2}\left(\frac{1+3b}{1+2b}\right)$	$\mu_X - x_M > 0$
20	Rayleigh	$\sigma(2\ln 2)^{1/2}$	$\sigma\sqrt{\frac{\pi}{2}}$	$\sigma\sqrt{\ln 8}$	$1-\exp\left(-\frac{\pi}{4}\right)$	$1-\frac{1}{2\sqrt{2}}$	0.1024
21	Stoppa	$y_0\left(1-2^{-1/\theta}\right)^{-1/\alpha}$	$y_0\theta\,\mathrm{B}(1-1/\alpha,\theta)$	$y_0\left\{1-\left[\frac{\mathrm{B}(1-1/\alpha,2\theta)}{\mathrm{B}(1-1/\alpha,\theta)}\right]^{1/\theta}\right\}^{-1/\alpha}$	$\left[1-\theta^{-\alpha}\mathrm{B}(1-1/\alpha,\theta)^{-\alpha}\right]^{\theta}$	$\frac{\mathrm{B}(1-1/\alpha,2\theta)}{\mathrm{B}(1-1/\alpha,\theta)}$	$\mu_X - x_M < 0$
22	Topp-Leone	$1-\left(1-2^{-1/a}\right)^{1/2}$	$1-\frac{1}{2}\mathrm{B}\left(\frac{1}{2},1+\alpha\right)$	$1-\left\{1-2^{-1/a}\left[\frac{\mathrm{B}(1/2,1+2\alpha)-2}{\mathrm{B}(1/2,1+\alpha)-2}\right]^{1/a}\right\}^{1/2}$	$\left[1-\frac{1}{4}\mathrm{B}\left(\frac{1}{2},1+\alpha\right)\right]^{\alpha}$	$\frac{1}{2}\left[\frac{\mathrm{B}(1/2,1+2\alpha)-2}{\mathrm{B}(1/2,1+\alpha)-2}\right]$	$\left(0,\frac{13}{90}\right)$
23	Triangular	$a\left(1-\frac{1}{\sqrt{2}}\right)$	$\frac{a}{3}$	$a\left(1-\sqrt{\frac{3}{10}}\right)$	$\frac{5}{9}$	$\frac{7}{10}$	0.1444

	Distribution (1)	Median (2) $Q(0.5)$	Mean (3) $\mu_Y = E[Y]$	Barycentric value (4) $y_B = Q(\mu_X)$	Mean's location (5) $x_M = F_Y(\mu_Y)$	Barycenter (6) $\mu_X = E[X]$	*x*-distance (7) $\mu_X - x_M$
24	Tukey Lambda I	$\frac{1}{\lambda}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda}+\frac{1}{\lambda}\left[\frac{1}{2}-\frac{1}{1+\lambda}+\frac{2}{2+\lambda}\right]^{\lambda} -\frac{1}{\lambda}\left[\frac{1}{2}+\frac{1}{1+\lambda}-\frac{2}{2+\lambda}\right]^{\lambda}$	-	$\frac{1}{2}-\frac{1}{1+\lambda}+\frac{2}{2+\lambda}$	-
25	Tukey Lambda III	λ_1	λ_1	$\lambda_1+\frac{1}{\lambda_2}\left[\frac{1}{2}+\frac{\lambda_3}{\lambda_1\lambda_2\left(2+3\lambda_3+\lambda_3^2\right)}\right]^{\lambda_3} -\frac{1}{\lambda_2}\left[\frac{1}{2}-\frac{\lambda_3}{\lambda_1\lambda_2\left(2+3\lambda_3+\lambda_3^2\right)}\right]^{\lambda_3}$	-	$\frac{1}{2}+\frac{\lambda_3}{\lambda_1\lambda_2\left(2+3\lambda_3+\lambda_3^2\right)}$	-
26	Tukey Lambda IV	$\lambda_1+\frac{2^{-\lambda_3}-2^{-\lambda_4}}{\lambda_2}$	$\frac{1}{2}+\frac{1}{\lambda_2(1+\lambda_3)}-\frac{1}{\lambda_2(1+\lambda_4)}$	[...]	-	$\left[\lambda_1\lambda_2/2+\left(2+\lambda_3\right)^{-1} -\left(2+3\lambda_4+\lambda_4^2\right)^{-1}\right]\left[\lambda_1\lambda_2 +\left(1+\lambda_3\right)^{-1}-\left(1+\lambda_4\right)^{-1}\right]^{-1}$	-
27	Tukey Lambda, Generalized	$\lambda\left[1+2^{-\lambda_1}-2^{-\lambda_2}\right]$	$\lambda\left(1+\frac{1}{1+\lambda_1}-\frac{1}{1+\lambda_2}\right)$	[...]	-	$\{(1+\lambda_1)[4+\lambda_2(4+\lambda_1) (3+\lambda_2)]\}\{2(2+\lambda_1) (2+\lambda_2)[1+(2+\lambda_1)\lambda_2]\}^{-1}$	-
28	Uniform	$\frac{a+b}{2}$	$\frac{a+b}{2}$	$\frac{2}{3}\left[\frac{(a+b)^2-ab}{a+b}\right]$	$\frac{1}{2}$	$\frac{1}{3}\left(1+\frac{b}{a+b}\right)$	$\frac{1}{6}\left(\frac{b-a}{a+b}\right)>0$
29	U-Quadratic	$\frac{a+b}{2}$	$\frac{a+b}{2}$	$\frac{a+b}{2}\left[1+\left(\frac{3}{7}\right)^{1/3}\left(1-\frac{2a}{a+b}\right)^{4/3}\right]$	$\frac{1}{2}$	$\frac{5}{7}-\frac{3}{7}\frac{a}{(a+b)}$	$\frac{3}{14}\frac{(b-a)}{(a+b)}>0$
30	Weibull	$\beta\,(\ln 2)^{1/a}$	$\beta\,\Gamma\left(1+\frac{1}{a}\right)$	$\beta\left[-\ln\left(2^{-1-1/\alpha}\right)\right]^{1/\alpha}$	$1-\exp\left[-\Gamma\left(1+\frac{1}{a}\right)^{a}\right]$	$1-2^{-1-1/a}$	$\mu_X - x_M > 0$

Notes. For each of the distributions listed in Table 2, this table shows the median (Column 2), the mean (Column.3), the barycentric value (Column.4), the location of the mean (Column.5), the barycenter (Column.6), and the *x*-distance between the barycenter and the means location (Column 7). Symbols and functions used are listed in Appendix 1.

APPENDIX 3 THEORETICAL DISTRIBUTIONS: GRAPHICAL REPRESENTATIONS

With reference to Section 4, this annex contains the figures representing the application of the methodology introduced in Section 2 and 3 to thirty theoretical distributions of non-negative continuous random variables.

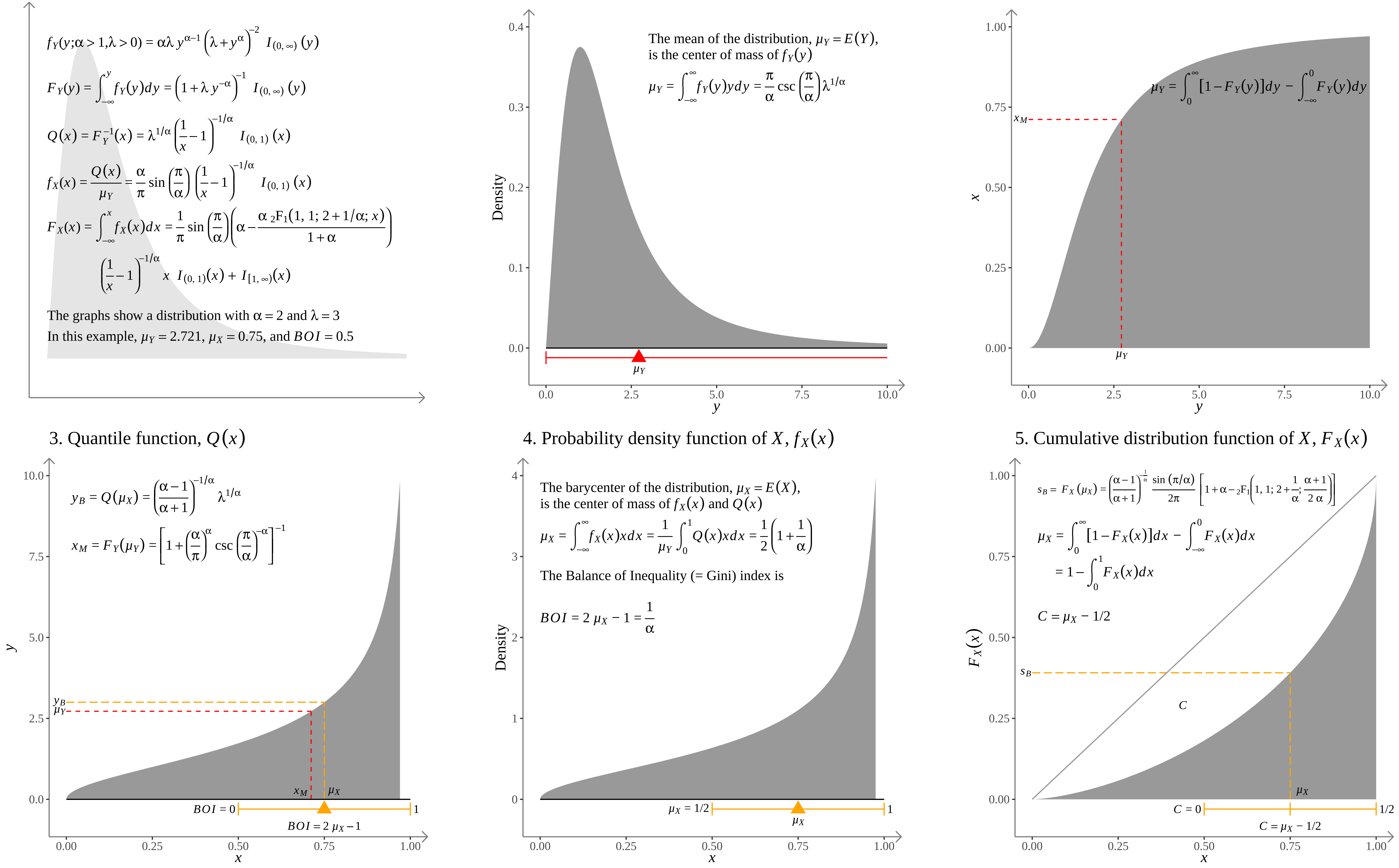


Figure A3.1 − Champernowne−Fisk distribution

Figure A3.2 − Davies distribution

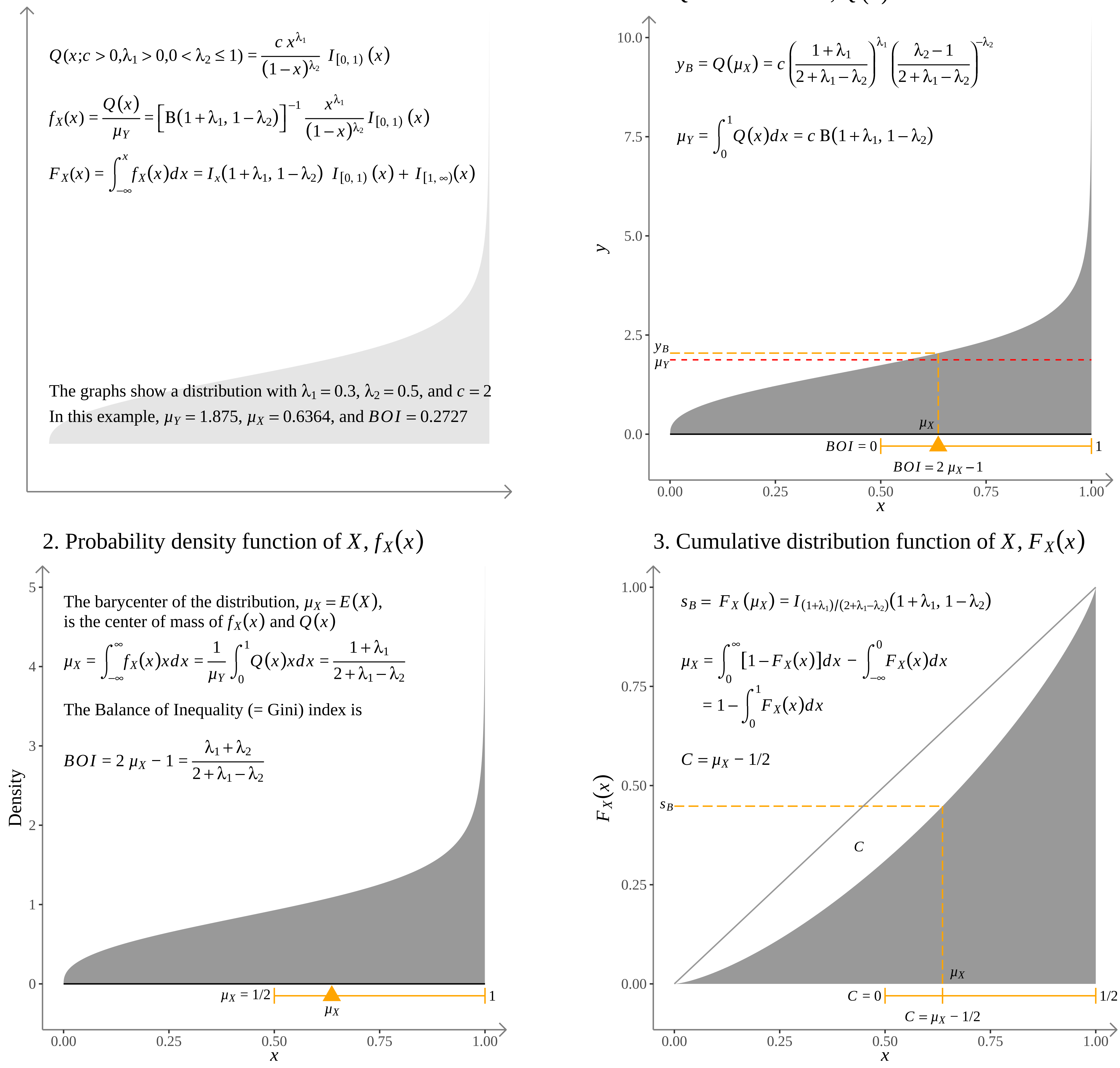
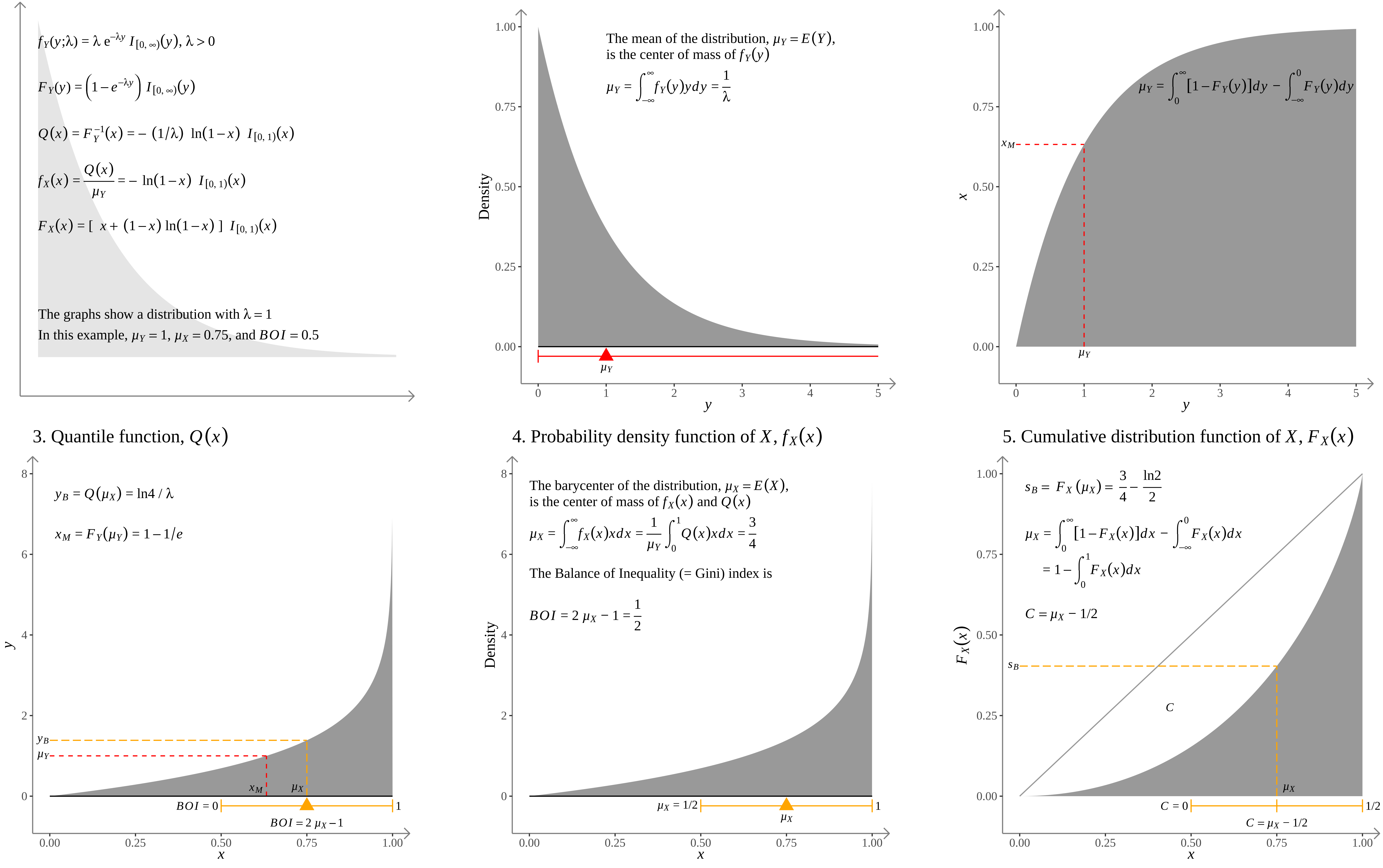


