

Chapter 2

Classical Averaging Functions

Abstract This chapter presents the classical means, starting with the weighted arithmetic and power means, and then continuing to the quasi-arithmetic means. The topics of generating functions, comparability and weights selection are covered. Several interesting classes of non-quasi-arithmetic means are presented, including Gini, Bonferroni, logarithmic and Bajraktarevic means. Methods of extension of symmetric bivariate means to the multivariate case are also discussed.

2.1 Semantics

Averaging is the most common way to combine the inputs. It is commonly used in voting, multicriteria and group decision making, constructing various performance scores, statistical analysis, etc. The basic rule is that the total score cannot be above or below any of the inputs. The aggregated value is seen as some sort of representative value of all the inputs.

We shall adopt the following generic definition [BPC07, Bul03, CKKM02, DP85, DP04].

Definition 2.1 (*Averaging aggregation*) An aggregation function f is averaging if for every \mathbf{x} it is bounded by the minimum and the maximum

$$\min(\mathbf{x}) \leq f(\mathbf{x}) \leq \max(\mathbf{x}).$$

We remind that due to monotonicity of aggregation functions, averaging functions are idempotent, and vice versa, see Note 1.12, p. 10. That is, an aggregation function f is averaging if and only if $f(t, \dots, t) = t$ for any $t \in \mathbb{I}$.

Formally, the minimum and maximum functions can be considered as averaging, however they are the limiting cases, right on the border with conjunctive and disjunctive functions. There are also some types of mixed aggregation functions, such as uninorms or nullnorms, that include averaging functions as particular cases.

The term *mean* is often used synonymously with the average to denote a measure of central tendency of the data.¹ A prototypical example of means is the arithmetic mean, often referred to as “the” mean.

However, the concepts of other means like the geometric and harmonic means, as well as the mode and midrange were already known to the Greeks [Eis71, Rub71]. Attempts to axiomatise the means were made in the 1920–1930s. Chisini [Chi29] gave a definition of a mean with respect to a function of n arguments f as a number M such that if every number x_i is replaced by M we get an equality

$$f(x_1, \dots, x_n) = f(M, \dots, M).$$

This definition was later criticised (e.g., by Gini [Gin58]) because it neither implies that the solution to the above equation exists, nor that M is in the range of the arguments, one of the fundamental requirements stipulated by Cauchy [Cau21]. Gini [Gin58] proposes to define a mean of several quantities as some value obtained as a result of a certain mathematical procedure with the given quantities, which either coincides with one of those quantities or is bounded by the largest and the smallest quantity.

Gini [Gin58] separates the means into analytical and positional. The analytical means are the ones obtained by application of a certain formula that acts on the values of the arguments. In contrast, positional means are computed by taking into account the relative position of the arguments. The arithmetic mean is an example of analytical means whereas the median is a positional mean. This classification is rather fuzzy, as Gini states himself, because of significant overlap.

Another distinction made by Gini is between hard means (where the result depends on all the inputs) and “moving” means (where changes to some inputs may not affect the value of the mean). In our terms the hard means correspond to strictly increasing averages. The representative examples of both classes are again the arithmetic mean and the median.

2.1.1 Measure of Orness

The *measure of orness*, also called the *degree of orness* or *attitudinal character*, is an important numerical characteristic of averaging aggregation functions. It was first defined in 1974 by Dujmovic [Duj73, Duj74], and then rediscovered several times, see [FR94, Yag88], mainly in the context of OWA functions (Sect. 3.1). It is applicable to any averaging function (and even to some other aggregation functions, like ST-OWA [FSM03]).

¹In some languages there are no distinct terms that refer separately to the average and the mean, e.g. *moyenne* (Fr), *medie* (It.), *promedio* (or *media*) (Sp.), *sredniaya* (Ru.), *Gemiddelde* (Dut.), *Gjennomsnitt* (No.), whereas in others there are, e.g. *Durchschnitt* or *Mittelwerte* (Ger.). For etymology of the words *mean* and *average* see [Eis71]. We will use both terms synonymously.

Basically, the measure of orness measures how far a given averaging function is from the max function, which is the weakest disjunctive function. The measure of orness is computed for any averaging function [Duj74, FSM03, FR94] using

Definition 2.2 (*Measure of orness*) Let f be an averaging aggregation function. Then its measure of orness is

$$orness(f) = \frac{\int_{\mathbb{I}^n} f(\mathbf{x})d\mathbf{x} - \int_{\mathbb{I}^n} \min(\mathbf{x})d\mathbf{x}}{\int_{\mathbb{I}^n} \max(\mathbf{x})d\mathbf{x} - \int_{\mathbb{I}^n} \min(\mathbf{x})d\mathbf{x}}. \quad (2.1)$$

Clearly, $orness(\max) = 1$ and $orness(\min) = 0$, and for any f , $orness(f) \in [0, 1]$. The calculation of the integrals of max and min functions was performed in [Duj73] and results in simple equations

$$\int_{[0,1]^n} \max(\mathbf{x})d\mathbf{x} = \frac{n}{n+1} \text{ and } \int_{[0,1]^n} \min(\mathbf{x})d\mathbf{x} = \frac{1}{n+1}. \quad (2.2)$$

Hence (since both max and min are homogeneous functions) we get

$$\int_{[0,b]^n} \max(\mathbf{x})d\mathbf{x} = b^{n+1} \frac{n}{n+1} \text{ and } \int_{[0,b]^n} \min(\mathbf{x})d\mathbf{x} = b^{n+1} \frac{1}{n+1}, \quad (2.3)$$

and then for $a, b \geq 0$:

$$\begin{aligned} \int_{[a,b]^n} \max(\mathbf{x})d\mathbf{x} &= a(b-a)^n + \frac{(b-a)^{n+1}n}{n+1} = (b-a)^n \frac{a+bn}{n+1} \text{ and} \\ \int_{[a,b]^n} \min(\mathbf{x})d\mathbf{x} &= a(b-a)^n + \frac{(b-a)^{n+1}}{n+1} = (b-a)^n \frac{an+b}{n+1} \end{aligned} \quad (2.4)$$

A different measure of orness, the average orness value, is proposed in [FSM03].

Definition 2.3 (*Average orness value*) Let f be an averaging aggregation function. Then its average orness value is

$$\overline{orness}(f) = \int_{\mathbb{I}^n} \frac{f(\mathbf{x}) - \min(\mathbf{x})}{\max(\mathbf{x}) - \min(\mathbf{x})} d\mathbf{x}. \quad (2.5)$$

Both the measure of orness and the average orness value are $\frac{1}{2}$ for weighted arithmetic means, and later we will see that both quantities coincide for OWA functions. However computation of the average orness value for other averaging functions is more involved (typically performed by numerical methods), therefore we will use mainly the measure of orness in Definition 2.2.

There are some alternative definitions of the measure of orness aimed at simplifying calculations (especially for quasi-arithmetic means, [Liu06]); we will discuss them in the relevant sections. Recently the concept of the orness measure has been

revisited in [KSP14] in the context of OWA functions, we discuss the details in Sect. 3.2.1.

The measure of orness was used by Dujmovic [Duj07] to classify the averages into seven classes, including disjunction, hard and soft partial disjunction, neutral averaging, soft and hard partial conjunction, and conjunction. The mentioned classes are ordered according to the orness, starting from 1 (the max function), to some threshold value $\frac{1}{2} < \alpha < 1$, then between α and $\frac{1}{2}$, $\frac{1}{2}$ (neutrality), to some other threshold value $\frac{1}{2} > \beta > 0$, from β to 0, and then 0 (the min function).

The hard partial conjunctions model mandatory requirements (in the decision making domain), and soft partial conjunctions model desired but not mandatory requirements. The hard partial disjunction models sufficient requirements, and soft partial disjunction models desired but not sufficient requirements.

2.2 Classical Means

While means are treated synonymously with averaging functions, the classical treatment of means (see, e.g., [Bul03]) excludes certain types of averaging functions, which have been developed quite recently, in particular ordered weighted averaging and various integrals. On the other hand some classical means (e.g., some Gini means) lack monotonicity, and therefore are not aggregation functions. Some of these averages will be considered in Chap. 7. In this chapter we will concentrate on various classical means, and present other types of averaging, or mean-type functions in separate chapters.

2.2.1 Arithmetic Mean

The arithmetic mean is the most widely used averaging function.

Definition 2.4 (*Arithmetic mean*) The arithmetic mean, or the average of n values, is the function

$$M(\mathbf{x}) = \frac{1}{n}(x_1 + x_2 + \cdots + x_n) = \frac{1}{n} \sum_{i=1}^n x_i.$$

Since M is properly defined for any number of arguments, it is an extended aggregation function, see Definition 1.6.

Main properties

- The arithmetic mean M is a strictly increasing aggregation function;
- M is a symmetric function;

- M is an additive function, i.e., $M(\mathbf{x} + \mathbf{y}) = M(\mathbf{x}) + M(\mathbf{y})$ for all $\mathbf{x}, \mathbf{y} \in \mathbb{I}^n$ such that $\mathbf{x} + \mathbf{y} \in \mathbb{I}^n$;
- M is a homogeneous function, i.e., $M(\lambda \mathbf{x}) = \lambda M(\mathbf{x})$ for all $\mathbf{x} \in \mathbb{I}^n$ and for all $\lambda \in \mathbb{R}$;
- The orness measure $orness(M) = \frac{1}{2}$;
- M is a Lipschitz continuous function, with the Lipschitz constant in any $\|\cdot\|_p$ norm (see p. 22) $n^{-1/p}$, the smallest Lipschitz constant of all aggregation functions.

When the inputs are not symmetric, it is a common practice to associate each input with a weight, a number $w_i \in [0, 1]$ which reflects the relative contribution of this input to the total score. For example, in shareholders' meetings, the strength of each vote is associated with the number of shares this shareholder possesses. The votes are usually just added to each other, and after dividing by the total number of shares, we obtain a weighted arithmetic mean. Weights can also represent the reliability of an input or its importance.

Weights are not the only way to obtain asymmetric functions, we will study other methods in Sect. 2.4 and in Chap. 5. Recall from Chap. 1 the definition of a weighting vector:

Definition 2.5 (*Weighting vector*) A vector $\mathbf{w} = (w_1, \dots, w_n)$ is called a weighting vector if $w_i \in [0, 1]$ and $\sum_{i=1}^n w_i = 1$.

Definition 2.6 (*Weighted arithmetic mean (WAM)*) Given a weighting vector \mathbf{w} , the weighted arithmetic mean is the function

$$M_{\mathbf{w}}(\mathbf{x}) = w_1 x_1 + w_2 x_2 + \dots + w_n x_n = \sum_{i=1}^n w_i x_i = \langle \mathbf{w}, \mathbf{x} \rangle.$$

Main properties

- A weighted arithmetic mean $M_{\mathbf{w}}$ is a strictly increasing aggregation function, if all $w_i > 0$;
- $M_{\mathbf{w}}$ is an asymmetric (unless $w_i = 1/n$ for all $i \in \{1, \dots, n\}$) idempotent function;
- $M_{\mathbf{w}}$ is an additive function, i.e., $M_{\mathbf{w}}(\mathbf{x} + \mathbf{y}) = M_{\mathbf{w}}(\mathbf{x}) + M_{\mathbf{w}}(\mathbf{y})$ for all $\mathbf{x}, \mathbf{y} \in \mathbb{I}^n$ such that $\mathbf{x} + \mathbf{y} \in \mathbb{I}^n$;
- $M_{\mathbf{w}}$ is a homogeneous function, i.e., $M_{\mathbf{w}}(\lambda \mathbf{x}) = \lambda M_{\mathbf{w}}(\mathbf{x})$ for all $\mathbf{x} \in \mathbb{I}^n$ and for all $\lambda \in \mathbb{R}$;
- Jensen's inequality: for any convex function² $g : \mathbb{I} \rightarrow [-\infty, \infty]$, $g(M_{\mathbf{w}}(\mathbf{x})) \leq M_{\mathbf{w}}(g(x_1), \dots, g(x_n))$.
- $M_{\mathbf{w}}$ is a Lipschitz continuous function, in fact it is a kernel aggregation function (see p. 23);

²A function g is convex if and only if $g(\alpha t_1 + (1 - \alpha)t_2) \leq \alpha g(t_1) + (1 - \alpha)g(t_2)$ for all $t_1, t_2 \in \text{Dom}(g)$ and $\alpha \in [0, 1]$.

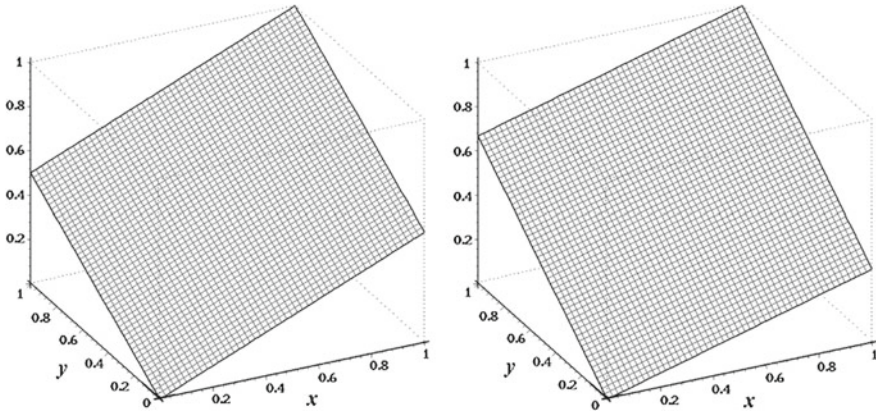


Fig. 2.1 3D plots of weighted arithmetic means $M_{(\frac{1}{2}, \frac{1}{2})}$ and $M_{(\frac{1}{3}, \frac{2}{3})}$

- $M_{\mathbf{w}}$ is a shift-invariant function (see p. 17);
- The orness measure $orness(M_{\mathbf{w}}) = \frac{1}{2}$ ³;
- $M_{\mathbf{w}}$ is a special case of the Choquet integral (see Sect. 4.1) with respect to an additive fuzzy measure (Fig. 2.1).

Geometric and harmonic means

Weighted arithmetic means are good for averaging inputs that can be added together. Frequently the inputs are not added but multiplied. For example, when averaging the rates of investment return over several years the use of the arithmetic mean is incorrect. This is because the rate of return (say 10 %) signifies that in one year the investment was multiplied by a factor 1.1. If the return is 20 % in the next year, then the total is multiplied by 1.2, which means that the original investment is multiplied by a factor of 1.1×1.2 . The average return is calculated by using the geometric mean of 1.1 and 1.2, which gives ≈ 1.15 .

Definition 2.7 (*Geometric mean*) The geometric mean is the function

³It is easy to check that

$$\int_{[0,1]^n} M(\mathbf{x}) d\mathbf{x} = \frac{1}{n} \left(\int_0^1 x_1 dx_1 + \cdots + \int_0^1 x_n dx_n \right) = \frac{n}{n} \int_0^1 t dt = \frac{1}{2}.$$

Substituting the above value in (2.1) we obtain $orness(M) = \frac{1}{2}$. Following, for a weighted arithmetic mean we also obtain

$$\int_{[0,1]^n} M_{\mathbf{w}}(\mathbf{x}) d\mathbf{x} = w_1 \int_0^1 x_1 dx_1 + \cdots + w_n \int_0^1 x_n dx_n = \sum_{i=1}^n w_i \int_0^1 t dt = \frac{1}{2}.$$

$$G(\mathbf{x}) = \sqrt[n]{x_1 x_2 \dots x_n} = \left(\prod_{i=1}^n x_i \right)^{1/n}.$$

Definition 2.8 (*Weighted geometric mean (WGM)*) Given a weighting vector \mathbf{w} , the weighted geometric mean is the function

$$G_{\mathbf{w}}(\mathbf{x}) = \prod_{i=1}^n x_i^{w_i}.$$

Definition 2.9 (*Harmonic mean*) The harmonic mean is the function

$$H(\mathbf{x}) = n \left(\sum_{i=1}^n \frac{1}{x_i} \right)^{-1}.$$

Definition 2.10 (*Weighted harmonic mean (WHM)*) Given a weighting vector \mathbf{w} , the weighted harmonic mean is the function

$$H_{\mathbf{w}}(\mathbf{x}) = \left(\sum_{i=1}^n \frac{w_i}{x_i} \right)^{-1}.$$

Note 2.11 If the weighting vector \mathbf{w} is given without normalization, i.e., $W = \sum_{i=1}^n w_i \neq 1$, then one can either normalize it first by dividing each component by W , or use the alternative expressions for weighted geometric and harmonic means

$$G_{\mathbf{w}}(\mathbf{x}) = \left(\prod_{i=1}^n x_i^{w_i} \right)^{1/W},$$

$$H_{\mathbf{w}}(\mathbf{x}) = W \left(\sum_{i=1}^n \frac{w_i}{x_i} \right)^{-1}.$$

Geometric-Arithmetic Mean Inequality

The following result is an extended version of the well known geometric-arithmetic means inequality

$$H_{\mathbf{w}}(\mathbf{x}) \leq G_{\mathbf{w}}(\mathbf{x}) \leq M_{\mathbf{w}}(\mathbf{x}), \quad (2.6)$$

for any vector \mathbf{x} and weighting vector \mathbf{w} , with equality if and only if $\mathbf{x} = (t, t, \dots, t)$.

Another curious relation between these three means is that for $n = 2$ we have $G(x, y) = \sqrt{M(x, y) \cdot H(x, y)}$ (Figs. 2.2 and 2.3).

Power means

A further generalization of the arithmetic mean is a family called power means (also called *root-power means*). This family is defined by

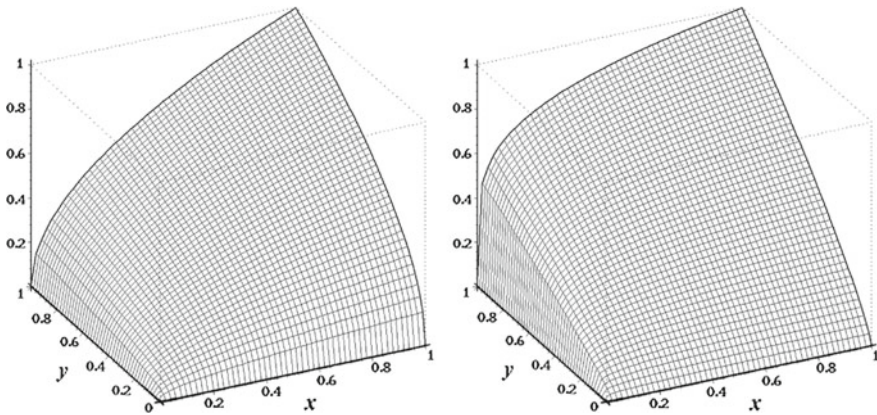


Fig. 2.2 3D plots of weighted geometric means $G_{(\frac{1}{2}, \frac{1}{2})}$ and $G_{(\frac{1}{5}, \frac{4}{5})}$

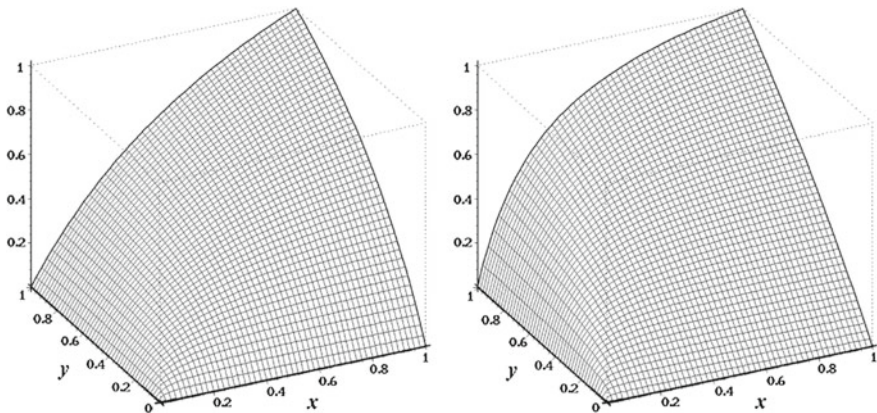


Fig. 2.3 3D plots of weighted harmonic means $H_{(\frac{1}{2}, \frac{1}{2})}$ and $H_{(\frac{1}{5}, \frac{4}{5})}$

Definition 2.12 (*Power mean*) For $r \in \mathbb{R}$, the power mean is the function

$$M_{[r]}(\mathbf{x}) = \left(\frac{1}{n} \sum_{i=1}^n x_i^r \right)^{1/r},$$

if $r \neq 0$, and $M_{[0]}(\mathbf{x}) = G(\mathbf{x})$.⁴

Definition 2.13 (*Weighted power mean*) Given a weighting vector \mathbf{w} and $r \in \mathbb{R}$, the weighted power mean is the function

$$M_{\mathbf{w},[r]}(\mathbf{x}) = \left(\sum_{i=1}^n w_i x_i^r \right)^{1/r},$$

if $r \neq 0$, and $M_{\mathbf{w},[0]}(\mathbf{x}) = G_{\mathbf{w}}(\mathbf{x})$.