Figure A3.3 − Exponential distribution



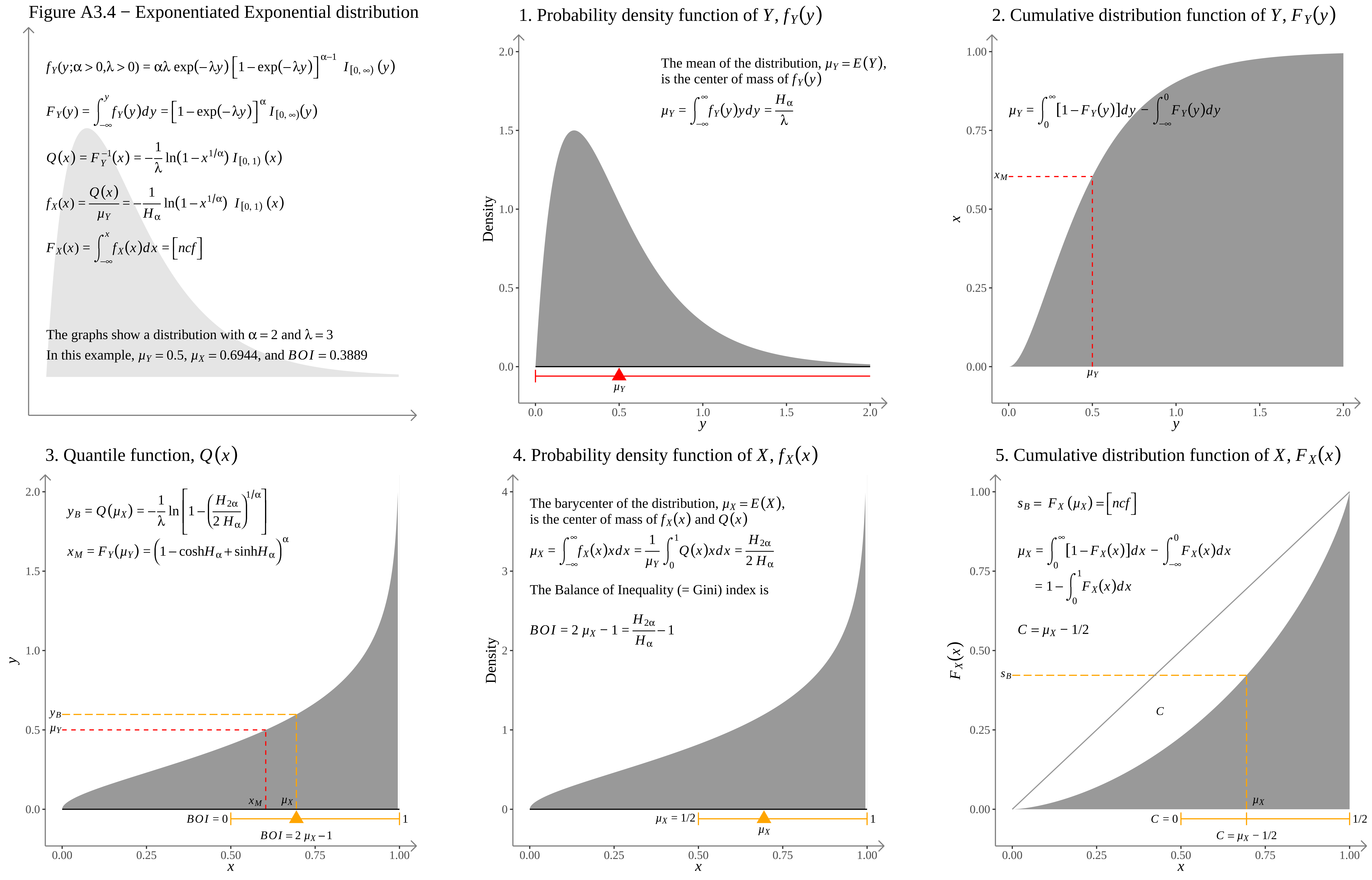


Figure A3.4 − Exponentiated Exponential distribution

Figure A3.5 − Extreme value distribution

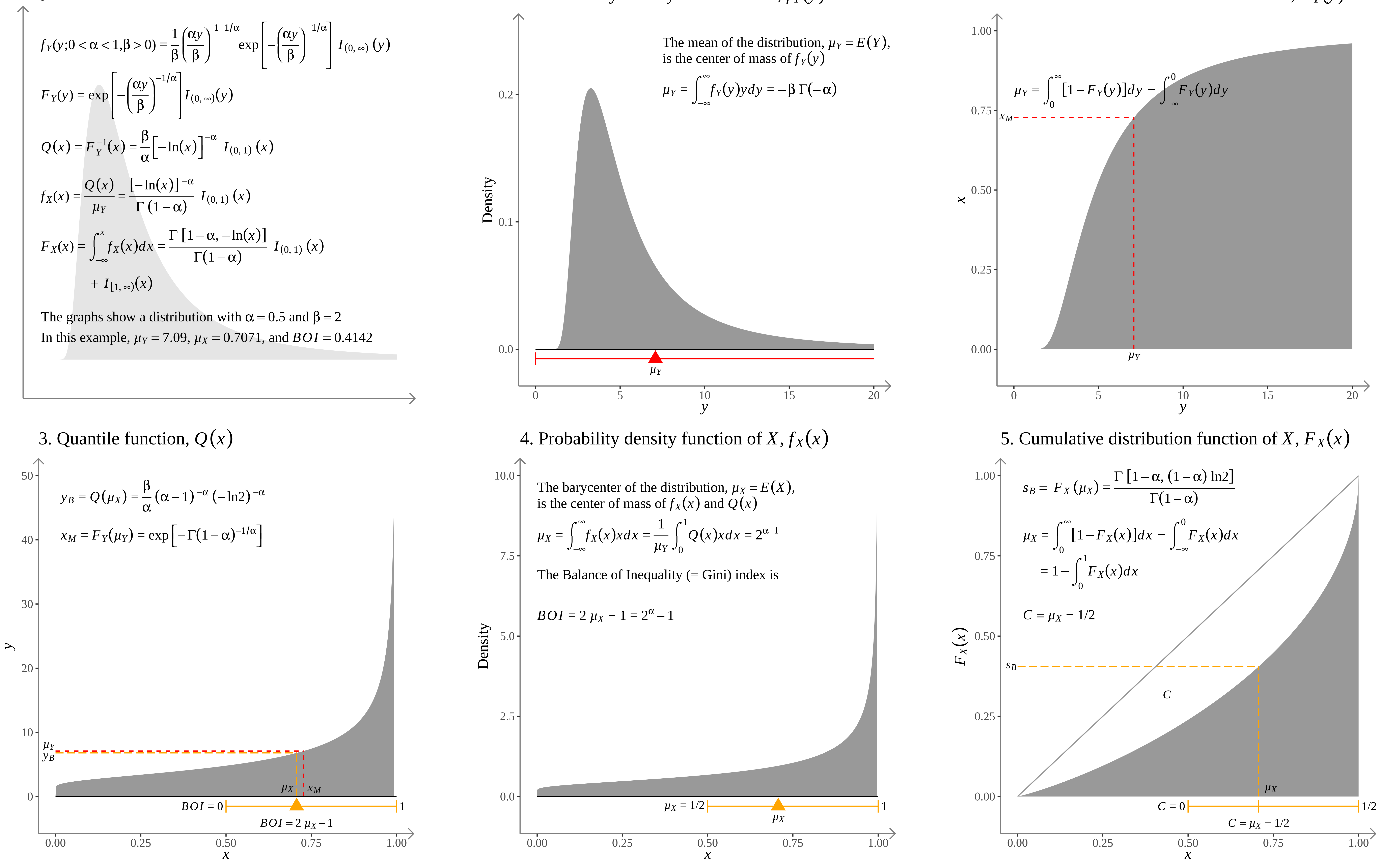


Figure A3.6 − Fréchet I distribution

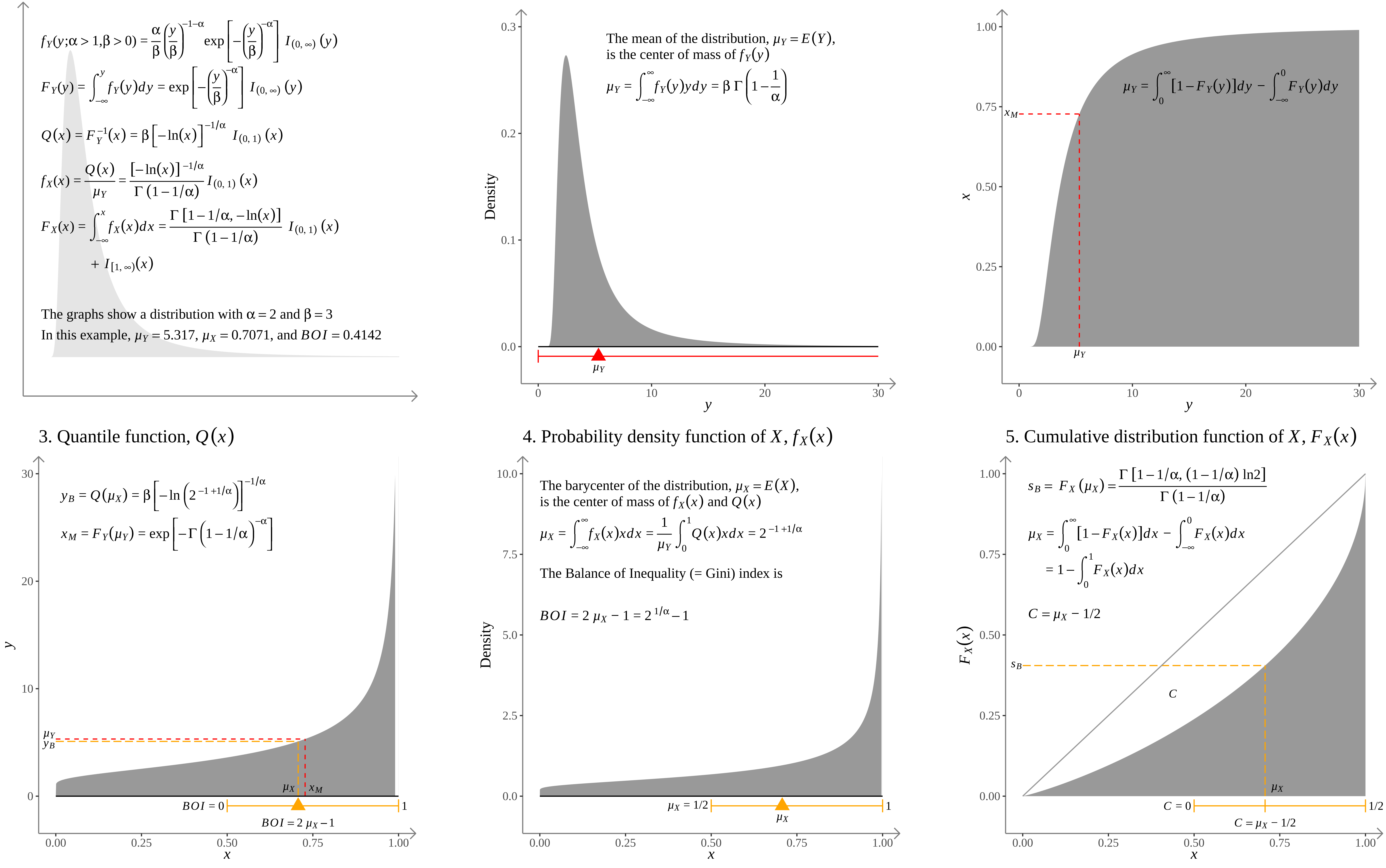