Note 2.14 The family of weighted power means is *augmented* to $r = -\infty$ and $r = \infty$ by using the limiting cases

$$M_{\mathbf{w},[-\infty]}(\mathbf{x}) = \lim_{r \rightarrow -\infty} M_{\mathbf{w},[r]}(\mathbf{x}) = \min(\mathbf{x}),$$

$$M_{\mathbf{w},[\infty]}(\mathbf{x}) = \lim_{r \rightarrow \infty} M_{\mathbf{w},[r]}(\mathbf{x}) = \max(\mathbf{x}).$$

However \min and \max are not themselves power means.

The limiting case of the weighted geometric mean is also obtained as

$$M_{\mathbf{w},[0]}(\mathbf{x}) = \lim_{r \rightarrow 0} M_{\mathbf{w},[r]}(\mathbf{x}) = G_{\mathbf{w}}(\mathbf{x}).$$

Of course, the family of weighted power means includes the special cases $M_{\mathbf{w},[1]}(\mathbf{x}) = M_{\mathbf{w}}(\mathbf{x})$, and $M_{\mathbf{w},[-1]}(\mathbf{x}) = H_{\mathbf{w}}(\mathbf{x})$. Another special case is the weighted *quadratic mean*

$$M_{\mathbf{w},[2]}(\mathbf{x}) = Q_{\mathbf{w}}(\mathbf{x}) = \sqrt{\sum_{i=1}^n w_i x_i^2}.$$

Main properties

- The weighted power mean $M_{\mathbf{w},[r]}$ is a strictly increasing aggregation function, if all $w_i > 0$ and $0 < r < \infty$;
- $M_{\mathbf{w},[r]}$ is a continuous function on $[0, \infty)^n$;

⁴We shall use square brackets in the notation $M_{[r]}$ for power means to distinguish them from quasi-arithmetic means M_g (see Sect. 2.3), where parameter g denotes a generating function rather than a real number. The same applies to the weighted power means.

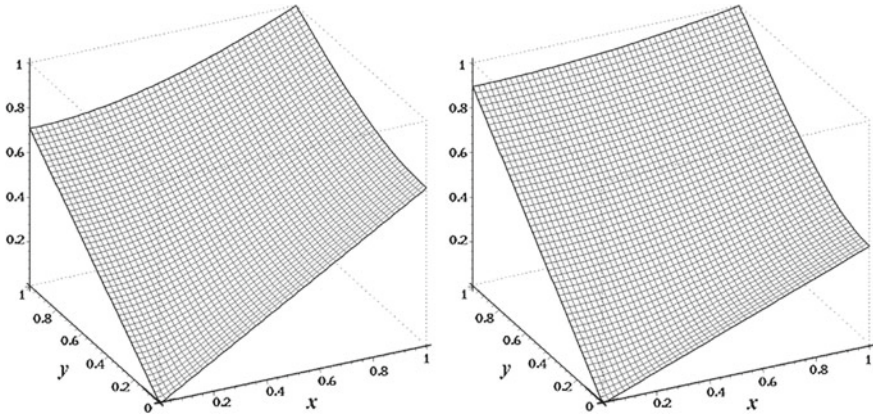


Fig. 2.4 3D plots of weighted quadratic mean $Q_{(\frac{1}{2}, \frac{1}{2})}$ and $Q_{(\frac{1}{3}, \frac{2}{3})}$

- $M_{\mathbf{w},[r]}$ is an asymmetric idempotent function (symmetric if all $w_i = \frac{1}{n}$);
- $M_{\mathbf{w},[r]}$ is a homogeneous function, i.e., $M_{\mathbf{w},[r]}(\lambda \mathbf{x}) = \lambda M_{\mathbf{w},[r]}(\mathbf{x})$ for all $\mathbf{x} \in [0, \infty)^n$ and for all $\lambda \in \mathbb{R}$; it is the only homogeneous weighted quasi-arithmetic mean (this class is introduced in Sect. 2.3);
- Weighted power means are comparable: $M_{\mathbf{w},[r]}(\mathbf{x}) \leq M_{\mathbf{w},[s]}(\mathbf{x})$ if $r \leq s$; this implies the geometric-arithmetic mean inequality;
- $M_{\mathbf{w},[r]}$ has absorbing element (always $a = 0$) if and only if $r \leq 0$ (and all weights w_i are positive);
- $M_{\mathbf{w},[r]}$ does not have a neutral element⁵ (Figs. 2.4 and 2.5).

Measure of orness

Calculations for the geometric mean yields

$$orness(G) = -\frac{1}{n-1} + \frac{n+1}{n-1} \left(\frac{n}{n+1} \right)^n,$$

but for other means explicit formulas are known only for special cases, e.g., $n = 2$

$$\int_{[0,1]^2} Q(\mathbf{x}) d\mathbf{x} = \frac{1}{3} \left(1 + \frac{1}{\sqrt{2}} \ln(1 + \sqrt{2}) \right) \approx 0.541075,$$

$$\int_{[0,1]^2} H(\mathbf{x}) d\mathbf{x} = \frac{4}{3} (1 - \ln(2)), \text{ and}$$

⁵The limiting cases \min ($r = -\infty$) and \max ($r = \infty$) which have neutral elements $e = 1$ and $e = 0$ respectively, are not themselves power means.

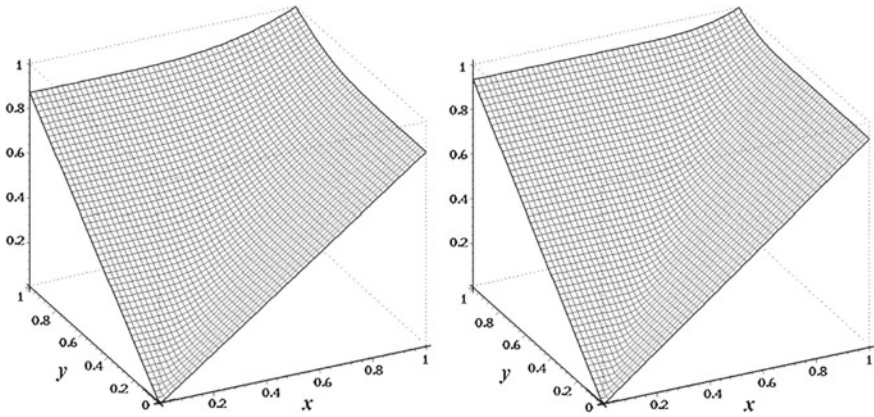


Fig. 2.5 3D plots of power means $M_{[5]}$ and $M_{[10]}$

$$\int_{[0,1]^2} M_{[-2]}(\mathbf{x}) d\mathbf{x} = \frac{2}{3}(2 - \sqrt{2}),$$

from which, when $n = 2$, $orness(Q) \approx 0.623225$, $orness(H) \approx 0.22741$, and $orness(M_{[-2]}) = 3 - 2\sqrt{2}$.

Definition 2.15 (*Dual weighted power mean*) Let $M_{\mathbf{w},[r]}$ be a weighted power mean on $[0, 1]$. The function

$$\bar{M}_{\mathbf{w},[r]}(\mathbf{x}) = 1 - M_{\mathbf{w},[r]}(\mathbf{1} - \mathbf{x})$$

is called the dual weighted power mean.

Note 2.16 The dual weighted power mean is obviously a mean (the class of means is closed under duality). The absorbent element, if any, becomes $a = 1$. The extensions of weighted power means satisfy $\bar{M}_{\mathbf{w},[\infty]}(\mathbf{x}) = M_{\mathbf{w},[-\infty]}(\mathbf{x})$ and $\bar{M}_{\mathbf{w},[-\infty]}(\mathbf{x}) = M_{\mathbf{w},[\infty]}(\mathbf{x})$. The weighted arithmetic mean $M_{\mathbf{w}}$ is self-dual.

2.3 Weighted Quasi-arithmetic Means

2.3.1 Definitions

Quasi-arithmetic means generalize power means. Consider a univariate continuous strictly monotone function $g : \mathbb{I} \rightarrow [-\infty, \infty]$, which we call a *generating function*. Of course, g is invertible, but it is not necessarily a bijection (i.e., its range may be $Ran(g) \subset [-\infty, \infty]$).

Definition 2.17 (*Quasi-arithmetic mean (QAM)*) For a given generating function g , the quasi-arithmetic mean is the function

$$M_g(\mathbf{x}) = g^{-1} \left(\frac{1}{n} \sum_{i=1}^n g(x_i) \right). \quad (2.7)$$

Its weighted analogue is given by

Definition 2.18 (*Weighted quasi-arithmetic mean (WQAM)*) For a given generating function g , and a weighting vector \mathbf{w} , the weighted quasi-arithmetic mean is the function

$$M_{\mathbf{w},g}(\mathbf{x}) = g^{-1} \left(\sum_{i=1}^n w_i g(x_i) \right). \quad (2.8)$$

The weighted power means are a subclass of weighted quasi-arithmetic means with the generating function

$$g(t) = \begin{cases} t^r, & \text{if } r \neq 0, \\ \log(t), & \text{if } r = 0. \end{cases}$$

Note 2.19 Observe that if $\text{Ran}(g) = [-\infty, \infty]$, then we have the summation $-\infty + \infty$ or $+\infty - \infty$ if $x_i = 0$ and $x_j = 1$ for some $i \neq j$. When this occurs, a convention, such as $-\infty + \infty = +\infty - \infty = -\infty$, is adopted, and continuity of $M_{\mathbf{w},g}$ is lost.