Figure A3.7 − Fréchet II distribution

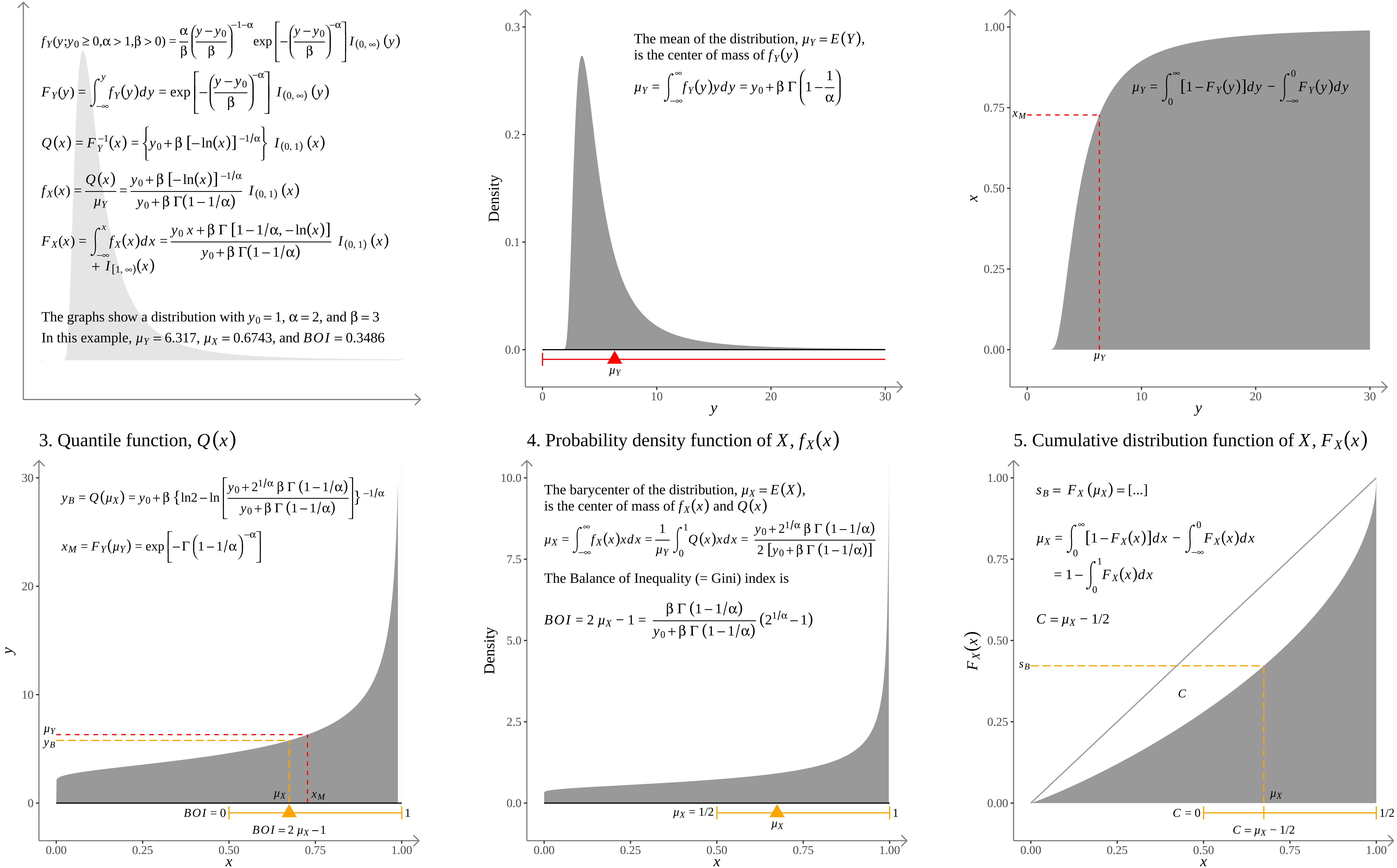


Figure A3.8 − Gumbel distribution

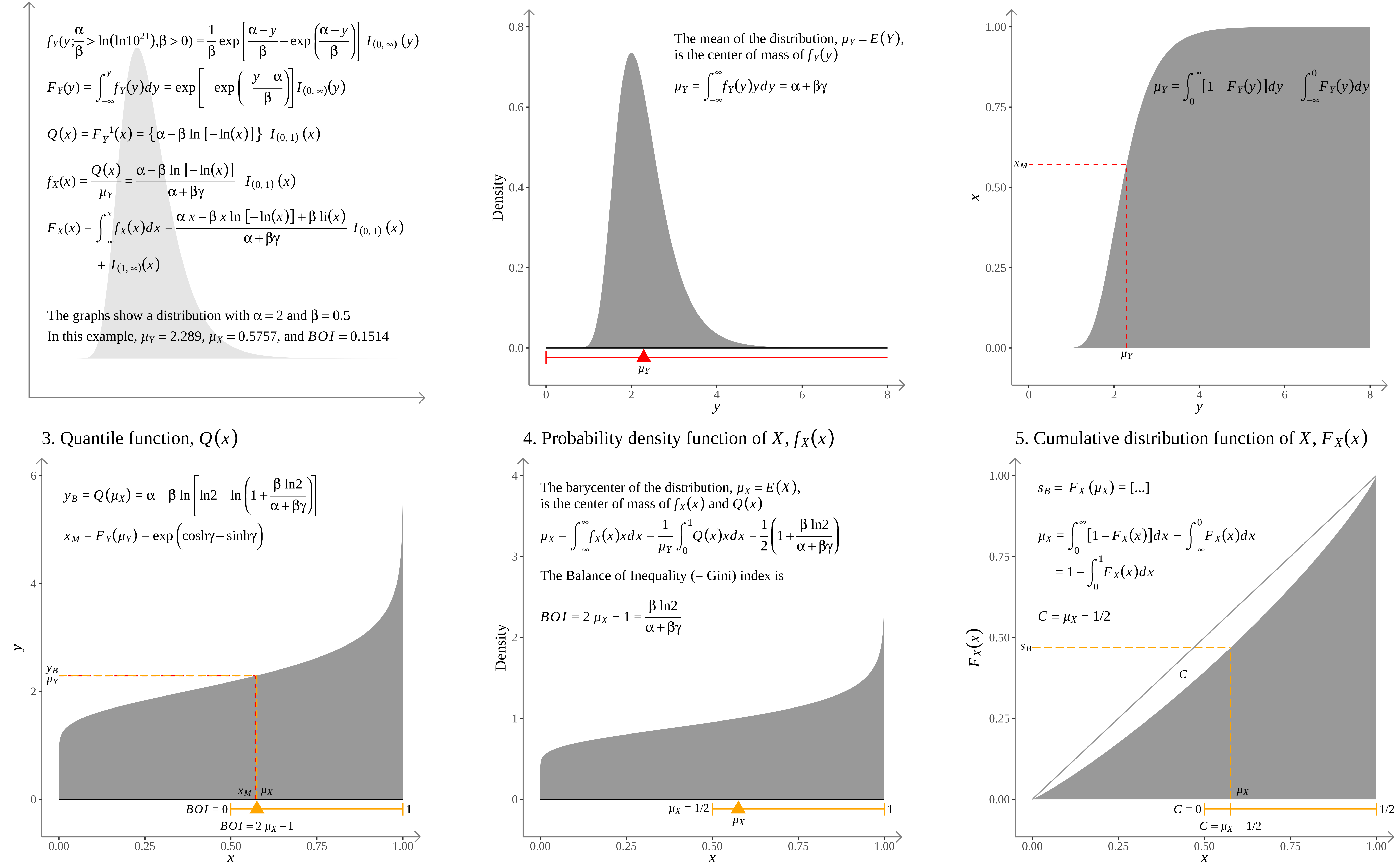


Figure A3.9 − Half Logistic distribution

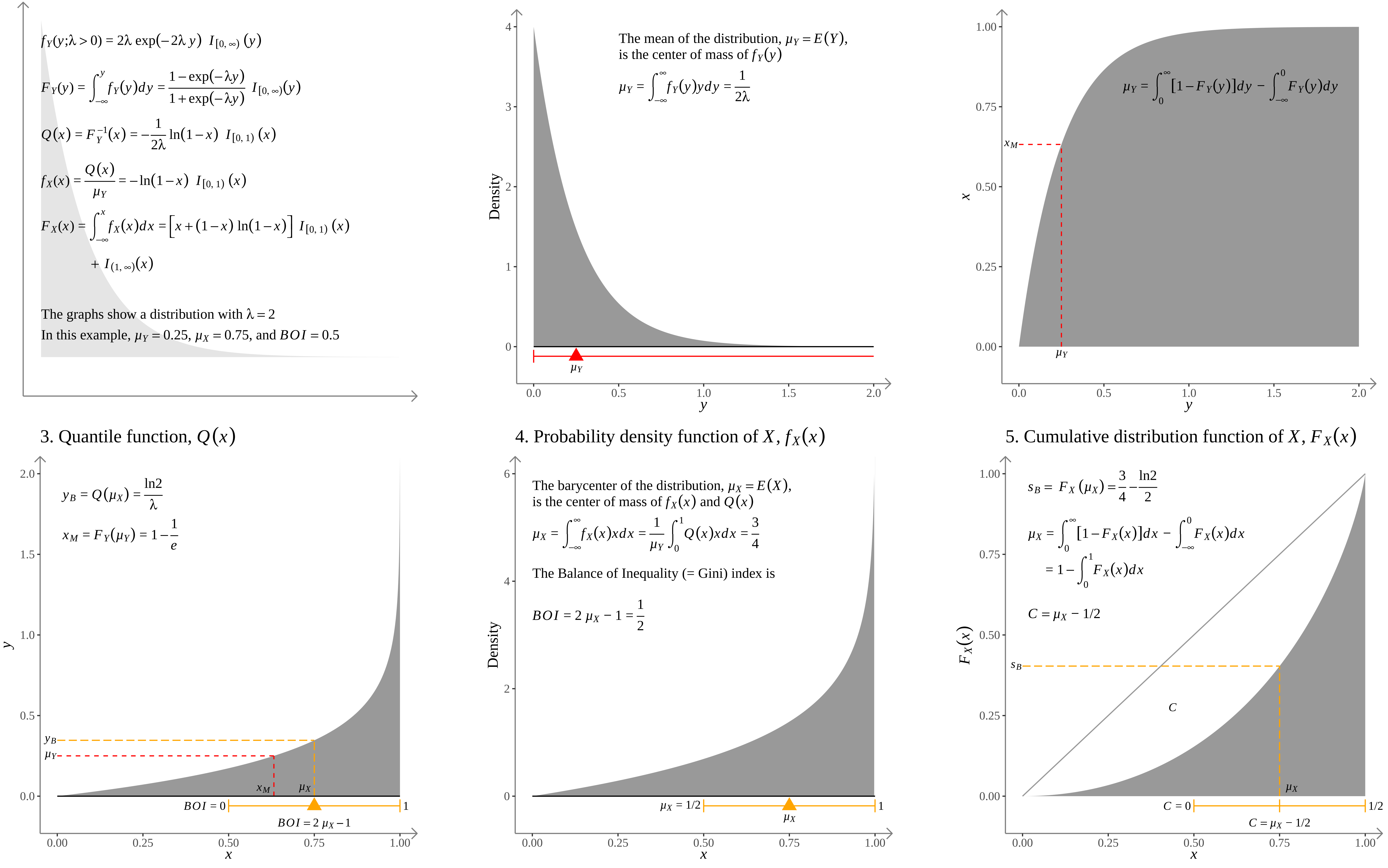
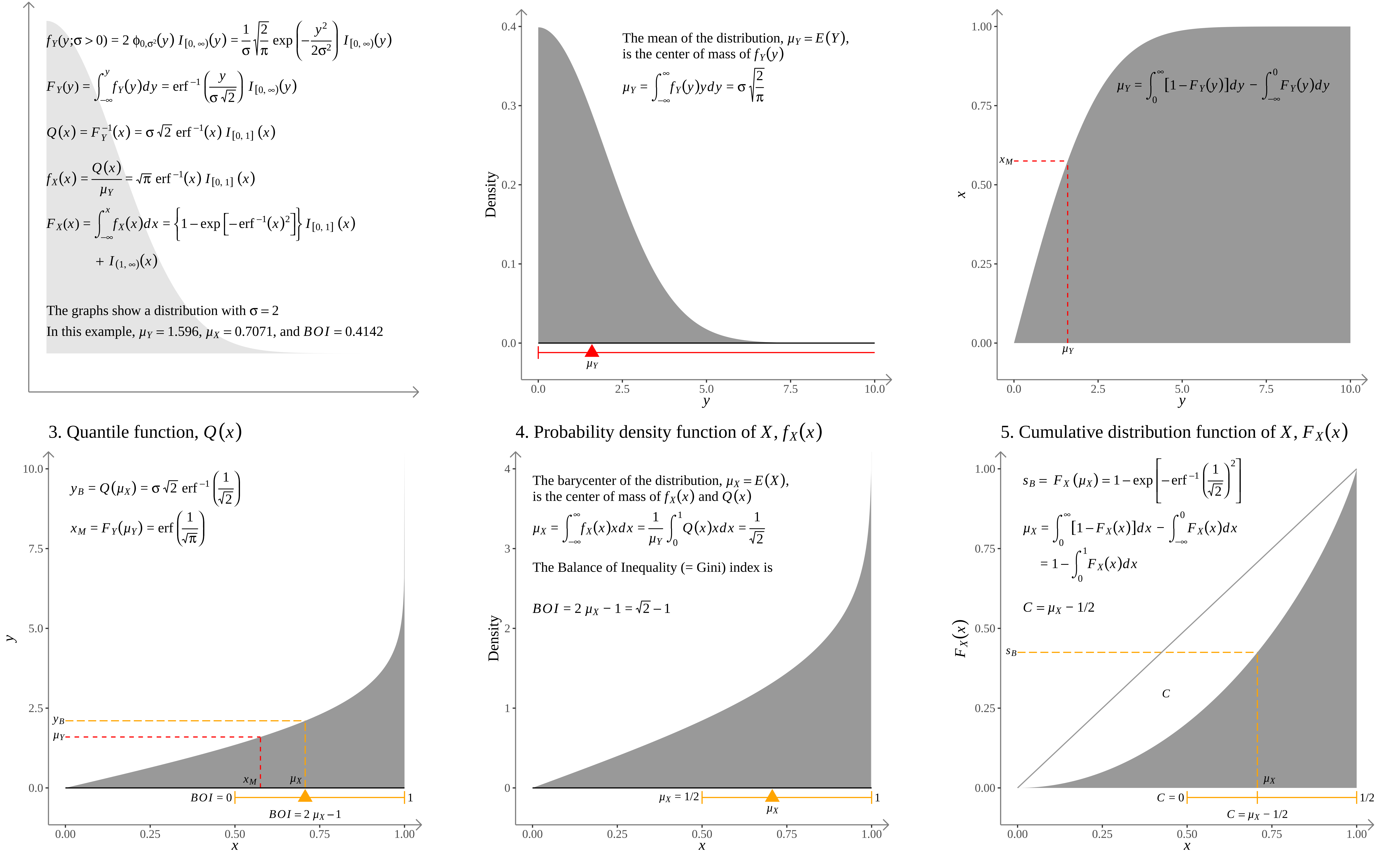