Note 2.20 If the weighting vector \mathbf{w} is not normalized, i.e., $W = \sum_{i=1}^n w_i \neq 1$, then weighted quasi-arithmetic means are expressed as

$$M_{\mathbf{w},g}(\mathbf{x}) = g^{-1} \left(\frac{1}{W} \sum_{i=1}^n w_i g(x_i) \right).$$

2.3.2 Main Properties

- Weighted quasi-arithmetic means are continuous if and only if $\text{Ran}(g) \neq [-\infty, \infty]$ [Kom01];
- Weighted quasi-arithmetic means on $[a, b]$ with strictly positive weights are strictly monotone increasing on $]a, b[^n$;
- The class of weighted quasi-arithmetic means is closed under duality. That is, given a strong negation N , the N -dual of a weighted quasi-arithmetic mean (on $[0, 1]$) $M_{\mathbf{w},g}$ is in turn a weighted quasi-arithmetic mean, given by $M_{\mathbf{w},g \circ N}$. For the standard negation, the dual of a weighted quasi-arithmetic mean is characterized by the generating function $h(t) = g(1 - t)$;
- The following result regarding self-duality holds (see, e.g., [PTC02]): Given a strong negation N , a weighted quasi-arithmetic mean $M_{\mathbf{w},g}$ is N -self-dual if and

only if N is the strong negation generated by g , i.e., if $N(t) = g^{-1}(g(0) + g(1) - g(t))$ for any $t \in [0, 1]$. This implies, in particular:

- Weighted quasi-arithmetic means, such that $g(0) = \pm\infty$ or $g(1) = \pm\infty$ are never N -self-dual (in fact, they are dual to each other);
- Weighted arithmetic means are always self-dual (i.e., N -self-dual with respect to the standard negation $N(t) = 1 - t$);
- The generating function is not defined uniquely, but up to an arbitrary linear transformation, i.e., if $g(t)$ is a generating function of some weighted quasi-arithmetic mean, then $ag(t) + b$, $a, b \in \mathbb{R}$, $a \neq 0$ is also a generating function of *the same mean*,⁶ provided $\text{Ran}(g) \neq [-\infty, \infty]$;
- There are incomparable quasi-arithmetic means. Two quasi-arithmetic means M_g and M_h satisfy $M_g \leq M_h$ if and only if either the composite $g \circ h^{-1}$ is convex and g is decreasing, or $g \circ h^{-1}$ is concave and g increasing;
- Consequently, $M_g \leq M_{Id}$ if and only if g^{-1} is a convex function. By using duality, $M_{Id} \leq M_g$ if and only if g^{-1} is concave.
- The only homogeneous weighted quasi-arithmetic means are weighted power means;
- Weighted quasi-arithmetic means do not have a neutral element.⁷ They may have an absorbing element only when all the weights are strictly positive and $g(a) = \pm\infty$ or $g(b) = \pm\infty$, and in such cases the corresponding absorbing elements are, respectively, a and b .
- Quasi-arithmetic means may or may not be Lipschitz continuous. Their Lipschitz properties are discussed in detail in [BCJ10].

Convexity and concavity of the generator

By comparing a quasi-arithmetic mean with the arithmetic mean $M = M_{Id}$ we can establish the convexity or concavity of its generator, namely the inverse g^{-1} is convex if and only if $M_g \leq M$. This is the case of a partial conjunction in the sense of Dujmovic [Duj07], i.e., $\text{orness}(M_g) \leq \text{orness}(M) = \frac{1}{2}$. For concave g^{-1} we have a partial disjunction $\text{orness}(M_g) \geq \frac{1}{2}$.

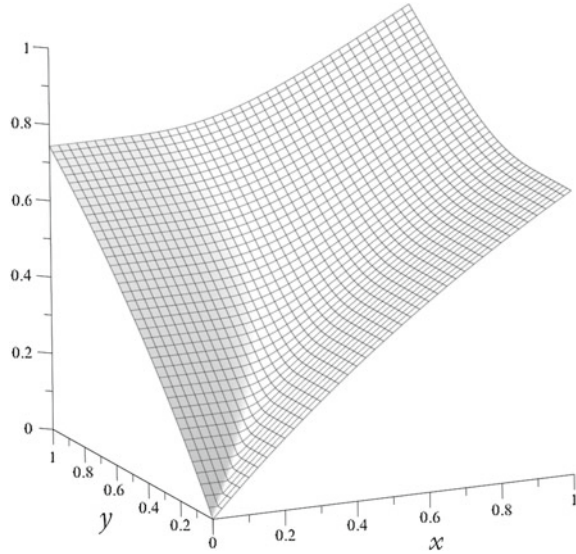
However the convexity (concavity) of the inverse g^{-1} does not imply concavity (convexity) of the mean M_g , although this is the case for many means, such as power and exponential means. As a counterexample consider a convex (on $[0,1]$) generator $g(t) = e^{-\frac{2}{t}}$, and the concave inverse $g^{-1}(t) = -\frac{2}{\ln(t)}$. The resulting mean $M_g \geq M$ but is neither convex nor concave as can be appreciated from Fig. 2.6. For example, $M_g(x, 0) = \frac{2}{\frac{2}{x} + \ln(2)}$ is concave whereas $M_g(x, 1) = \frac{2}{\ln(2) - \ln(e^{-\frac{2}{x}} + e^{-2})}$ is convex.

Recently the convexity (concavity) of the generator was connected to the supermodularity (submodularity) of the function M_g [CM09]. A function f is called supermodular if $f(\mathbf{x} \vee \mathbf{y}) + f(\mathbf{x} \wedge \mathbf{y}) \geq f(\mathbf{x}) + f(\mathbf{y})$, and submodular if the

⁶For this reason, one can assume that g is monotone increasing, as otherwise we can simply take $-g$.

⁷Observe that the limiting cases min and max are not quasi-arithmetic means.

Fig. 2.6 3D plot of the quasi-arithmetic mean with the generator $g(t) = e^{-\frac{2}{t}}$



inequality is reversed.⁸ The operations \vee, \wedge are the usual componentwise maximum and minimum. Supermodularity is related to the property of increasing differences: a function f is supermodular if and only if it has increasing differences (which in the bi-variate case reads $f(x+a, y+b) - f(x, y+b) \geq f(x+a, y) - f(x, y)$ for positive a, b). A twice differentiable function f is supermodular if and only if its mixed partial derivatives $\frac{\partial^2 f}{\partial x_i \partial x_j} \geq 0$ for all distinct i, j . Then M_g is supermodular if and only if g^{-1} is convex [CM09], and hence if and only if $M_g \leq M$ is a partial conjunction. The case of partial disjunction equivalent to submodular functions M_g and concave g^{-1} is dealt with by using duality. The result also applies to weighted quasi-arithmetic means.

Orness measure

Calculation of the orness measure using (2.1) is technically difficult as the closed form expression for the integrals is available only in special cases. Some alternative definitions of the orness measure were presented, such as the one in [Duj05]

$$\text{orness}(M_g) = \frac{g(b) - \int_a^b g(t)dt}{g(b) - g(a)},$$

or the one in [Liu06]

$$\text{orness}(M_g) = \frac{g^{-1}\left(\frac{\int_a^b g(t)dt}{b-a}\right) - a}{b-a}.$$

⁸This concept of supermodularity of a function on \mathbb{I}^n is different from supermodularity and submodularity of fuzzy measures in Definition 4.14.

2.3.3 Examples

Example 2.21 (Weighted power means) Weighted power means are a special case of weighted quasi-arithmetic means, with $g(t) = t^r$, $r \neq 0$ and $g(t) = \log(t)$ if $r = 0$. Note that the generating function $g(t) = \frac{t^r - 1}{r}$ defines exactly the same power mean (as a particular case of a linear transformation of g).

Example 2.22 (Harmonic and geometric means) These classical means, defined on p. 60, 61, are special cases of the power means, obtained when $g(t) = t^r$, $r = -1$ and $r = 0$ respectively.

Example 2.23 Let $g(t) = \log \frac{t}{1-t}$. The corresponding quasi-arithmetic mean M_g is given by

$$M_g(\mathbf{x}) = \begin{cases} \frac{\sqrt[n]{\prod_{i=1}^n x_i}}{\sqrt[n]{\prod_{i=1}^n x_i} + \sqrt[n]{\prod_{i=1}^n (1-x_i)}}, & \text{if } \{0, 1\} \not\subseteq \{x_1, \dots, x_n\} \\ 0, & \text{otherwise,} \end{cases}$$

that is, $M_g = \frac{G}{G+1-G_d}$ and with the convention $\frac{0}{0} = 0$.

Example 2.24 (Weighted trigonometric means) Let $g_1(t) = \sin(\frac{\pi}{2}t)$, $g_2(t) = \cos(\frac{\pi}{2}t)$, and $g_3(t) = \tan(\frac{\pi}{2}t)$ be the generating functions. The weighted trigonometric means are the functions

$$\begin{aligned} SM_{\mathbf{w}}(\mathbf{x}) &= \frac{2}{\pi} \arcsin \left(\sum_{i=1}^n w_i \sin\left(\frac{\pi}{2}x_i\right) \right), \\ CM_{\mathbf{w}}(\mathbf{x}) &= \frac{2}{\pi} \arccos \left(\sum_{i=1}^n w_i \cos\left(\frac{\pi}{2}x_i\right) \right) \text{ and} \\ TM_{\mathbf{w}}(\mathbf{x}) &= \frac{2}{\pi} \arctan \left(\sum_{i=1}^n w_i \tan\left(\frac{\pi}{2}x_i\right) \right). \end{aligned}$$

Their 3D plots are presented on Figs. 2.7 and 2.8.

Example 2.25 (Weighted exponential means) Let the generating function be

$$g(t) = \begin{cases} \gamma^t, & \text{if } \gamma \neq 1, \\ t, & \text{if } \gamma = 1. \end{cases}$$

The weighted exponential mean is the function

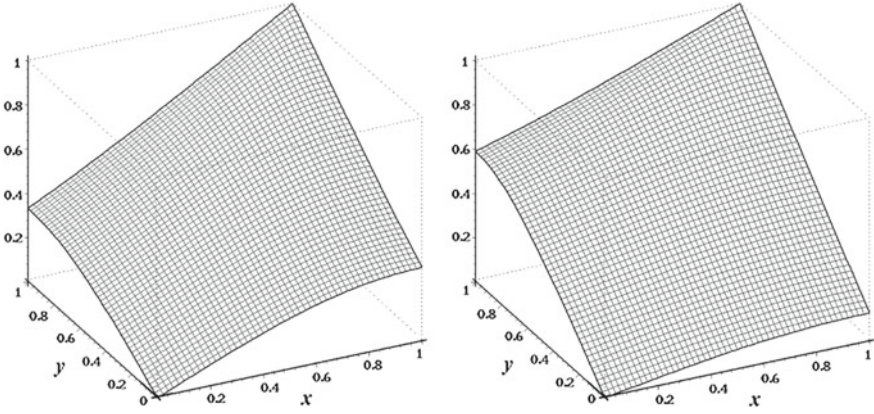


Fig. 2.7 3D plots of weighted trigonometric means $SM_{(\frac{1}{2}, \frac{1}{2})}$ and $SM_{(\frac{1}{3}, \frac{2}{3})}$

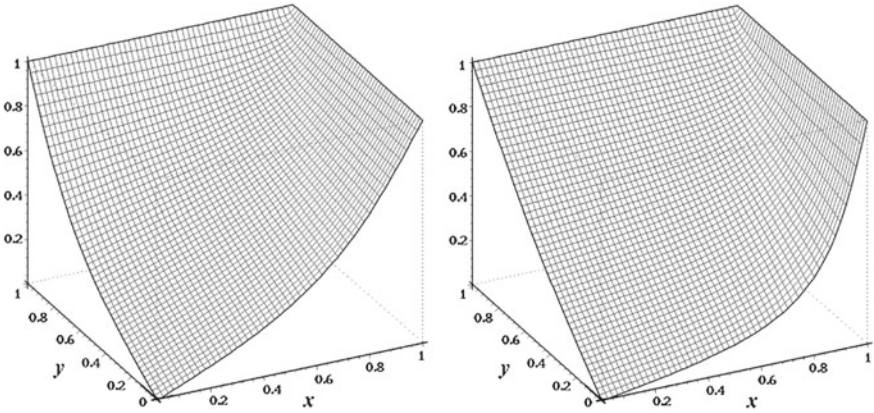


Fig. 2.8 3D plots of weighted trigonometric means $TM_{(\frac{1}{2}, \frac{1}{2})}$ and $TM_{(\frac{1}{3}, \frac{2}{3})}$

$$EM_{\mathbf{w}, \gamma}(\mathbf{x}) = \begin{cases} \log_{\gamma} \left(\sum_{i=1}^n w_i \gamma^{x_i} \right), & \text{if } \gamma \neq 1, \\ M_{\mathbf{w}}(\mathbf{x}), & \text{if } \gamma = 1. \end{cases}$$

3D plots of some weighted exponential means are presented on Figs. 2.9 and 2.10.