Figure A3.10 − Half Normal distribution



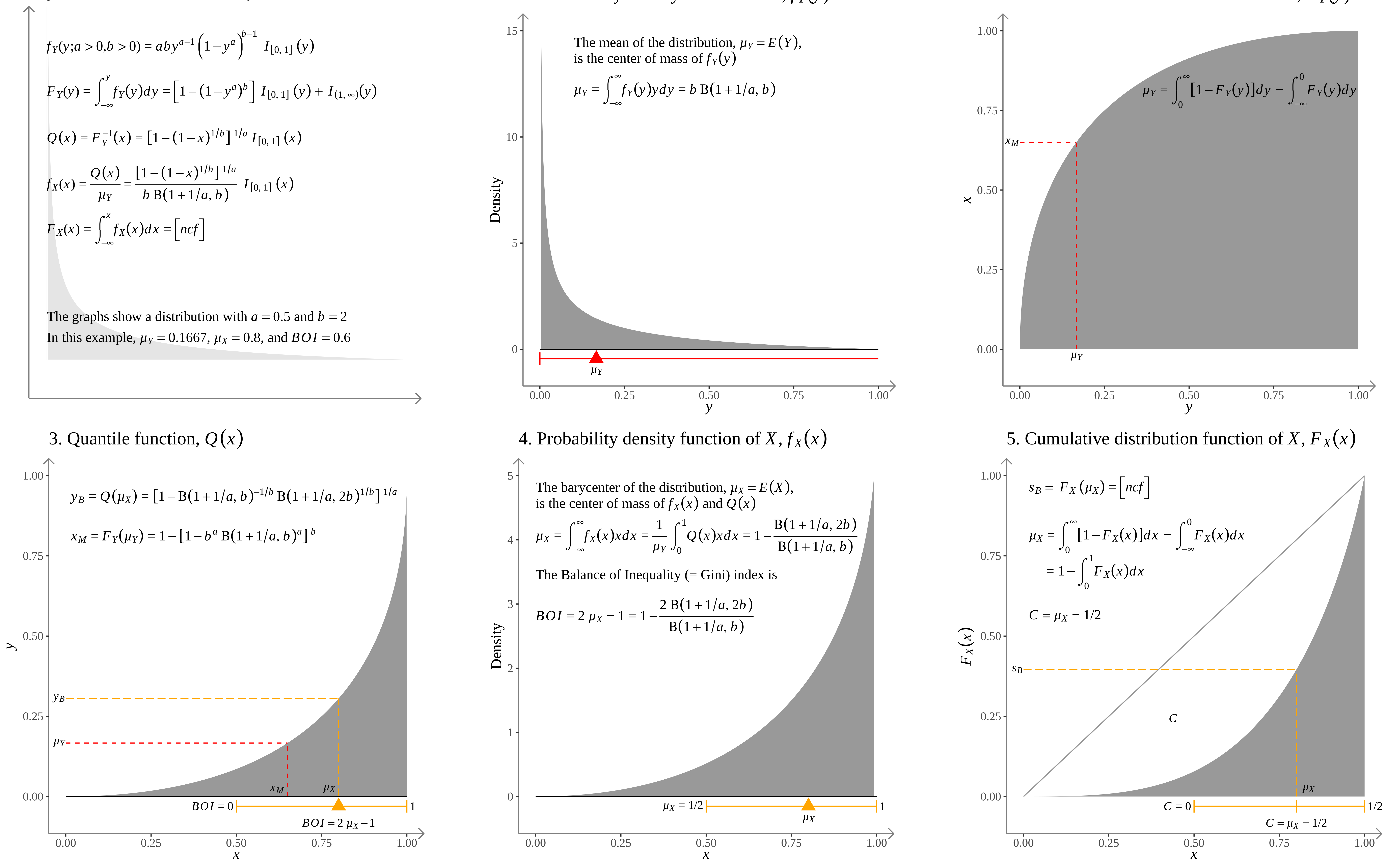


Figure A3.11 − Kumaraswamy distribution

Figure A3.12 − Log−Gompertz distribution

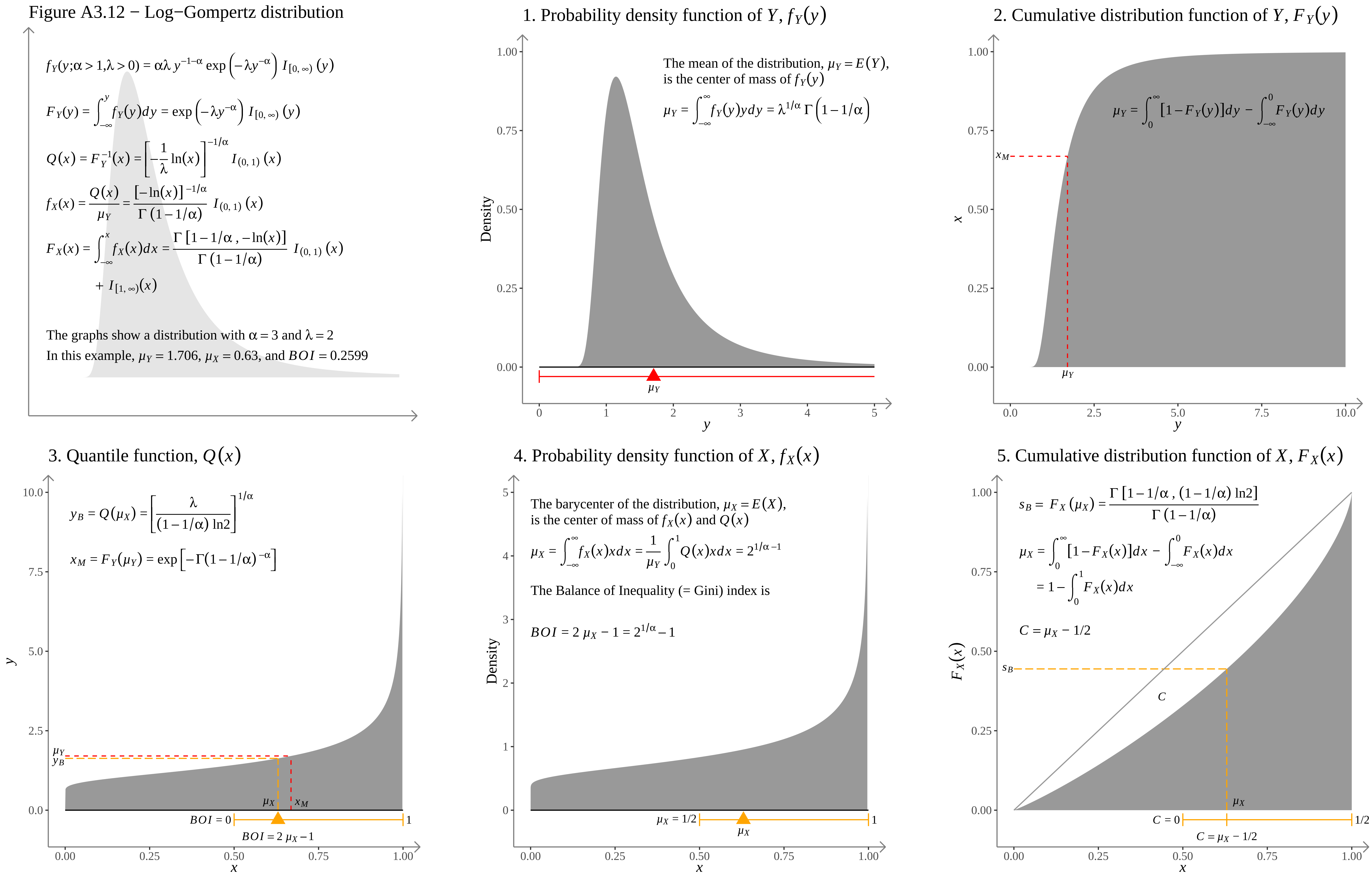


Figure A3.13 − Normal distribution symmetrically truncated

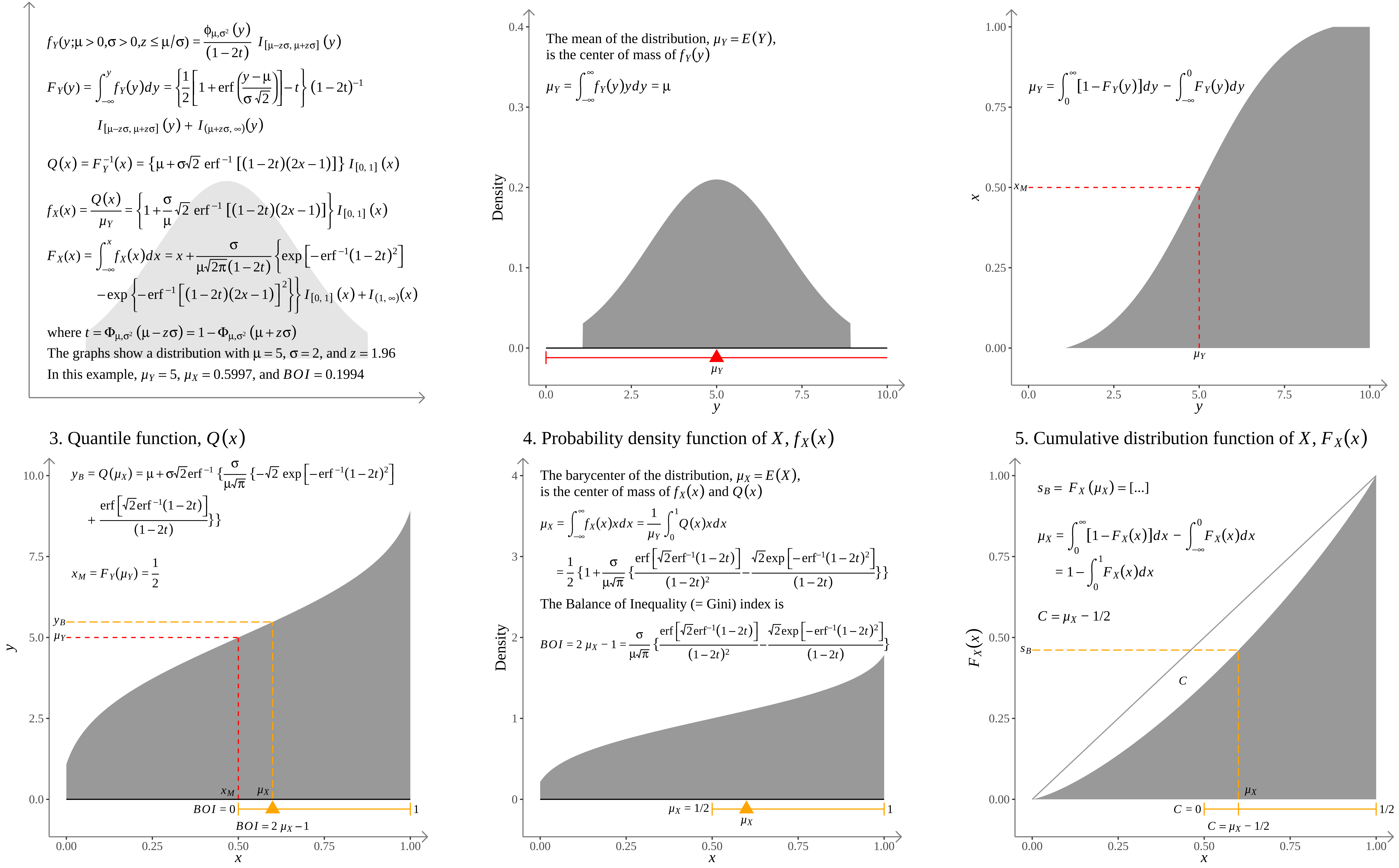


Figure A3.14 − Normal distribution symmetrically truncated 95%

$$f_Y(y;\mu>0,\frac{\mu}{\sigma}\geq 1.96) = \frac{\phi_{\mu,\sigma^2}(y)}{0.95}\, I_{[\mu-1.96\sigma,\, \mu+1.96\sigma]}(y)$$

$$F_Y(y) = \int_{-\infty}^{y} f_Y(y)dy = \frac{1}{0.95}\left\{\frac{1}{2}\left[1+\text{erf}\left(\frac{y-\mu}{\sigma\sqrt{2}}\right)\right]-0.025\right\}$$
$$I_{[\mu-1.96\sigma,\, \mu+1.96\sigma]}(y) + I_{(\mu+1.96\sigma,\, \infty)}(y)$$

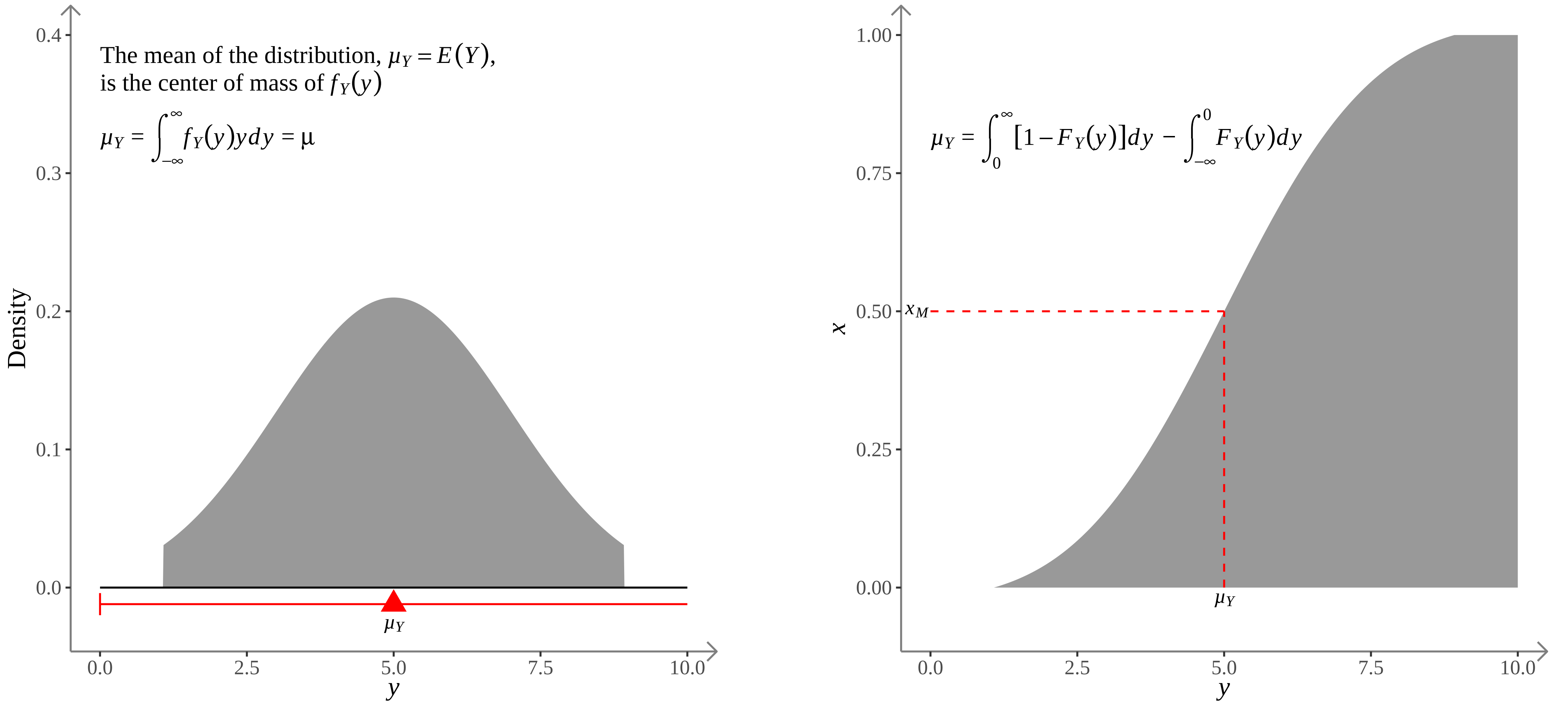
$$Q(x) = F_Y^{-1}(x) = \{\mu+\sigma\sqrt{2}\,\text{erf}^{-1}[0.95\,(2x-1)]\}\, I_{[0,1]}(x)$$

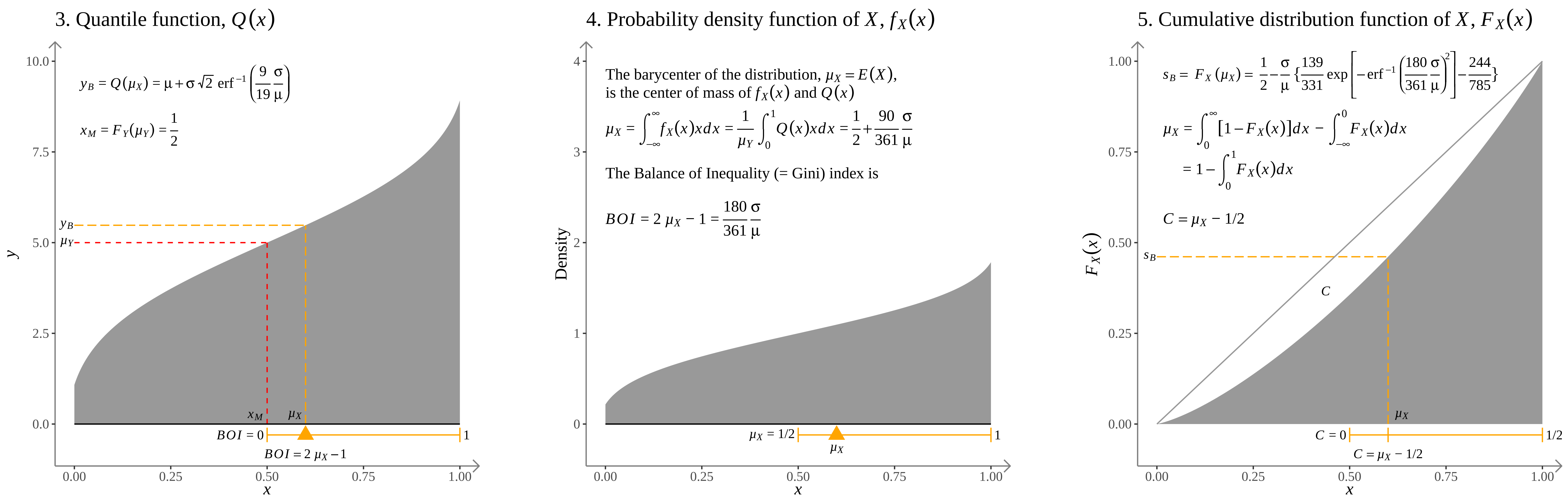
$$f_X(x) = \frac{Q(x)}{\mu_Y} = \left\{1+\frac{\sigma}{\mu}\sqrt{2}\,\text{erf}^{-1}[0.95\,(2x-1)]\right\} I_{[0,1]}(x)$$

$$F_X(x) = \int_{-\infty}^{x} f_X(x)dx = \left\{x-\frac{\sigma}{\mu}\left\{\frac{139}{331}\exp\left\{-\text{erf}^{-1}\left[0.95\,(2x-1)\right]^2\right\}\right.\right.$$
$$\left.\left.-\frac{55}{894}\right\}\right\} I_{[0,1]}(x) + I_{(1,\infty)}(x)$$

The graphs show a distribution with $\mu = 5$ and $\sigma = 2$

In this example, $\mu_Y = 5$, $\mu_X = 0.5997$, and $BOI = 0.1994$





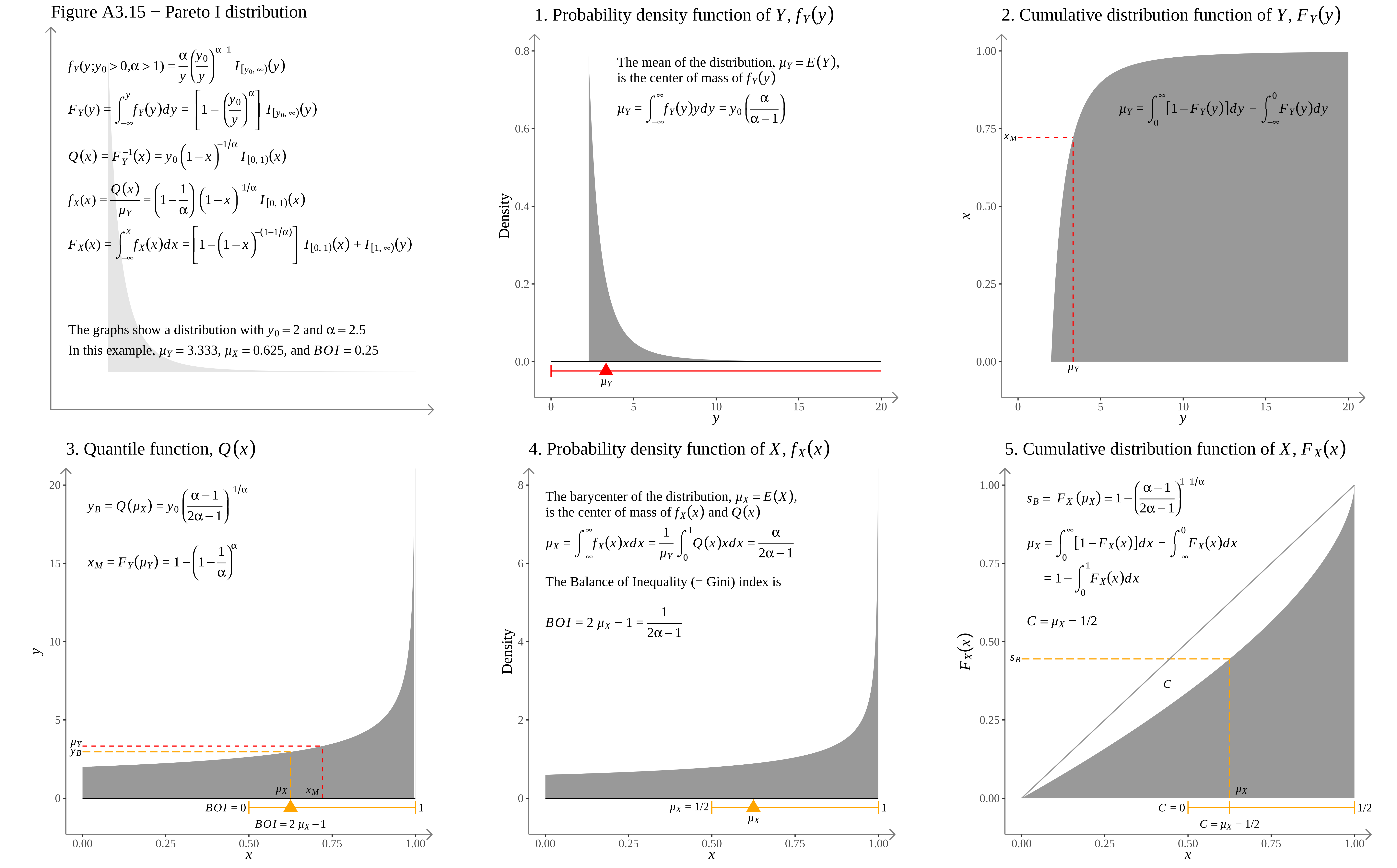


Figure A3.15 − Pareto I distribution

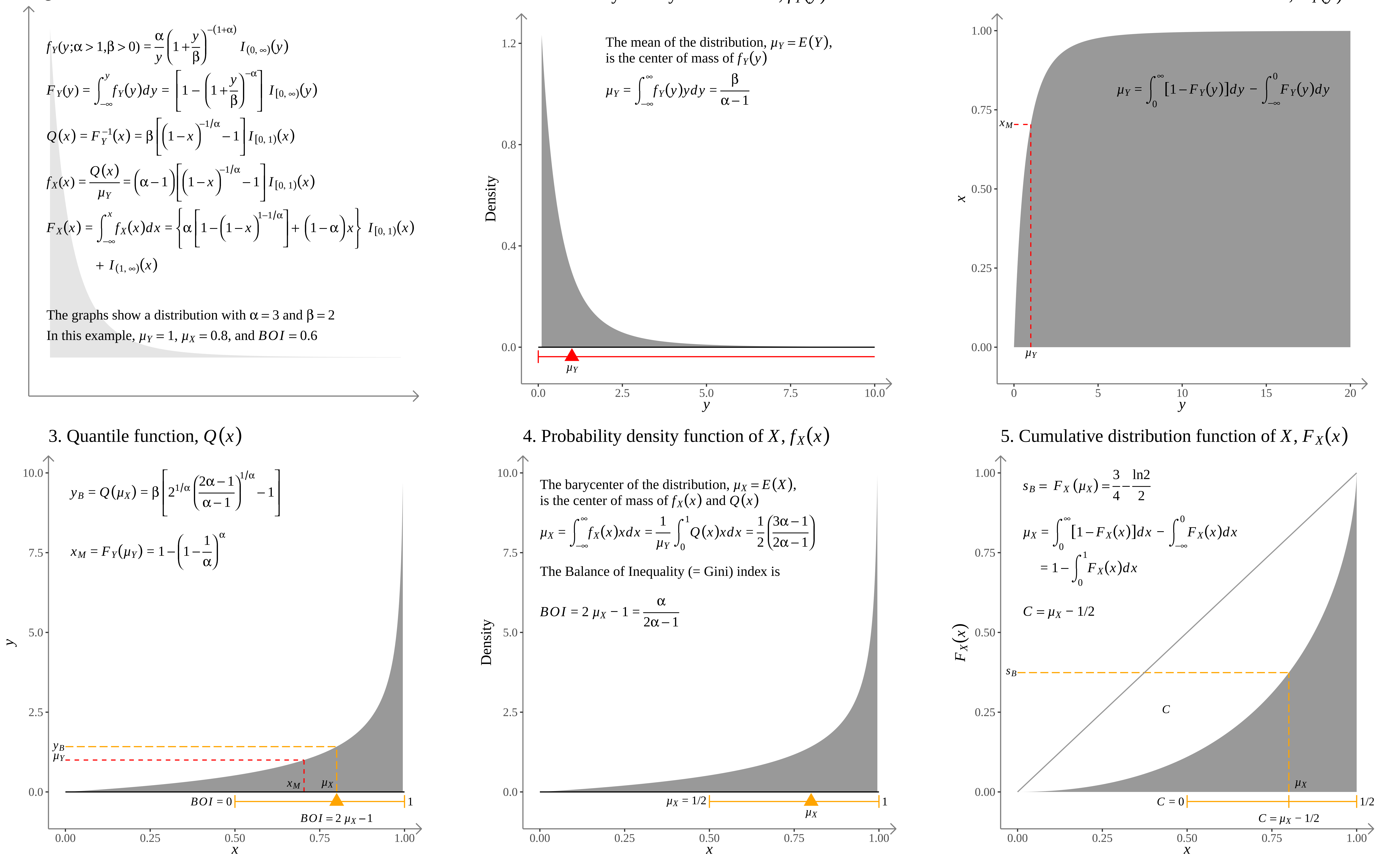
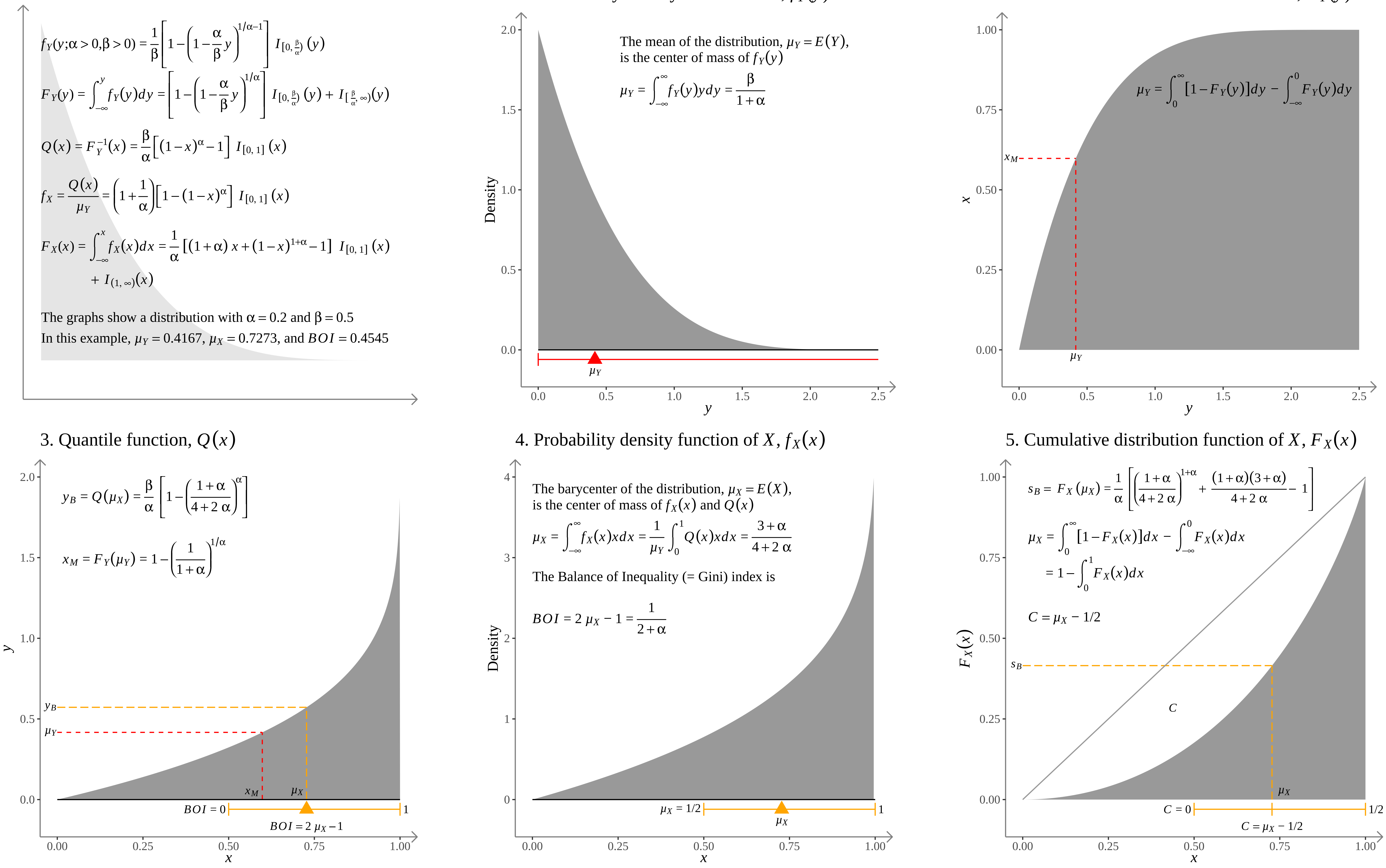