Example 2.26 There is another mean also known as exponential [Bul03], given for $\mathbf{x} \geq \mathbf{1}$ by

$$f(\mathbf{x}) = \exp \left(\left(\prod_{i=1}^n \log(x_i) \right)^{1/n} \right).$$

It is a quasi-arithmetic mean with a generating function $g(t) = \log(\log(t))$, and its inverse $g^{-1}(t) = \exp(\exp(t))$.

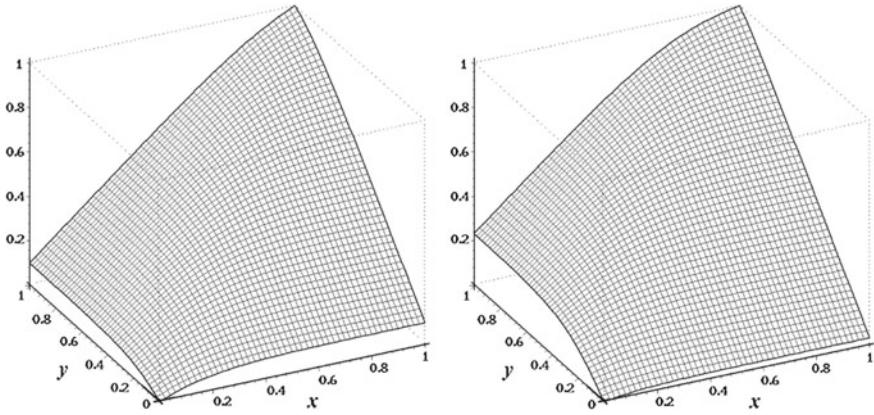


Fig. 2.9 3D plots of weighted exponential means $EM_{(\frac{1}{2}, \frac{1}{2}), 0.001}$ and $EM_{(\frac{1}{3}, \frac{1}{3}), 0.001}$

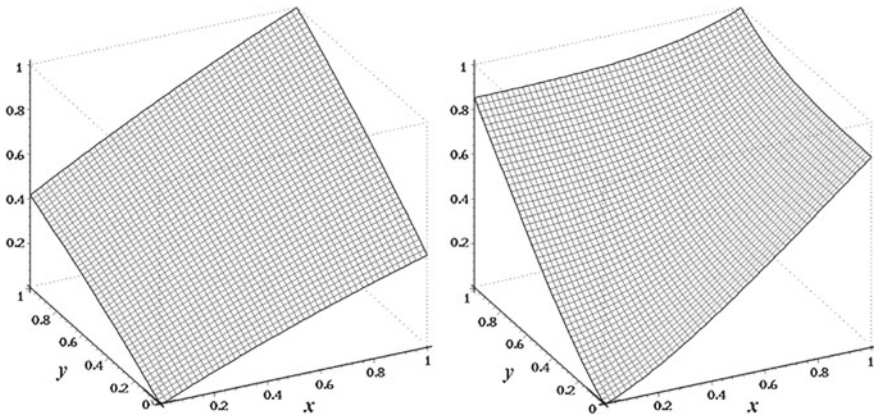


Fig. 2.10 3D plots of exponential means $EM_{(\frac{1}{2}, \frac{1}{2}), 0.5}$ and $EM_{(\frac{1}{2}, \frac{1}{2}), 100}$

In the domain $[0, 1]^n$ one can use a generating function $g(t) = \log(-\log(t))$, so that its inverse is $g^{-1}(t) = \exp(-\exp(t))$. This mean is discontinuous, since $\text{Ran}(g) = [-\infty, \infty]$. We obtain the expression

$$f(\mathbf{x}) = \exp\left(-\prod_{i=1}^n \left(-\log(x_i)\right)^{1/n}\right).$$

Example 2.27 (Weighted radical means) Let $\gamma > 0$, $\gamma \neq 1$, and let the generating function be

$$g(t) = \gamma^{1/t}.$$

The weighted radical mean is the function

$$RM_{\mathbf{w},\gamma}(\mathbf{x}) = \left(\log_{\gamma} \left(\sum_{i=1}^n w_i \gamma^{1/x_i} \right) \right)^{-1}.$$

3D plots of some radical means are presented on Fig. 2.11.

Example 2.28 (Weighted basis-exponential means) Weighted basis-exponential means are obtained by using the generating function $g(t) = t^t$ and $t \geq \frac{1}{e}$ (this generating function is decreasing on $[0, \frac{1}{e}[$ and increasing on $]\frac{1}{e}, \infty]$, hence the restriction). The value of this mean is such a value y that

$$y^y = \sum_{i=1}^n w_i x_i^{x_i}.$$

For practical purposes this equation has to be solved for y numerically.

Example 2.29 (Weighted basis-radical means) Weighted basis-radical means are obtained by using the generator $g(t) = t^{1/t}$ and $t \geq \frac{1}{e}$ (restriction for the same reason as in the Example 2.28). The value of this mean is such a value y that

$$y^{1/y} = \sum_{i=1}^n w_i x_i^{1/x_i}.$$

For practical purposes this equation has to be solved for y numerically.

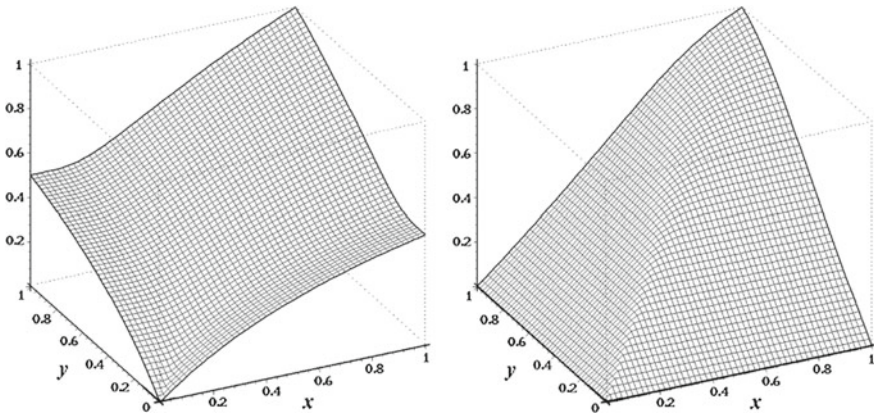


Fig. 2.11 3D plots of radical means $RM_{(\frac{1}{2}, \frac{1}{2}), 0.5}$ and $RM_{(\frac{1}{2}, \frac{1}{2}), 100}$

Example 2.30 (*Tsallis q -exponential means*) Another interesting generator was recently proposed in [RA14]. The authors used the q -exponentials e_q^t that are solutions to the differential equation $F' = F^q$ with the initial condition $F(0) = 1$. These are given by

$$F(t) = e_q^t = \lim_{r \rightarrow 1-q} (1 + rt)^{\frac{1}{r}} = (1 + (1 - q)t)^{\frac{1}{1-q}}$$

when $q \neq 1$ and e^t otherwise. The generators g themselves are derived as $g_1(t) = e_q^t - 1$ and $g_2(t) = e_q^{-(1-t)}$.

2.3.4 Calculation

Generating functions offer a nice way of calculating the values of weighted quasi-arithmetic means. Note that we can write

$$M_{\mathbf{w},g}(\mathbf{x}) = g^{-1}(M_{\mathbf{w}}(g(\mathbf{x}))),$$

where $g(\mathbf{x}) = (g(x_1), \dots, g(x_n))$. Thus calculation can be performed in three steps:

1. Transform all the inputs by calculating vector $g(\mathbf{x})$;
2. Calculate the weighted arithmetic mean of the transformed inputs;
3. Calculate the inverse g^{-1} of the computed mean.

However one needs to be careful with the limiting cases, for example when $g(x_i)$ becomes infinite. Typically this is an indication of existence of an absorbing element, this needs to be picked up by the computer subroutine. Similarly, special cases like $M_{[r]}(\mathbf{x})$, $r \rightarrow \pm\infty$ have to be accommodated (in these cases the subroutine has to return the minimum or the maximum).

2.3.5 Weighting Triangles

When we are interested in using weighted quasi-arithmetic means as extended aggregation functions, we need to have a clear rule as to how the weighting vectors are calculated for each dimension $n = 2, 3, \dots$. For symmetric quasi-arithmetic means we have a simple rule: for each n the weighting vector $\mathbf{w}^n = (\frac{1}{n}, \dots, \frac{1}{n})$. For weighted means we need the concept of a weighting triangle [Cal+00, MC97].

Definition 2.31 (*Weighting triangle*) A *weighting triangle* or *triangle of weights* is a set of numbers $w_i^n \in [0, 1]$, for $i = 1, \dots, n$ and $n \geq 1$, such that: $\sum_{i=1}^n w_i^n = 1$, for all $n \geq 1$. It will be represented in the following form

$$\begin{array}{cccc}
& & 1 & \\
& & w_1^2 & w_2^2 \\
& w_1^3 & w_2^3 & w_3^3 \\
w_1^4 & w_2^4 & w_3^4 & w_4^4 \\
& \dots & &
\end{array}$$

Weighting triangles will be denoted by Δw_i^n .

Example 2.32 A basic example is the “normalized” Pascal triangle

$$\begin{array}{cccccc}
& & & 1 & & \\
& & 1/2 & & 1/2 & \\
& 1/4 & & 2/4 & & 1/4 \\
& 1/8 & 3/8 & & 3/8 & 1/8 \\
1/16 & 4/16 & 6/16 & 4/16 & 1/16 & \\
& \dots & \dots & \dots & \dots &
\end{array}$$

The generic formula for the weighting vector of dimension n in this weighting triangle is ⁹

$$w^n = \frac{1}{2^{n-1}} \left(\binom{n-1}{0}, \binom{n-1}{1}, \dots, \binom{n-1}{n-1} \right)$$

for each $n \geq 1$.