Figure A3.16 − Pareto II distribution

Figure A3.17 − Generalized Pareto distribution



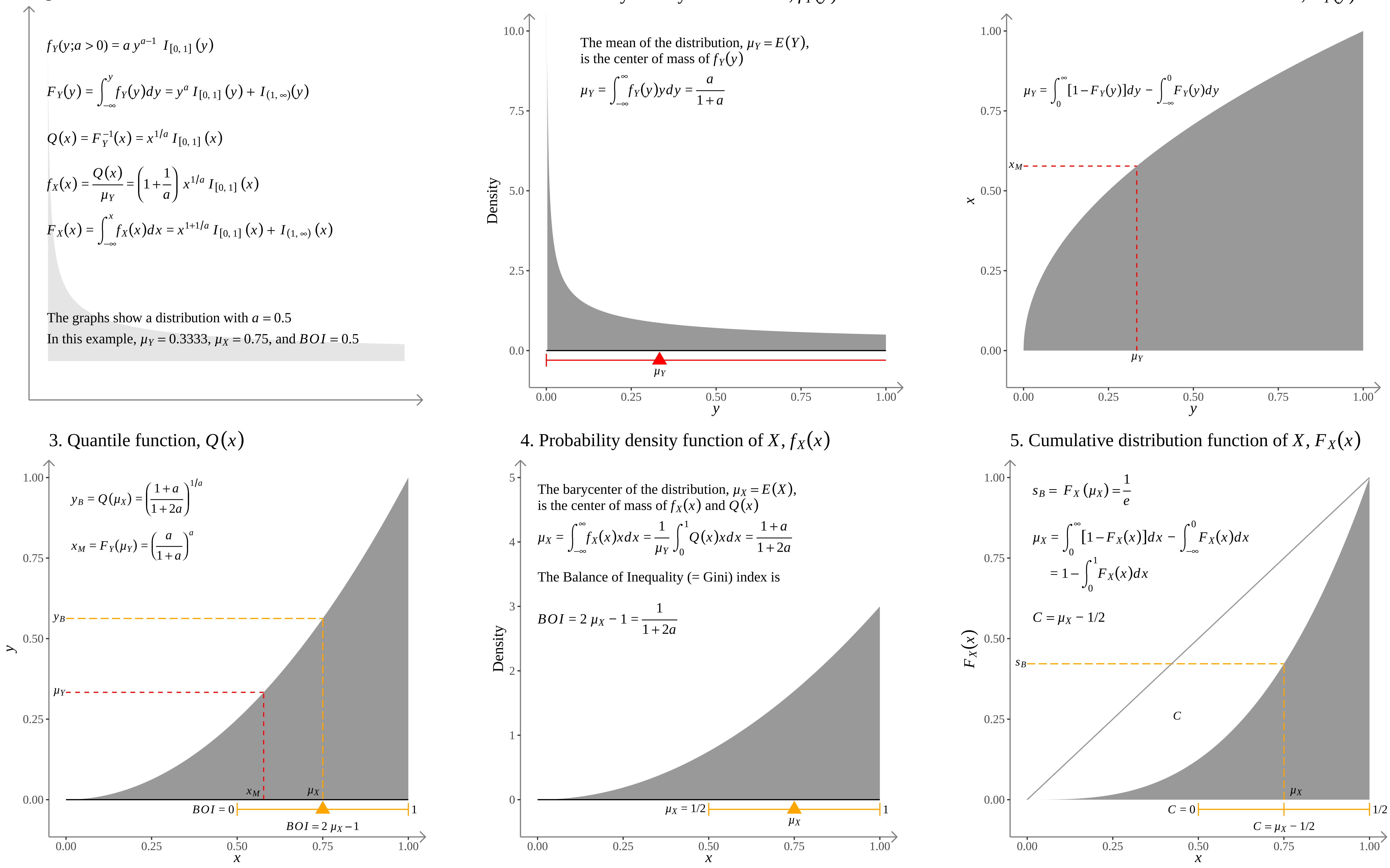


Figure A3.18 − Power function I distribution

Figure A3.19 − Power function II distribution

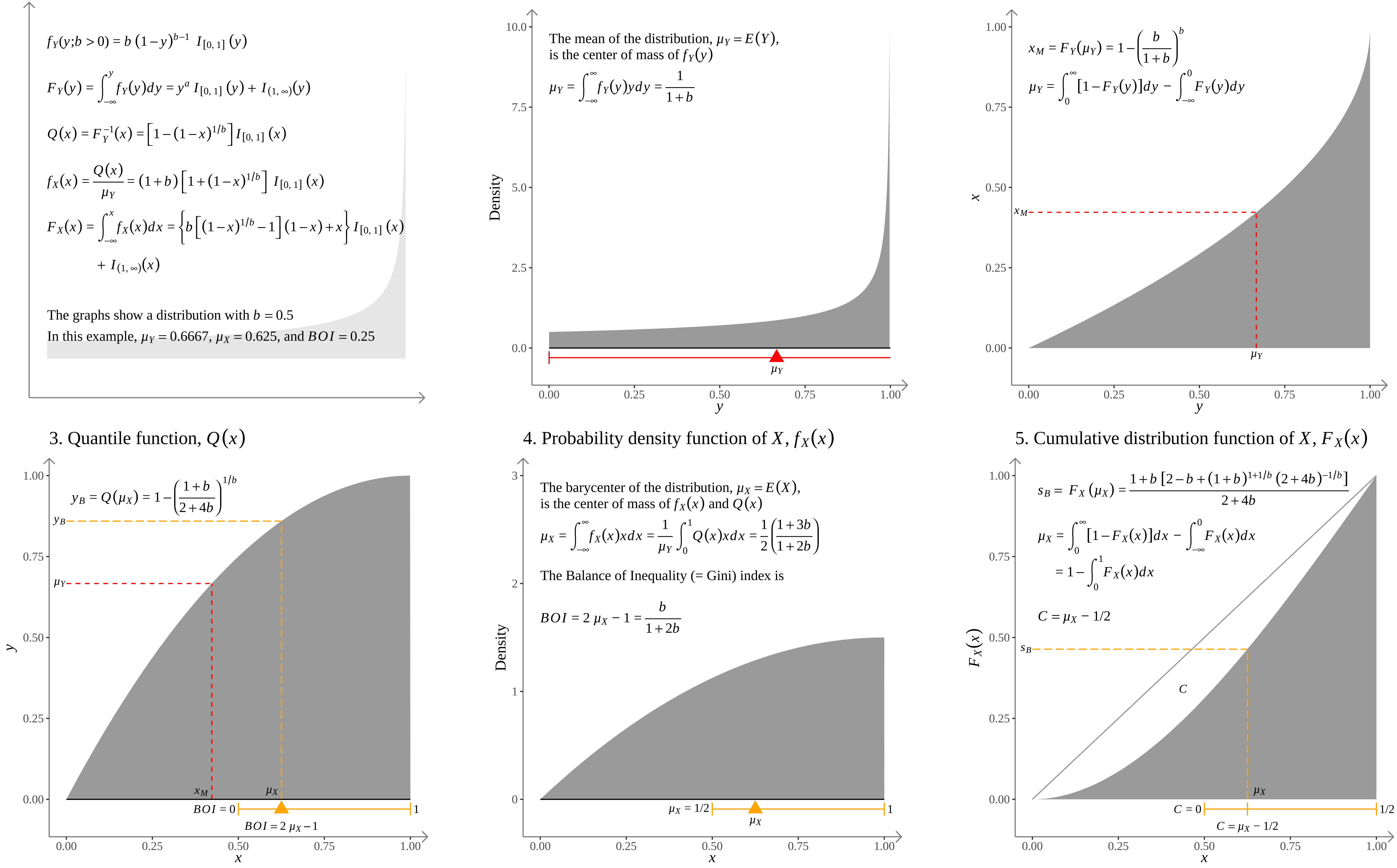


Figure A3.20 − Rayleigh distribution

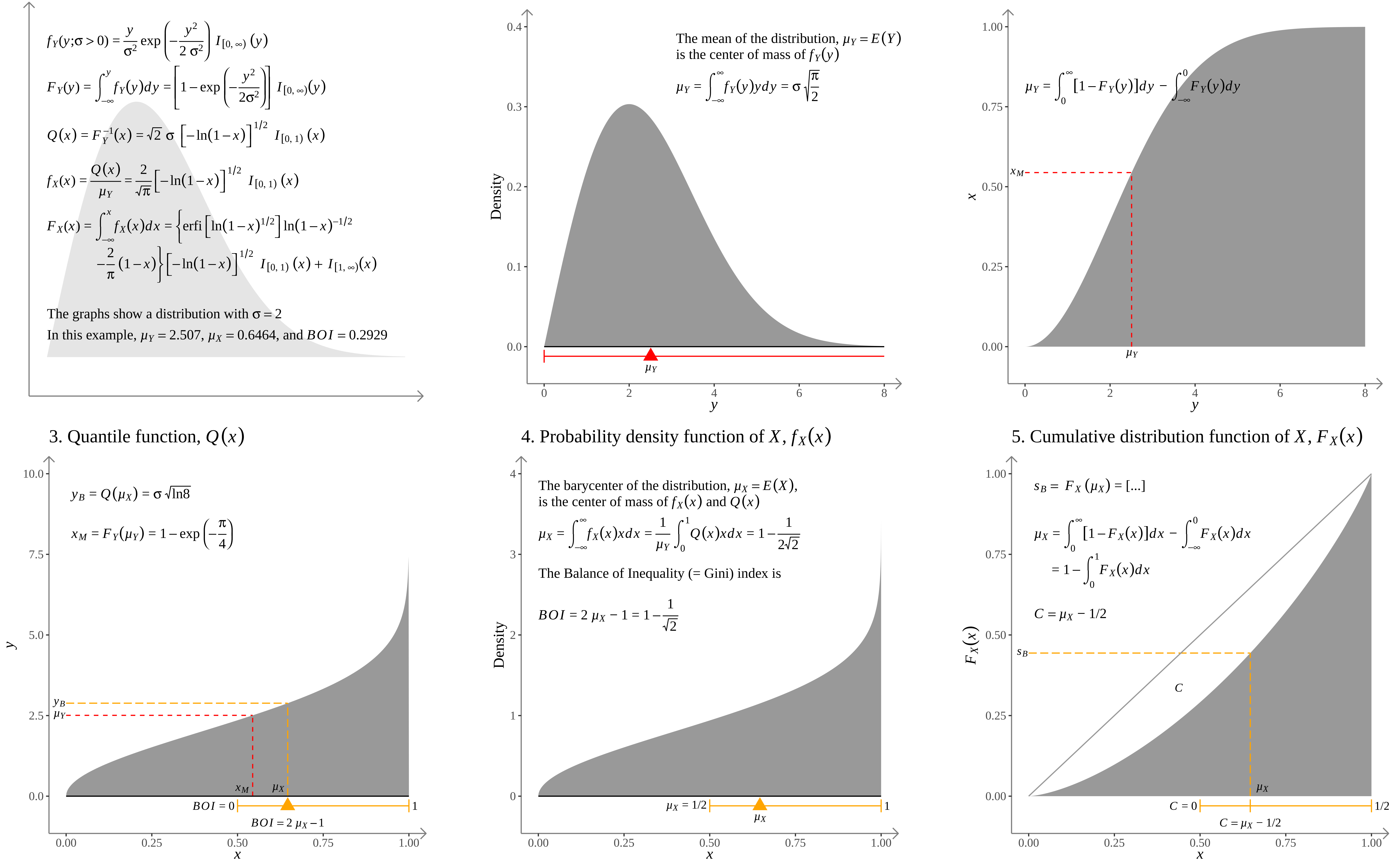
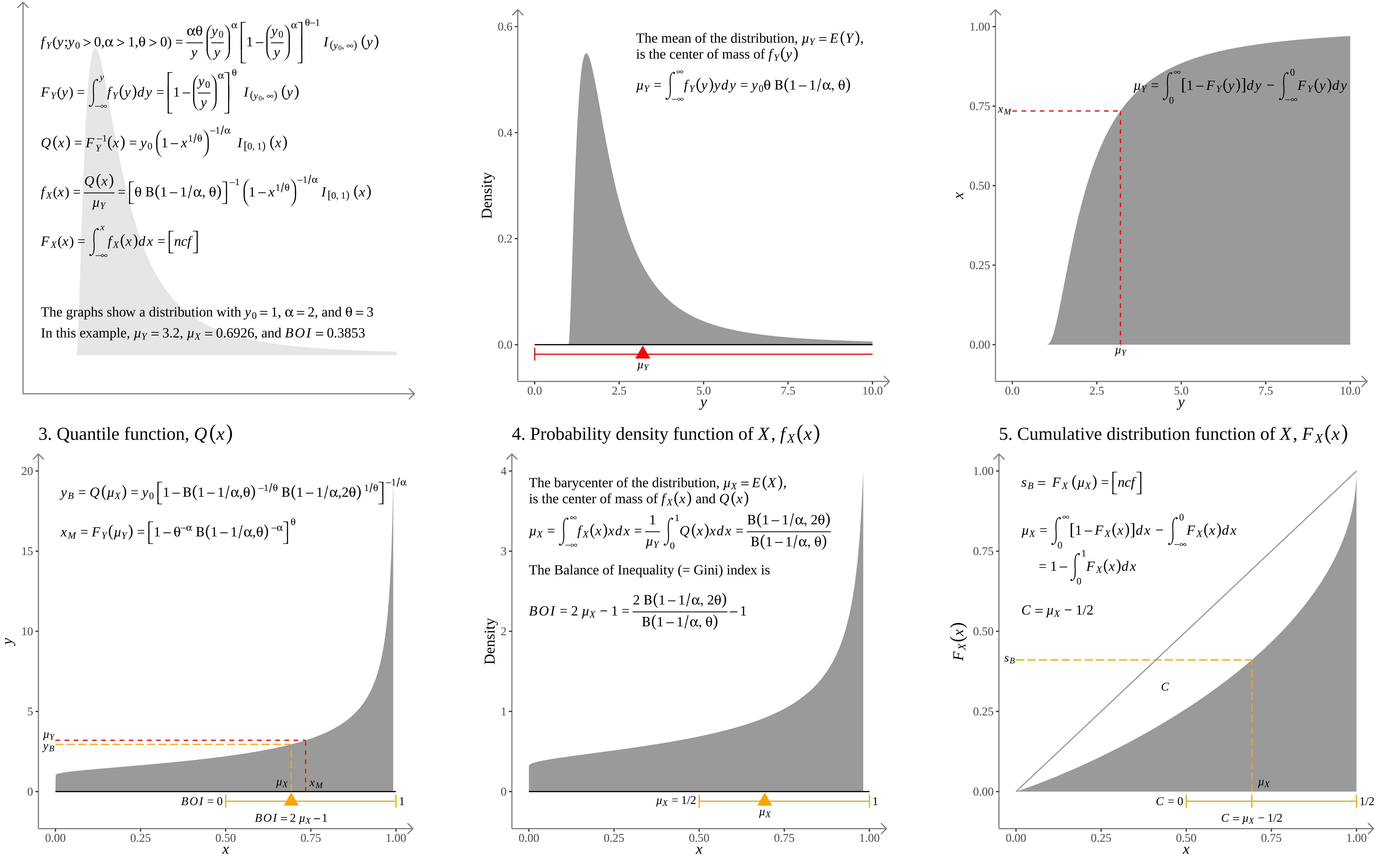


Figure A3.21 − Stoppa distribution



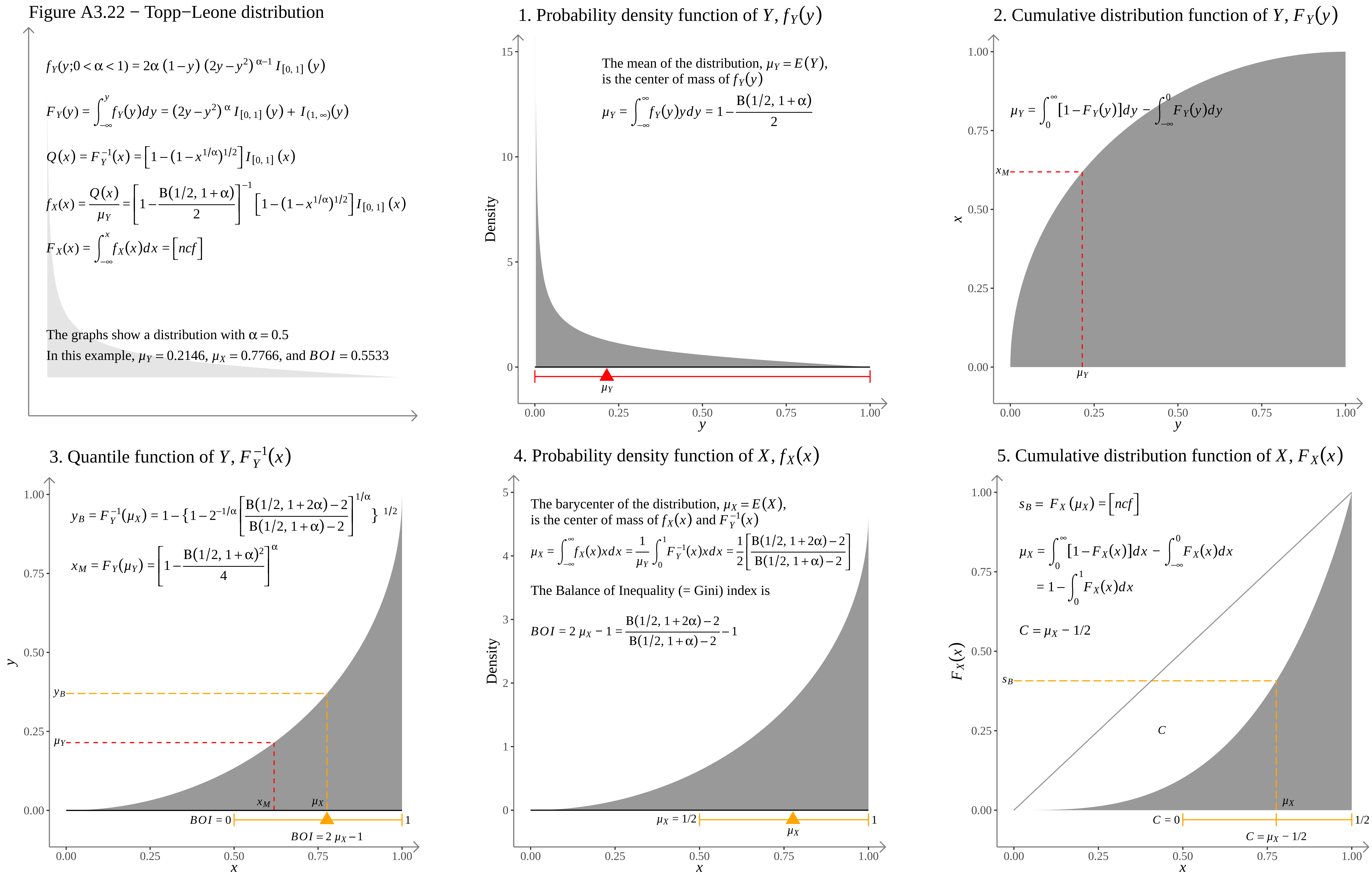


Figure A3.22 − Topp−Leone distribution

Figure A3.23 − Triangular distribution

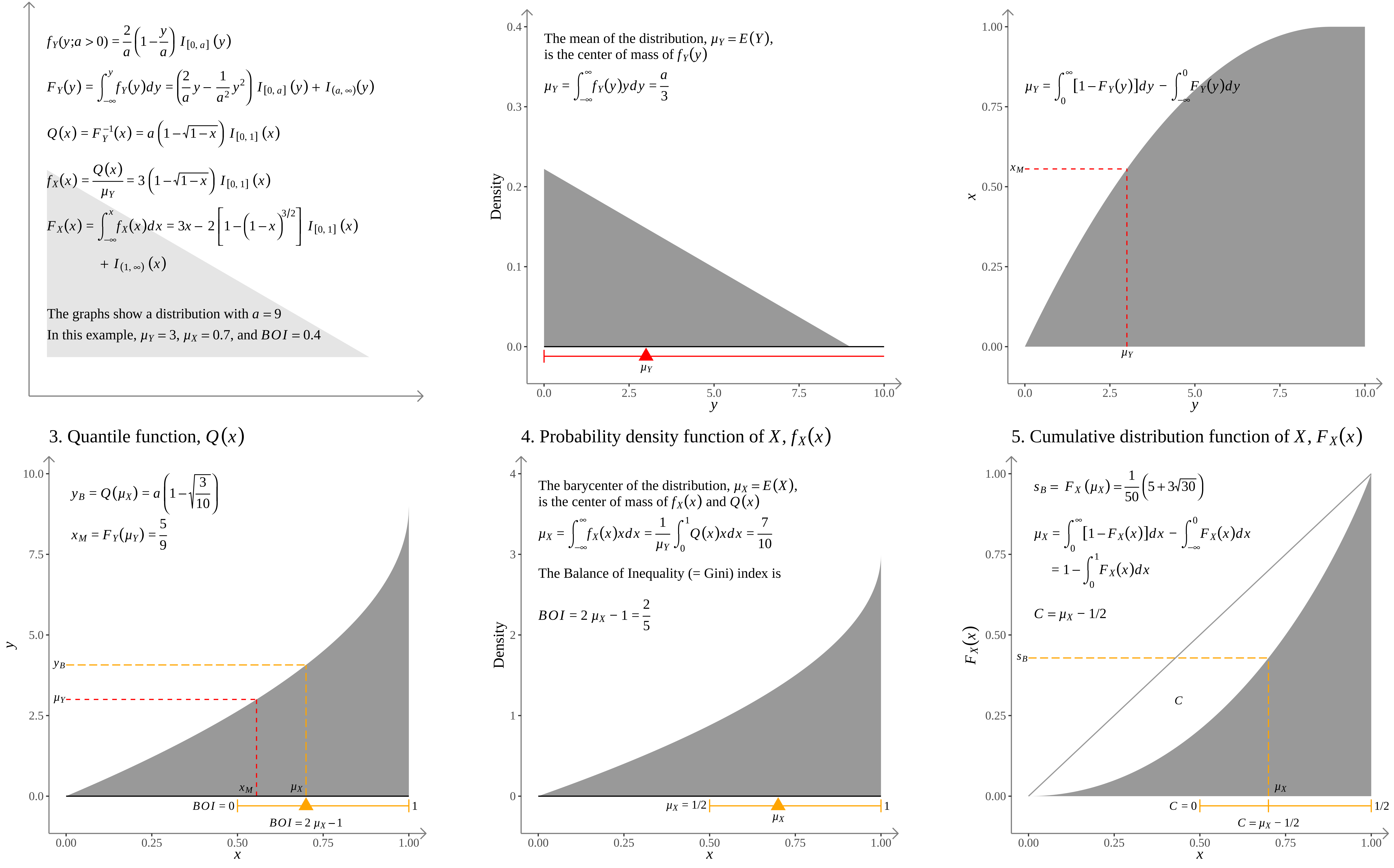


Figure A3.24 − Tukey Lambda I distribution

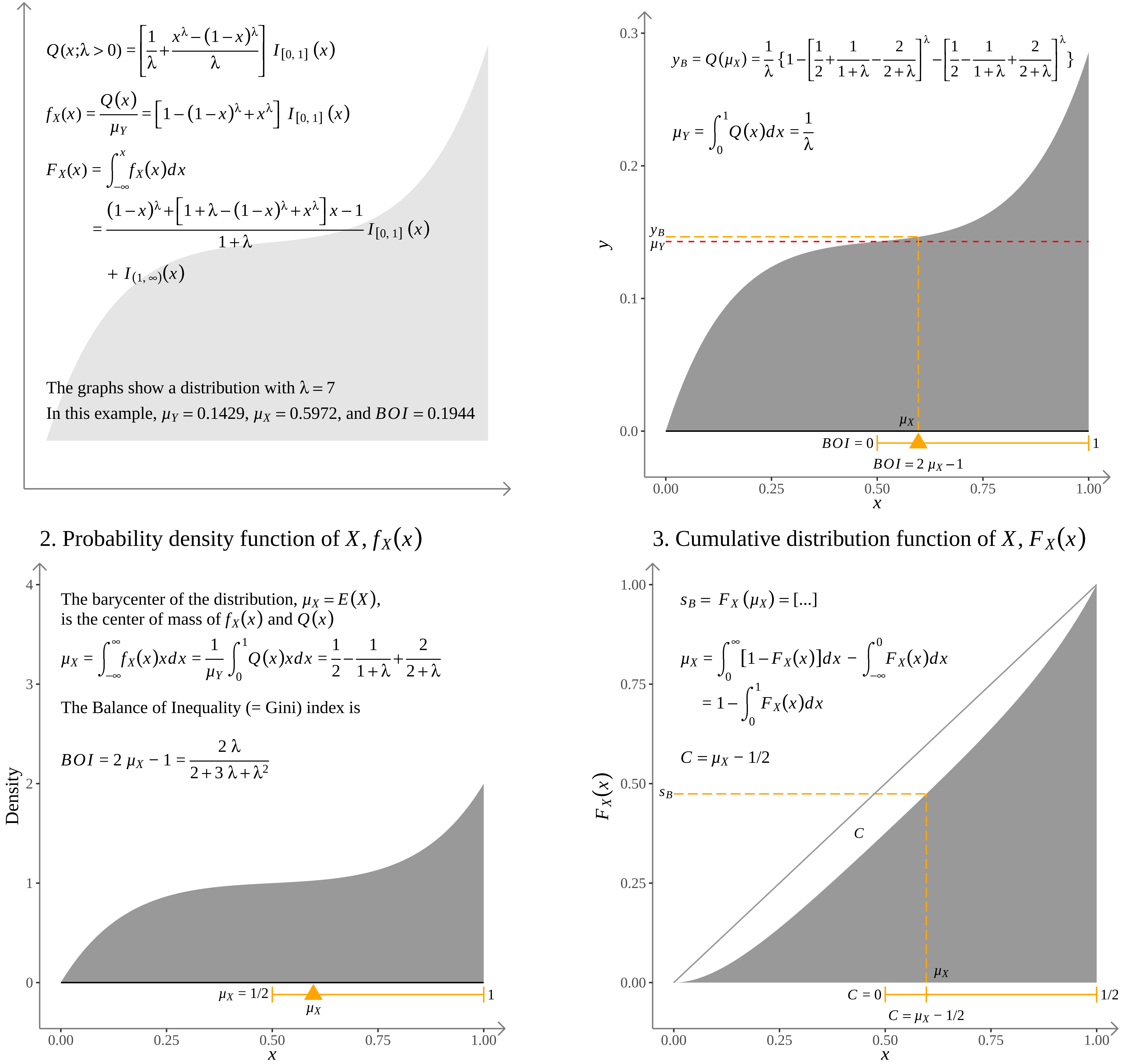


Figure A3.25 − Tukey Lambda III distribution

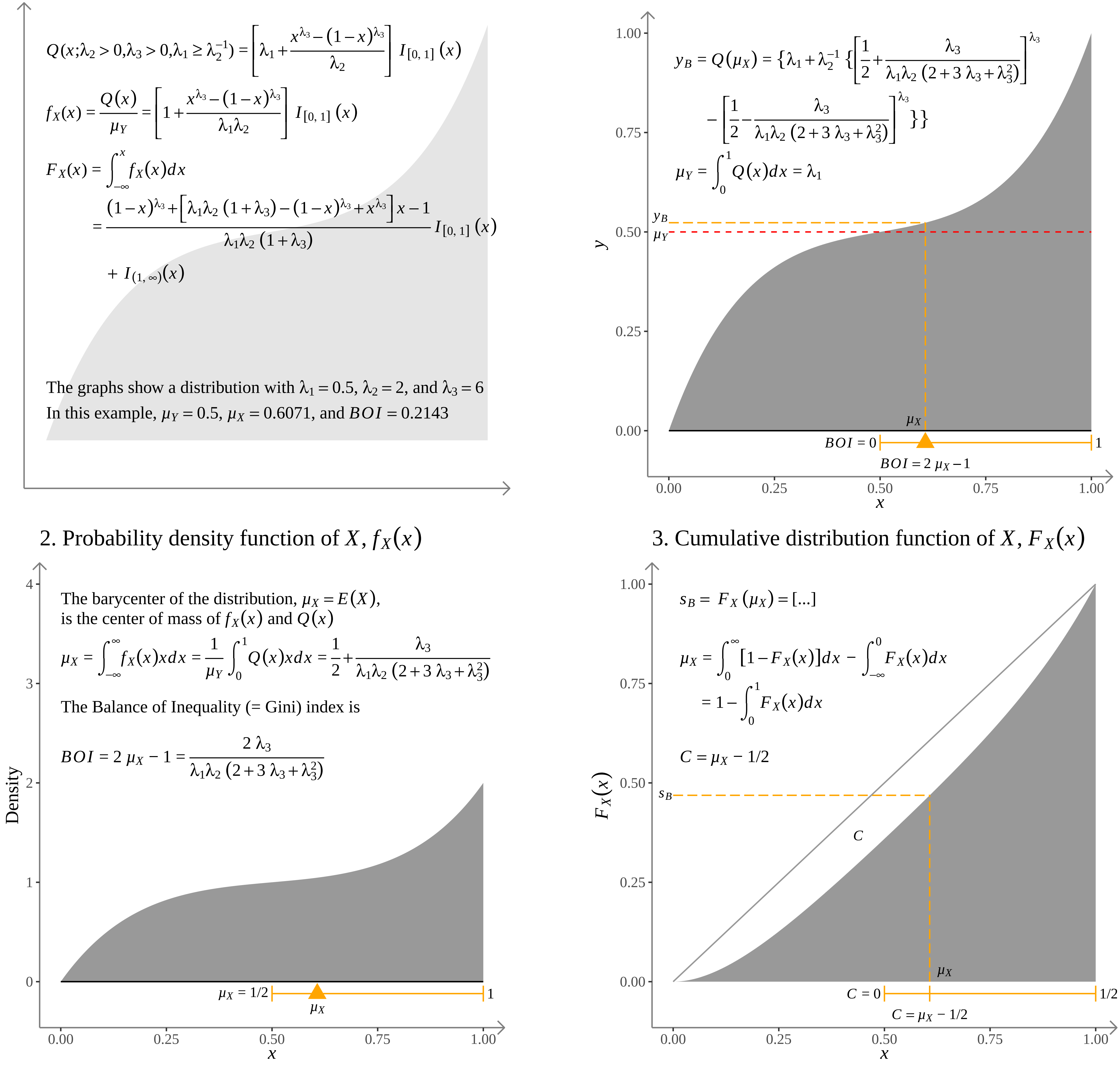


Figure A3.26 − Generalized Tukey Lambda distribution

$$Q(x;\lambda>0,\lambda_1>0,\lambda_2>0)=\lambda\left[1+x^{\lambda_1}-(1-x)^{\lambda_2}\right]\,I_{[0,\,1]}(x)$$

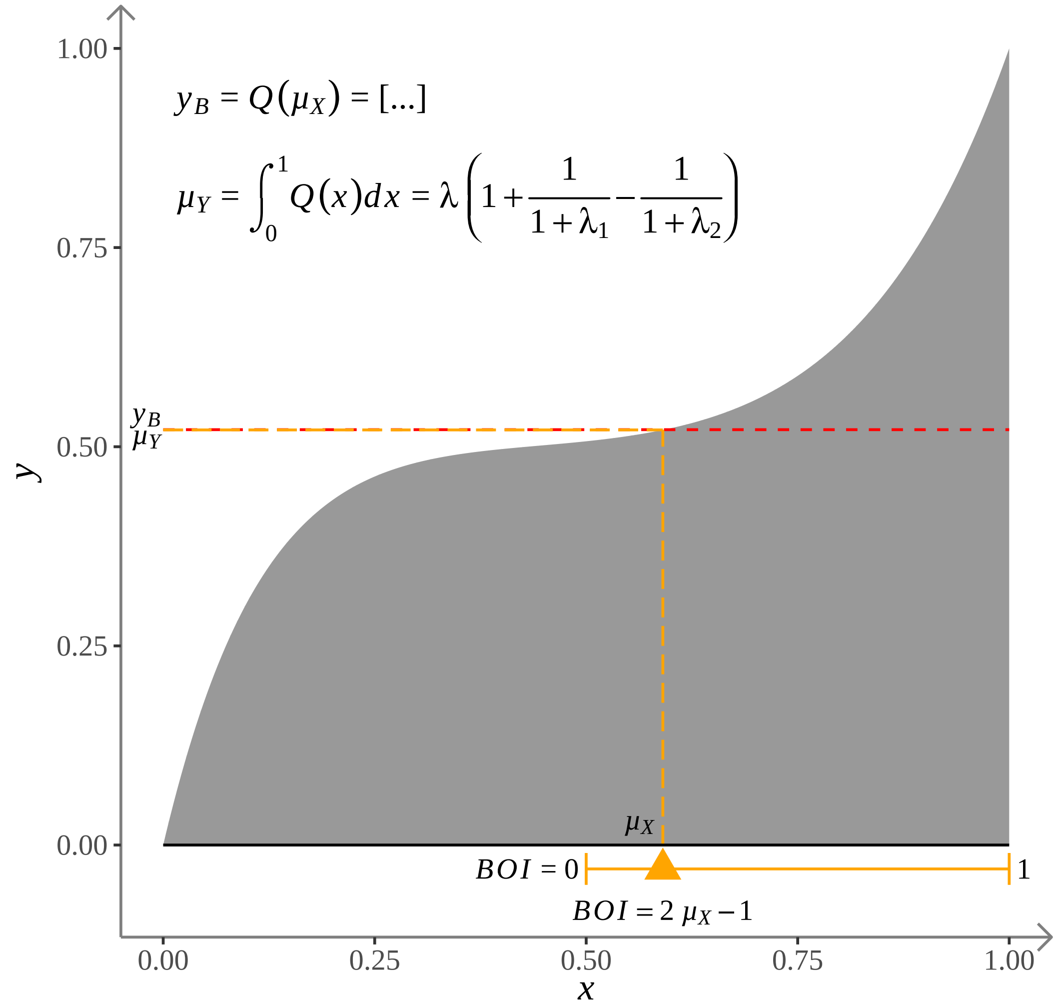
$$f_X(x)=\frac{Q(x)}{\mu_Y}=\frac{(1+\lambda_1)\,(1+\lambda_2)}{1+(2+\lambda_1)\,\lambda_2}\left[1+x^{\lambda_1}-(1-x)^{\lambda_2}\right]\,I_{[0,\,1]}(x)$$

$$F_X(x)=\int_{-\infty}^{x}f_X(x)dx=\frac{1}{1+(2+\lambda_1)\,\lambda_2}\Big\{(1+\lambda_2)\,x^{1+\lambda_1}+(1+\lambda_1)\left[(1-x)^{1+\lambda_2}+(1+\lambda_2)\,x-1\right]\Big\}\,I_{[0,\,1]}(x)+I_{(1,\,\infty)}(x)$$

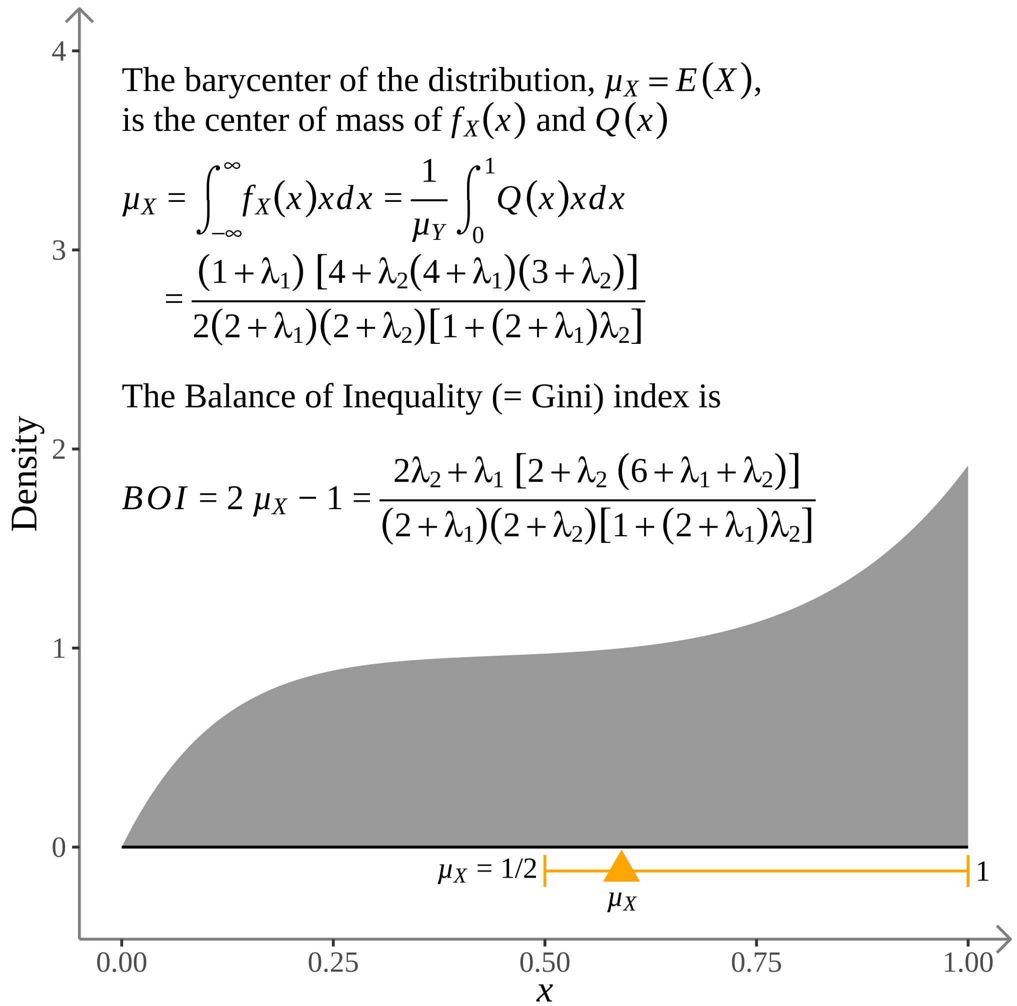
The graphs show a distribution with $\lambda=0.5$, $\lambda_1=6$, and $\lambda_2=9$