It is possible to generate weighting triangles in different ways [CM99]:

Proposition 2.33 *The following methods generate weighting triangles:*

1. Let $\lambda_1, \lambda_2, \dots \geq 0$ be a sequence of non-negative real numbers such that $\lambda_1 > 0$. Define the weights using

$$w_i^n = \frac{\lambda_{n-i+1}}{\lambda_1 + \dots + \lambda_n},$$

for all $i = 1, \dots, n$ and $n \geq 1$;

2. Let N be a strong negation.¹⁰ Generate the weights using N by

$$w_i^n = N\left(\frac{i-1}{n}\right) - N\left(\frac{i}{n}\right),$$

for all $i = 1, \dots, n$ and $n \geq 1$;

3. Let Q be a monotone non-decreasing function $Q : [0, 1] \rightarrow [0, 1]$ such that¹¹ $Q(0) = 0$ and $Q(1) = 1$. Generate the weights using function Q by [Yag91]

⁹Recall $\binom{n}{m} = \frac{n!}{m!(n-m)!}$.

¹⁰See Definition 1.51 on p. 18.

¹¹This is a so-called quantifier, see p. 120.

$$w_i^n = Q\left(\frac{i}{n}\right) - Q\left(\frac{i-1}{n}\right),$$

for all $i = 1, \dots, n$ and $n \geq 1$.

Another way to construct weighting triangles is by using fractal structures exemplified below. Such weighting triangles cannot be generated by any of the methods in Proposition 2.33.

Example 2.34 The following two triangles belong to the Sierpinski family [CM99]

$$\begin{array}{ccccccc} & & & & 1 & & \\ & & & & 1 \cdot \frac{1}{4} & 3 \cdot \frac{1}{4} & \\ & & & & 1 \cdot \frac{1}{4} & 3 \cdot \frac{1}{4^2} & 3^2 \cdot \frac{1}{4^2} \\ & & & & 1 \cdot \frac{1}{4} & 3 \cdot \frac{1}{4^2} & 3^2 \cdot \frac{1}{4^3} & 3^3 \cdot \frac{1}{4^3} \\ & & & & \dots & \dots & \dots \end{array}$$

and

$$\begin{array}{ccccccc} & & & & 1 & & \\ & & & & 1 \cdot \frac{1}{9} & 8 \cdot \frac{1}{9} & \\ & & & & 1 \cdot \frac{1}{9} & 8 \cdot \frac{1}{9^2} & 8^2 \cdot \frac{1}{9^2} \\ & & & & 1 \cdot \frac{1}{9} & 8 \cdot \frac{1}{9^2} & 8^2 \cdot \frac{1}{9^3} & 8^3 \cdot \frac{1}{9^3} \\ & & & & \dots & \dots & \dots \end{array}$$

In general, given two real numbers a, b such that $a < b$ and $\frac{a}{b} \neq \frac{b-a}{b}$, it is possible to define a weighting triangle by [CM99]

$$\begin{array}{ccccccc} & & & & 1 & & \\ & & & & \frac{a}{b} & \frac{b-a}{b} & \\ & & & & \frac{a}{b} & \frac{(b-a)}{b} \left(\frac{a}{b}\right) & \left(\frac{b-a}{b}\right)^2 \\ & & & & \frac{a}{b} & \frac{(b-a)}{b} \left(\frac{a}{b}\right) & \frac{(b-a)}{b} \left(\frac{b-a}{b}\right)^2 & \left(\frac{b-a}{b}\right)^3 \\ & & & & \dots & \dots & \dots \end{array}$$

This weighting triangle also belongs to the Sierpinski family, and a generic formula for the weights is

$$w_i^n = \frac{a}{b} \cdot \left(\frac{b-a}{b}\right)^{i-1}, \quad i = 1, \dots, n-1 \quad \text{and} \quad w_n^n = \left(\frac{b-a}{b}\right)^{n-1}.$$

The question of stability of weighted quasi-arithmetic means with respect to the number of arguments was recently addressed in [BJ13, GRMRB14, RGM13], see Sect. 6.5. This method connects the weighting vectors of various dimensions through

$$w_i^n = (1 - w_n^n)w_i^{n-1}, i = 1, 2, \dots, n-1,$$

as shown in Corollary 6.89.

Let us now mention two characterization theorems, which relate continuity, strict monotonicity and the properties of decomposability and bisymmetry to the class of weighted quasi-arithmetic means.

Theorem 2.35 (Kolmogorov-Nagumo) *An extended aggregation function F is continuous, decomposable,¹² and strictly monotone if and only if there is a monotone bijection $g : \mathbb{I} \rightarrow \mathbb{I}$, such that for each $n > 1$, f_n is a quasi-arithmetic mean M_g .*

The next result is a generalized version of Kolmogorov and Nagumo characterization, due to Aczél [Acz48].

Theorem 2.36 *An extended aggregation function F is continuous, bisymmetric,¹³ idempotent, and strictly monotone if and only if there is a monotone bijection $g : \mathbb{I} \rightarrow \mathbb{I}$, and a weighting triangle Δw_i^n with all positive weights, so that for each $n > 1$, f_n is a weighted quasi-arithmetic mean $M_{\mathbf{w}^n, g}$ (i.e., $f_n = M_{\mathbf{w}^n, g}$).*

Note 2.37 If we omit the strict monotonicity of F , we recover the class of non-strict means introduced by Fodor and Marichal [FM97].

2.3.6 Weights Dispersion

An important quantity associated with weighting vectors is their dispersion, also called entropy.

Definition 2.38 (*Weights dispersion (entropy)*) For a given weighting vector \mathbf{w} its measure of dispersion (entropy) is

$$Disp(\mathbf{w}) = - \sum_{i=1}^n w_i \log w_i, \quad (2.9)$$

with the convention $0 \cdot \log 0 = 0$.

The weights dispersion measures the degree to which a weighted aggregation function f takes into account all inputs. For example, in the case of weighted means, among the two weighting vectors $\mathbf{w}_1 = (0, 1)$ and $\mathbf{w}_2 = (0.5, 0.5)$ the second one may be preferable, since the corresponding weighted mean uses information from two sources rather than a single source, and is consequently less sensitive to input inaccuracies.

¹²See Definition 1.43. Continuity and decomposability imply idempotency.

¹³See Definition 1.44.

A useful normalization of this measure is

$$-\frac{1}{\log n} \sum_{i=1}^n w_i \log w_i.$$

Along with the orness value (p. 57), the weights entropy is an important parameter in choosing weighting vectors of both quasi-arithmetic means and OWA functions (see p. 106).

There are also other entropy measures (e.g., Rényi entropy) frequently used in studies of weighted aggregation functions, e.g., [Yag95].¹⁴

2.3.7 How to Choose Weights

Choosing weights of weighted arithmetic means

In each application the weighting vector of the weighted arithmetic mean will be different. We examine the problem of choosing the weighting vector which fits best some empirical data, the pairs (\mathbf{x}_k, y_k) , $k = 1, \dots, K$. Our goal is to determine the best weighted arithmetic mean that minimizes the norm of the differences between the predicted ($f(\mathbf{x}_k)$) and observed (y_k) values. We will use the least squares or least absolute deviation criterion, as discussed on p. 34. In the first case we have the following optimization problem

$$\begin{aligned} \min \quad & \sum_{k=1}^K \left(\sum_{i=1}^n w_i x_{ik} - y_k \right)^2 \\ \text{s.t.} \quad & \sum_{i=1}^n w_i = 1, w_i \geq 0, i = 1, \dots, n. \end{aligned} \quad (2.10)$$

It is easy to recognize a standard quadratic programming problem (QP), with a convex objective function. There are plenty of standard methods for its solution.

We mentioned on p. 34 that one can use a different fitting criterion, such as the least absolute deviation (LAD) criterion, which translates into a different optimization problem

$$\begin{aligned} \min \quad & \sum_{k=1}^K \left| \sum_{i=1}^n w_i x_{ik} - y_k \right| \\ \text{s.t.} \quad & \sum_{i=1}^n w_i = 1, w_i \geq 0, i = 1, \dots, n. \end{aligned} \quad (2.11)$$

This problem is subsequently converted into a linear programming problem (LP).

¹⁴These measures of entropy can be obtained by relaxing the subadditivity condition which characterizes Shannon entropy [TY05].

Particular attention is needed for the case when the quadratic (resp. linear) programming problems have singular matrices. Such cases appear when there are few data, or when the input values are linearly dependent. While modern quadratic and linear programming methods accommodate for such cases, the minimization problem will typically have multiple solutions. An additional criterion is then used to select one of these solutions, and typically this criterion relates to the dispersion of weights, or the entropy [Tor02], as defined in Definition 2.38. Torra [Tor02] proposes to solve an auxiliary univariate optimization problem to maximize weights dispersion, subject to a given value of (2.10).

Specifically, one solves the problem

$$\begin{aligned} \min \quad & \sum_{i=1}^n w_i \log w_i \\ \text{s.t.} \quad & \sum_{i=1}^n w_i = 1, w_i \geq 0, i = 1, \dots, n, \\ & \sum_{k=1}^K \left(\sum_{i=1}^n w_i x_{ik} - y_k \right)^2 = A, \end{aligned} \tag{2.12}$$

where A is the value of the solution of problem (2.10). It turns out that if Problem (2.10) has multiple solutions, they are expressed in parametric form as linear combinations of one another. Further, the objective function in (2.12) is convex. Therefore problem (2.12) is a convex programming problem subject to linear constraints, and it can be solved by standard methods, see [Tor02].

A different additional criterion is the so-called measure of orness (discussed in Sect. 2.1), which measures how far a given averaging function is from the max function, which is the weakest disjunctive function. It is applicable to any averaging function, and is frequently used as an additional constraint or criterion when constructing these functions. However, for any weighted arithmetic mean, the measure of orness is always $\frac{1}{2}$, therefore this parameter does not discriminate between arithmetic means with different weighting vectors.

Preservation of ordering of the outputs

We recall from Sect. 1.6, p. 32, that sometimes one not only has to fit an aggregation function to the numerical data, but also preserve the ordering of the outputs. That is, if $y_j \leq y_k$ then we expect $f(\mathbf{x}_j) \leq f(\mathbf{x}_k)$.

First, arrange the data, so that the outputs are in non-decreasing order, i.e., $y_k \leq y_{k+1}, k = 1, \dots, K - 1$. Define the additional linear constraints

$$\langle \mathbf{x}_{k+1} - \mathbf{x}_k, \mathbf{w} \rangle = \sum_{i=1}^n w_i (x_{i,k+1} - x_{ik}) \geq 0,$$

$k = 1, \dots, K - 1$. We add the above constraints to problem (2.10) or (2.11) and solve it. The addition of an extra $K - 1$ constraints neither changes the structure of the optimization problem, nor drastically affects its complexity.

Choosing weights of weighted quasi-arithmetic means

Consider the case of weighted quasi-arithmetic means, when a given generating function g is given. As before, we have a data set (\mathbf{x}_k, y_k) , $k = 1, \dots, K$, and we are interested in finding the weighting vector \mathbf{w} that fits the data best. When we use the least squares, as discussed on p. 34, we have the following optimization problem

$$\begin{aligned} \min \quad & \sum_{k=1}^K \left(g^{-1} \left(\sum_{i=1}^n w_i g(x_{ik}) \right) - y_k \right)^2 \\ \text{s.t.} \quad & \sum_{i=1}^n w_i = 1, w_i \geq 0, i = 1, \dots, n. \end{aligned} \quad (2.13)$$

This is a nonlinear optimization problem, but it can be reduced to quadratic programming by the following artifice. Let us apply g to y_k and the inner sum in (2.13). We obtain

$$\begin{aligned} \min \quad & \sum_{k=1}^K \left(\sum_{i=1}^n w_i g(x_{ik}) - g(y_k) \right)^2 \\ \text{s.t.} \quad & \sum_{i=1}^n w_i = 1, w_i \geq 0, i = 1, \dots, n. \end{aligned} \quad (2.14)$$

We recognize a standard quadratic programming problem (QP), with a convex objective function. This approach was discussed in detail in [Bel03, Bel05, BMV04, Tor02]. There are plenty of standard methods of solution.

If one uses the least absolute deviation (LAD) criterion (p. 34) we obtain a different optimization problem

$$\begin{aligned} \min \quad & \sum_{k=1}^K \left| \sum_{i=1}^n w_i g(x_{ik}) - g(y_k) \right| \\ \text{s.t.} \quad & \sum_{i=1}^n w_i = 1, w_i \geq 0, i = 1, \dots, n. \end{aligned} \quad (2.15)$$

This problem is subsequently converted into a linear programming problem (LP).

As in the case of weighted arithmetic means, in the presence of multiple optimal solutions, one can use an additional criterion of the dispersion of weights [Tor02].

One recent result relates the choice of the weighting vectors to stability of the aggregation functions with respect to changes to the number of arguments [BJ13, BJGRM15, GMRR12, GRMRB14]. Section 6.5 discusses this type of stability, and

here we just mention that the notion of stability can be translated into the (nonlinear) constraints on the coefficients of the weighting vectors.

Preservation of ordering of the outputs

Similarly to what we did for weighted arithmetic means (see also Sect. 1.6, p. 32), we will require that the ordering of the outputs is preserved, i.e., if $y_j \leq y_k$ then we expect $f(\mathbf{x}_j) \leq f(\mathbf{x}_k)$. We arrange the data, so that the outputs are in non-decreasing order, $y_k \leq y_{k+1}$, $k = 1, \dots, K - 1$. Then we define the additional linear constraints

$$\langle \mathbf{g}(x_{k+1}) - \mathbf{g}(\mathbf{x}_k), \mathbf{w} \rangle = \sum_{i=1}^n w_i (g(x_{i,k+1}) - g(x_{ik})) \geq 0,$$

$k = 1, \dots, K - 1$. We add the above constraints to problem (2.14) or (2.15) and solve it. The addition of extra $K - 1$ constraints does not change the structure of the optimization problem, nor drastically affects its complexity.

Choosing generating functions

Consider now the case when the generating function g is also unknown, and hence needs to be found based on the data. We study two cases: (a) when g is given algebraically, with one or more unknown parameters to estimate (e.g., $g_p(t) = t^p$, p unknown) and (b) when no specific algebraic form of g is given.

In the first case we solve the problem

$$\begin{aligned} \min_{p, \mathbf{w}} \quad & \sum_{k=1}^K \left(\sum_{i=1}^n w_i g_p(x_{ik}) - g_p(y_k) \right)^2 \\ \text{s.t.} \quad & \sum_{i=1}^n w_i = 1, w_i \geq 0, i = 1, \dots, n, \\ & \text{conditions on } p. \end{aligned} \tag{2.16}$$

While this general optimization problem is non-convex and nonlinear (i.e., difficult to solve), we can convert it to a bi-level optimization problem

$$\begin{aligned} \min_p \quad & \left[\min_{\mathbf{w}} \sum_{k=1}^K \left(\sum_{i=1}^n w_i g_p(x_{ik}) - g_p(y_k) \right)^2 \right] \\ \text{s.t.} \quad & \sum_{i=1}^n w_i = 1, w_i \geq 0, i = 1, \dots, n, \\ & \text{plus conditions on } p. \end{aligned} \tag{2.17}$$

The problem at the inner level is the same as (2.14) with a fixed g_p , which is a QP problem. At the outer level we have a global optimization problem with respect to a single parameter p . It is solved by using one of the standard methods. We recommend deterministic Pijavski-Shubert method.

Example 2.39 Determine the weights and the generating function of a family of weighted power means. We have $g_p(t) = t^p$, and hence solve bi-level optimization problem

$$\begin{aligned} \min_{p \in [-\infty, \infty]} & \left[\min_{\mathbf{w}} \sum_{k=1}^K \left(\sum_{i=1}^n w_i x_{ik}^p - y_k^p \right)^2 \right] \\ \text{s.t. } & \sum_{i=1}^n w_i = 1, w_i \geq 0, i = 1, \dots, n. \end{aligned} \quad (2.18)$$

Of course, for numerical purposes we need to limit the range for p to a finite interval, and treat all the limiting cases $p \rightarrow \pm\infty$, $p \rightarrow 0$ and $p \rightarrow -1$.

A different situation arises when the parametric form of g is not given. The approach proposed in [Bel03] is based on approximation of g with a monotone linear spline (see [Bel00]), as

$$g(t) = \sum_{j=1}^J c_j B_j(t), \quad (2.19)$$

where B_j are appropriately chosen basis functions, and c_j are spline coefficients. The monotonicity of g is ensured by imposing linear restrictions on spline coefficients, in particular non-negativity, as in [Bel00]. Further, since the generating function is defined up to an arbitrary linear transformation, one has to fix a particular g by specifying two interpolation conditions, like $g(a) = 0$, $g(b) = 1$, $a, b \in]0, 1[$, and if necessary, properly model asymptotic behavior if $g(0)$ or $g(1)$ are infinite.

After rearranging the terms of the sum, the problem of identification becomes (subject to linear conditions on \mathbf{c}, \mathbf{w})

$$\min_{\mathbf{c}, \mathbf{w}} \sum_{k=1}^K \left(\sum_{j=1}^J c_j \left[\sum_{i=1}^n w_i B_j(x_{ik}) - B_j(y_k) \right] \right)^2. \quad (2.20)$$

For a fixed \mathbf{c} (i.e., fixed g) we have a quadratic programming problem to find \mathbf{w} , and for a fixed \mathbf{w} , we have a quadratic programming problem to find \mathbf{c} . However if we consider both \mathbf{c}, \mathbf{w} as variables, we obtain a difficult global optimization problem. We convert it into a bi-level optimization problem

$$\min_{\mathbf{c}} \min_{\mathbf{w}} \sum_{k=1}^K \left(\sum_{j=1}^J c_j \left[\sum_{i=1}^n w_i B_j(x_{ik}) - B_j(y_k) \right] \right)^2, \quad (2.21)$$

where at the inner level we have a QP problem and at the outer level we have a non-linear problem with multiple local minima. When the number of spline coefficients J is not very large (< 10), this problem can be efficiently solved by using deterministic

global optimization methods from Sect. 1.7.4. If the number of variables is small and J is large, then reversing the order of minimization (i.e., using $\min_{\mathbf{w}} \min_{\mathbf{c}}$) is more efficient.

2.4 Other Means

Besides weighted quasi-arithmetic means, there exist very large families of other means, some of which we will mention in this section. A comprehensive reference to the topic of means is [Bul03]. However we must note that not all these means are monotone functions, so technically they are not aggregation functions. Still some members of these families are aggregation functions, and we will mention the sufficient conditions for monotonicity, if available.

In Chap. 7 we will formulate less restrictive monotonicity conditions and will examine some of the mentioned means in the context of weak monotonicity. We will also present generalisations and special cases of some of the means mentioned below in Chap. 6. Most of the mentioned means do not require $\mathbf{x} \in [0, 1]^n$, but we will assume $\mathbf{x} \geq \mathbf{0}$. The values at $\mathbf{x} = \mathbf{0}$ are defined by continuity.

2.4.1 Gini Means

Definition 2.40 (*Lehmer mean*) Let $q \in \mathbb{R}$ and $\mathbf{w} \in \mathbb{R}^n$, $\mathbf{w} \geq \mathbf{0}$. Weighted Lehmer mean is the function

$$L_{\mathbf{w}}^q(\mathbf{x}) = \frac{\sum_{i=1}^n w_i x_i^{q+1}}{\sum_{i=1}^n w_i x_i^q} \quad (2.22)$$

Definition 2.41 (*Gini mean*) Let $p, q \in \mathbb{R}$ and $\mathbf{w} \in \mathbb{R}^n$, $\mathbf{w} \geq \mathbf{0}$. Weighted Gini mean is the function

$$G_{\mathbf{w}}^{p,q}(\mathbf{x}) = \begin{cases} \left(\frac{\sum_{i=1}^n w_i x_i^p}{\sum_{i=1}^n w_i x_i^q} \right)^{1/p-q} & \text{if } p \neq q, \\ \left(\prod_{i=1}^n x_i^{w_i x_i^p} \right)^{1/\sum_{i=1}^n w_i x_i^p} & \text{if } p = q. \end{cases} \quad (2.23)$$

Properties

- $G_{\mathbf{w}}^{p,q} = G_{\mathbf{w}}^{q,p}$, so we assume $p \geq q$;
- $\lim_{p \rightarrow q} G_{\mathbf{w}}^{p,q} = G_{\mathbf{w}}^{q,q}$;

- $\lim_{p \rightarrow \infty} G_{\mathbf{w}}^{p,q}(\mathbf{x}) = \max(\mathbf{x})$;
- $\lim_{q \rightarrow -\infty} G_{\mathbf{w}}^{p,q}(\mathbf{x}) = \min(\mathbf{x})$;
- If $p_1 \leq p_2, q_1 \leq q_2$, then $G_{\mathbf{w}}^{p_1,q_1} \leq G_{\mathbf{w}}^{p_2,q_2}$.

Special cases

- Setting $q = 0$ and $p \geq 0$ leads to weighted power means $G_{\mathbf{w}}^{p,0} = M_{\mathbf{w},[p]}$.
- Setting $p = 0$ and $q \leq 0$ also leads to weighted power means $G_{\mathbf{w}}^{0,q} = M_{\mathbf{w},[q]}$.
- Setting $q = p - 1$ leads to *counter-harmonic* means, also called *Lehmer means*.