In this example, $\mu_Y=0.5214$, $\mu_X=0.5906$, and $BOI=0.1812$

1. Quantile function, $Q(x)$



2. Probability density function of X, $f_X(x)$



3. Cumulative distribution function of X, $F_X(x)$

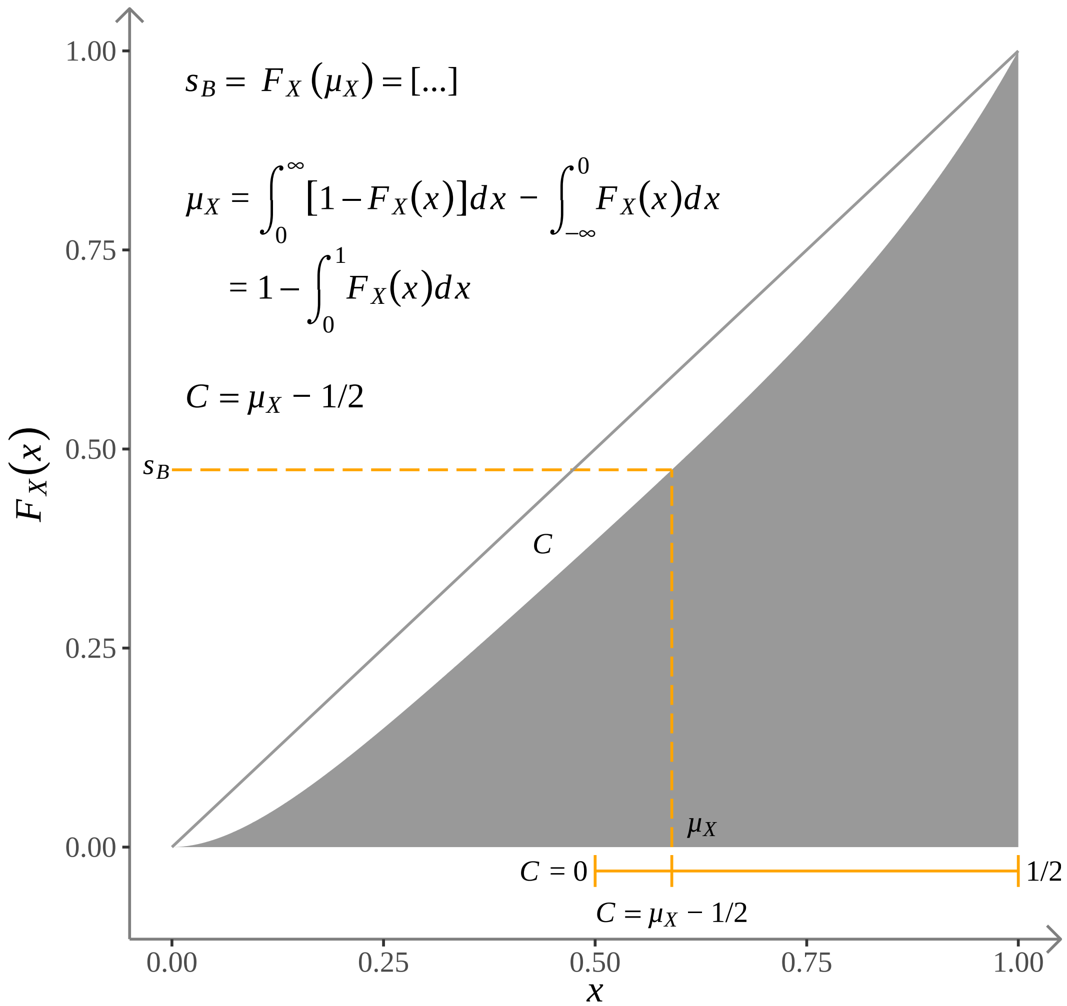


Figure A3.27 − Uniform distribution

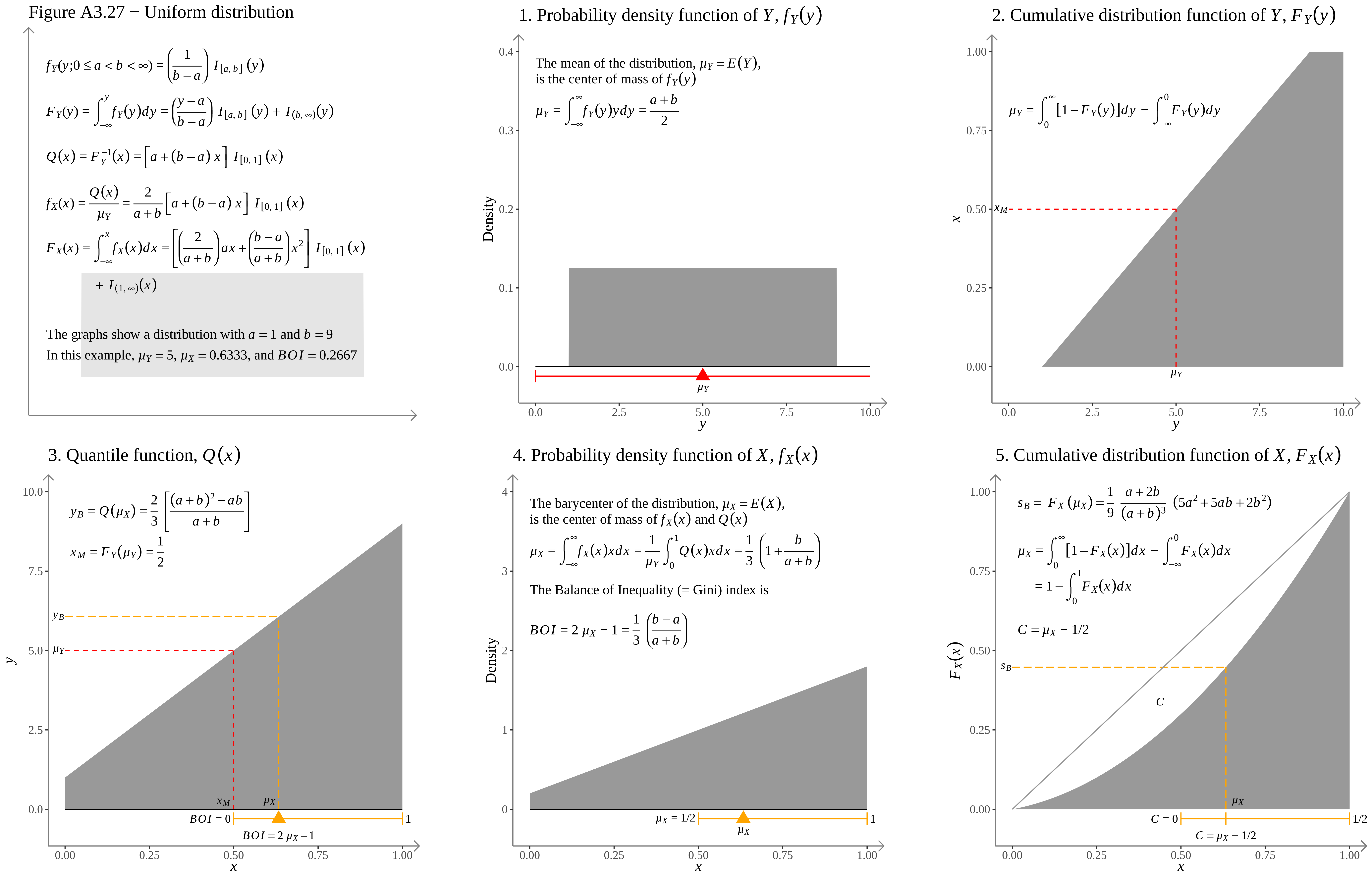


Figure A3.28 − U−Quadratic distribution

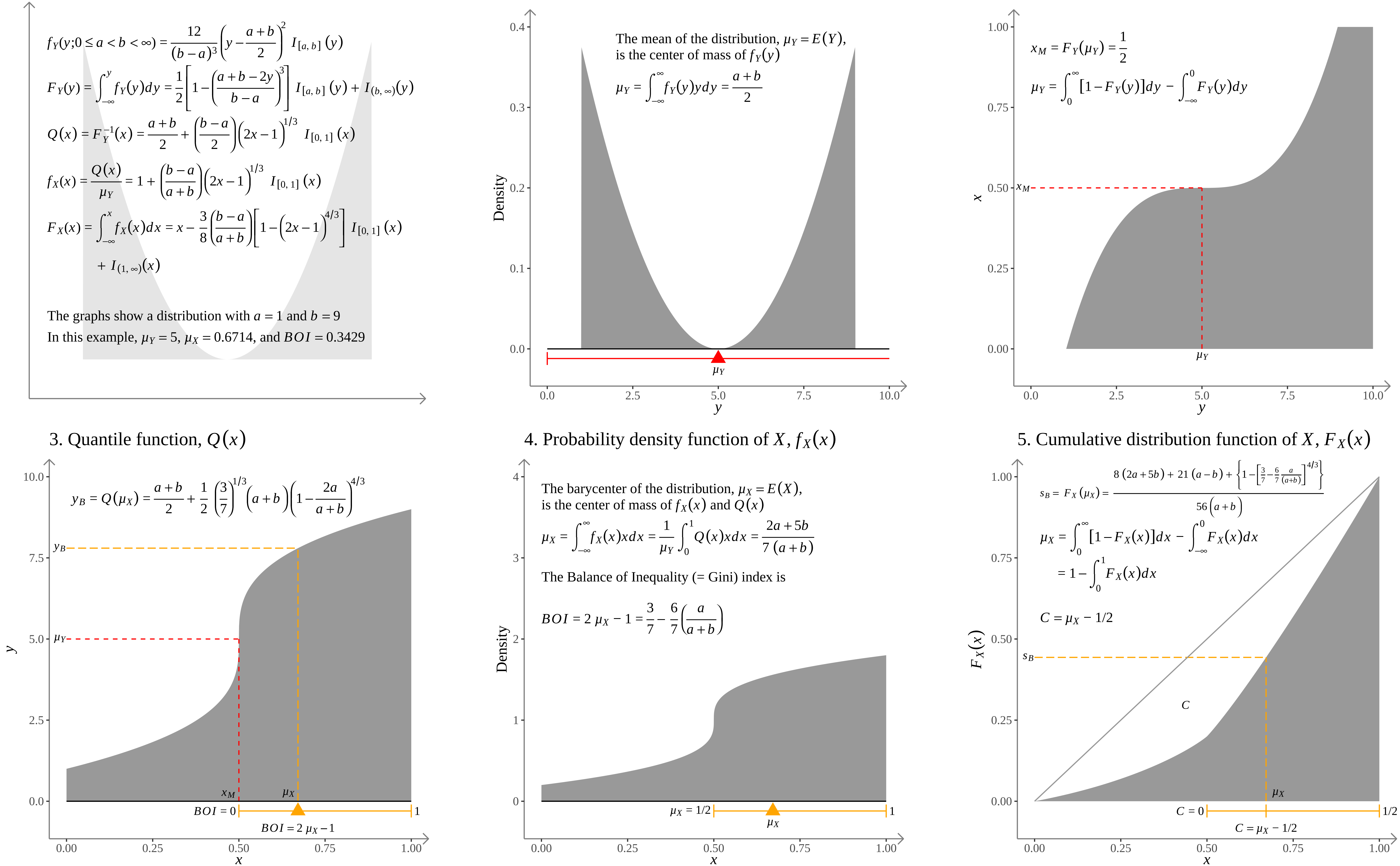
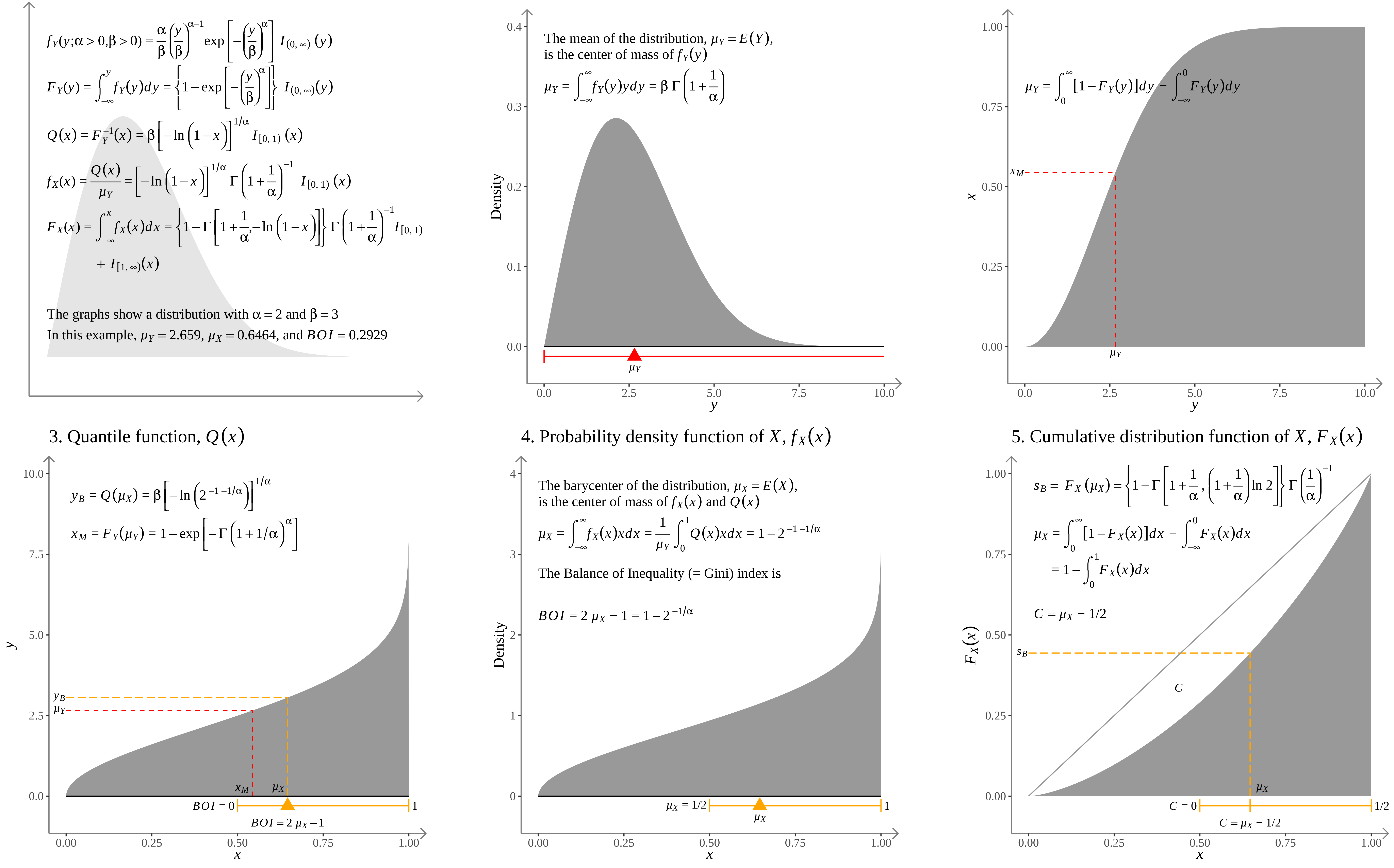


Figure A3.29 − Weibull distribution



APPENDIX 4 EMPIRICAL APPLICATION: INCOME INEQUALITY IN LIS COUNTRIES

Table A4.1 – Empirical application: Income inequality in LIS countries

(1) Country	(2) Year	(3) Number of observations: Total	(4) Number of observations: Negative values (Not available)	(5) Population size: Before rounding personal weights	(6) Population size: After rounding personal weights	(7) Change	(8) Rounded personal weights: Minimum value	(9) Rounded personal weights: Maximum value	(10) Barycenter	(11) Balance of Inequality (= Gini) index
India	2011	204,365	- (-)	1,211,296,909	1,211,299,163	2,254	154	156,648	0.9451	0.8903
Egypt	2012	49,174	- (13)	80,044,895	80,046,346	1,451	93	18,950	0.9370	0.8741
South Africa	2015	37,975	- (-)	54,948,875	54,948,709	-166	7	64,709	0.9345	0.8690
Palestine	2017	20,175	- (-)	4,733,357	4,733,357	-	13	1,249	0.9294	0.8589
Guatemala	2014	54,802	- (24)	15,993,629	15,993,691	62	13	2,398	0.9089	0.8178
Viet Nam	2013	23,583	- (-)	91,895,642	91,895,642	-	321	27,736	0.9058	0.8112
Paraguay	2016	37,713	- (-)	6,754,408	6,754,408	-	10	1,177	0.8874	0.7748
Mexico	2016	257,658	1,950 (-)	122,643,890	122,643,890	-	7	5,386	0.8859	0.7718
Peru	2016	130,526	979 (98)	31,893,611	31,892,968	-643	2	1,463	0.8830	0.7661
Israel	2016	29,739	14 (-)	8,201,662	8,201,795	133	9	1,603	0.8766	0.7532
Germany	1974	135,088	- (1,223)	62,101,366	62,096,565	-4,801	117	4,791	0.8765	0.7529
Panama	2016	42,168	- (48)	4,026,826	4,026,826	-	3	1,029	0.8757	0.7514
Georgia	2016	9,267	- (85)	3,643,149	3,643,197	48	96	1,651	0.8657	0.7314
United States of America	1974	34,244	- (932)	210,124,530	210,124,530	-	100	27,781	0.8602	0.7205
Brazil	2016	447,122	- (-)	204,412,569	204,412,458	-111	4	11,125	0.8531	0.7061
China	2013	61,162	1 (260)	1,355,286,257	1,355,284,887	-1,370	4,562	339,169	0.8476	0.6952
Italy	1986	25,068	- (234)	62,715,920	62,716,555	635	310	39,962	0.8429	0.6858
Taiwan	2016	50,569	- (-)	25,962,025	25,962,371	346	121	815	0.8407	0.6814
Japan	2013	7,276	- (-)	127,103,390	127,103,354	-36	3,265	579,565	0.8391	0.6781
United Kingdom	1974	18,974	- (5,160)	56,224,001	56,219,962	-4,039	2,963	2,963	0.8376	0.6753
United States of America	2016	185,412	159 (-)	319,310,408	319,310,186	-222	84	11,310	0.8301	0.6602
Ireland	2016	12,612	- (-)	4,802,277	4,802,240	-37	24	6,353	0.8255	0.6511
Uruguay	2016	118,568	- (745)	3,478,072	3,478,072	-	9	46	0.8197	0.6393
Australia	2010	42,595	157 (64)	21,472,970	21,472,860	-110	2	2,875	0.8135	0.6271
Serbia	2016	17,893	- (14)	6,755,343	6,755,137	-206	190	2,635	0.8056	0.6113
United Kingdom	2016	44,145	15 (-)	64,421,005	64,421,005	-	221	39,675	0.8055	0.6110
Italy	2016	16,464	2 (-)	60,243,342	60,243,440	98	492	20,161	0.8038	0.6077
Spain	2016	34,911	77 (-)	46,038,417	46,038,316	-101	34	18,528	0.8037	0.6074
Greece	2016	54,041	44 (234)	10,634,925	10,634,731	-194	8	3,910	0.8027	0.6053
Canada	2016	62,149	164 (-)	35,158,296	35,158,332	36	10	7,801	0.8003	0.6006

(1) Country	(2) Year	(3) Number of observations	(4)	(5) Population size	(6)	(7)	(8) Rounded personal weights	(9)	(10) Barycenter	(11) Balance of Inequality (= Gini) index
		Total	Negative values (Not available)	Before rounding personal weights	After rounding personal weights	Change	Minimum value	Maximum value		
Sweden	1975	29,277	- (299)	8,207,622	8,208,797	1,175	30	2,642	0.7937	0.5874
Lithuania	2016	11,136	- (9)	2,847,904	2,847,934	30	4	3,471	0.7936	0.5872
Poland	2016	99,230	- (-)	38,003,531	38,002,971	-560	0	1,299	0.7922	0.5844
Netherlands	2016	29,716	79 (-)	16,836,205	16,836,008	-197	0	10,695	0.7889	0.5776
France	2010	41,221	47 (-)	63,715,341	63,715,308	-33	1	8,893	0.7874	0.5747
Luxembourg	2013	9,982	16 (6)	507,498	507,610	112	3	335	0.7866	0.5731
Germany	2016	45,731	- (-)	82,205,143	82,205,141	-2	0	30,742	0.7829	0.5659
Estonia	2016	15,320	13 (-)	1,301,903	1,301,915	12	4	1,190	0.7804	0.5608
Russia	2016	367,080	- (-)	144,690,376	144,690,111	-265	2	58,058	0.7803	0.5606
Belgium	2016	14,028	6 (-)	11,162,308	11,162,300	-8	69	4,872	0.7791	0.5582
Austria	2016	12,876	- (-)	8,640,974	8,640,955	-19	115	4,761	0.7771	0.5542
Norway	2016	522,940	13,304 (-)	5,196,919	5,196,919	-	1	10	0.7742	0.5484
Iceland	2010	8,855	- (-)	300,766	300,562	-204	9	115	0.7731	0.5462
Switzerland	2016	18,700	- (-)	8,280,847	8,280,892	45	21	4,133	0.7714	0.5427
Denmark	2016	187,596	350 (-)	5,733,361	5,815,476	82,115	31	31	0.7618	0.5236
Slovenia	2015	11,228	- (-)	2,025,509	2,025,601	92	63	722	0.7569	0.5139
Finland	2016	24,818	1 (-)	5,418,579	5,418,731	152	1	1,681	0.7562	0.5125
Sweden	2005	36,918	56 (-)	8,882,224	8,881,614	-610	2	520	0.7476	0.4951
Slovakia	2016	16,031	8 (67)	5,255,973	5,255,926	-47	13	1,691	0.7455	0.4911
Hungary	2015	6,237	- (-)	9,897,541	9,897,739	198	68	14,069	0.7413	0.4826

Notes. This table shows an empirical application of the methodology for the estimation of the distributions' barycenter and Balance of Inequality (=Gini) index by using weighted observations introduced in Section 7. For each country listed in Col. (1), Col. (10) shows the estimation of the population's barycenter of the total individual income distribution, and Col. (11) shows the corresponding Balance of Inequality (=Gini) index with reference to the year shown in Col. (2). Col. (3) and (4) respectively show the total number of observations for each country and the number of negative (and missing) values. Personal weights are rounded to the nearest integer. Col. (5) and (6) show the estimated population size before and after rounding, respectively, and Col. (7) shows the change in the estimated population size. Col. (8) and (9) show the minimum and maximum personal weight after rounding. The estimates are made by using the total personal income variable (*pitotal*) in the Luxembourg Income Study (LIS) Database provided by the LIS Cross-National Data Center.