For example, when $n = 2$, $G_{(\frac{1}{2}, \frac{1}{2})}^{q+1,q}(x_1, x_2) = \frac{x_1^{q+1} + x_2^{q+1}}{x_1^q + x_2^q}$, $q \in \mathbb{R}$.

- When $q = 1$ we obtain the *contraharmonic* mean $G_{(\frac{1}{2}, \frac{1}{2})}^{2,1}(x_1, x_2) = \frac{x_1^2 + x_2^2}{x_1 + x_2}$.

Note 2.42 Counter-harmonic means (and hence Gini means in general) are not monotone, except in some special cases (e.g., power means) (Figs. 2.12, 2.13, 2.14 and 2.15).

2.4.2 Bonferroni Means

Definition 2.43 (*Bonferroni mean*) [Bon50] Let $p, q \geq 0$ and $\mathbf{x} \geq \mathbf{0}$. Bonferroni mean is the function

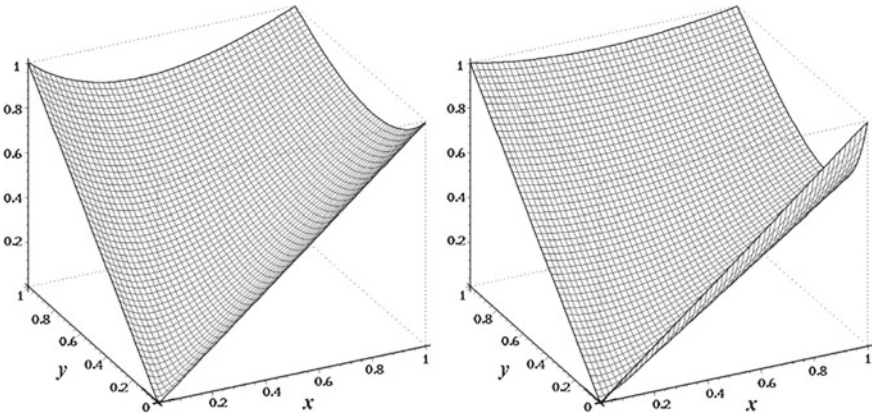


Fig. 2.12 3D plots of weighted Gini means $G_{(\frac{1}{2}, \frac{1}{2})}^{2,1}$ and $G_{(\frac{1}{5}, \frac{4}{5})}^{2,1}$ (both are weighted contraharmonic means). Note lack of monotonicity

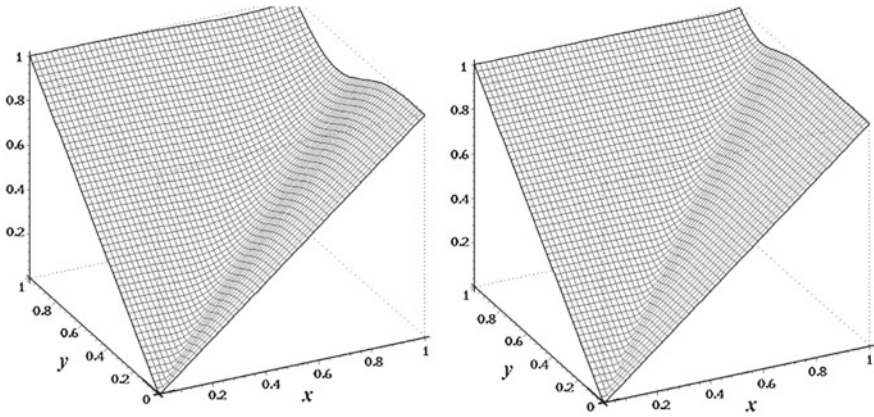


Fig. 2.13 3D plots of weighted Gini means $G_{(\frac{1}{3}, \frac{4}{3})}^{5,4}$ and $G_{(\frac{1}{3}, \frac{4}{3})}^{10,9}$ (both are weighted counter-harmonic means)

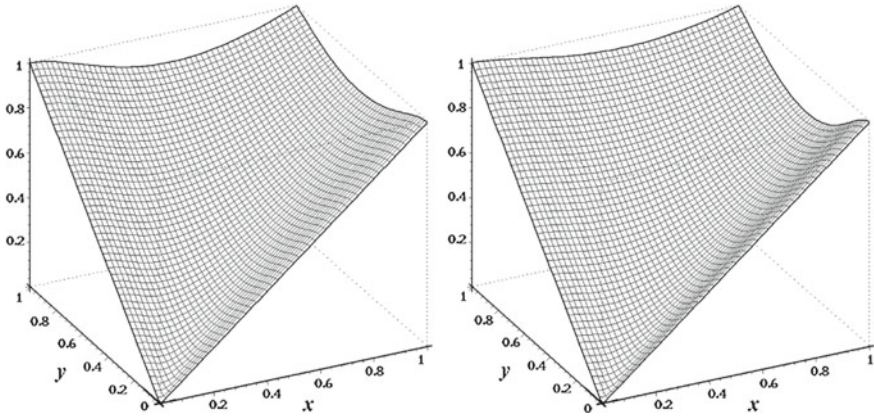


Fig. 2.14 3D plot of weighted Gini means $G_{(\frac{1}{2}, \frac{1}{2})}^{2,2}$ and $G_{(\frac{1}{3}, \frac{2}{3})}^{2,2}$

$$B^{p,q}(\mathbf{x}) = \left(\frac{1}{n(n-1)} \sum_{i,j=1, i \neq j}^n x_i^p x_j^q \right)^{1/(p+q)}. \quad (2.24)$$

Extension to $B^{p,q,r}(\mathbf{x})$ and products of a larger number of inputs is obvious.

It is an aggregation function. We will discuss recent generalisations of the Bonferroni mean in Sect. 6.4 (Figs. 2.16 and 2.17).

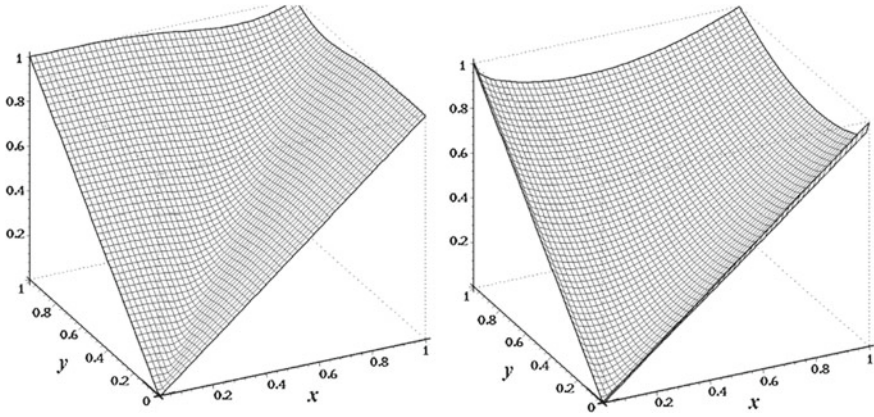


Fig. 2.15 3D plots of weighted Gini means $G_{(\frac{1}{2}, \frac{1}{2})}^{5,5}$ and $G_{(\frac{1}{2}, \frac{1}{2})}^{3,0.5}$

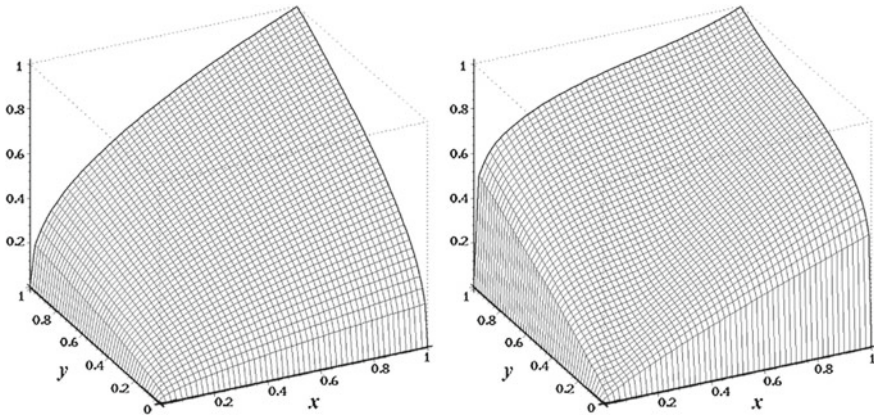


Fig. 2.16 3D plots of Bonferroni means $B^{3,2}$ and $B^{10,2}$

2.4.3 Heronian Mean

Definition 2.44 (*Heronian mean*) Heronian mean is the function

$$HR(\mathbf{x}) = \frac{2}{n(n+1)} \sum_{i=1}^n \sum_{j=i}^n \sqrt{x_i x_j}. \quad (2.25)$$

It is an aggregation function. For $n = 2$ we have $HR = \frac{1}{3}(2M + G)$.

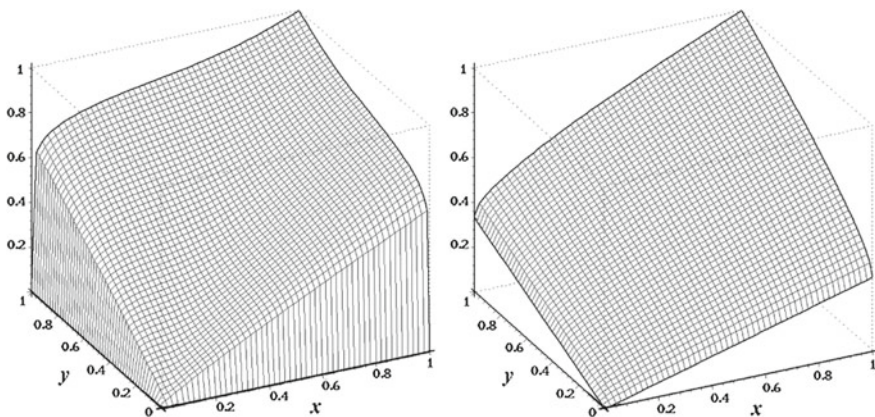


Fig. 2.17 3D plots of Bonferroni mean $B^{5,0.5}$ and the Heronian mean

2.4.4 Generalized Logarithmic Means

Definition 2.45 (*Generalized logarithmic mean*) Let $n = 2$, $x, y > 0$, $x \neq y$ and $p \in [-\infty, \infty]$. The generalized logarithmic mean is the function

$$L^p(x, y) = \begin{cases} \frac{y-x}{\log y - \log x}, & \text{if } p = -1, \\ \frac{1}{e} \left(\frac{y^y}{x^x} \right)^{1/(y-x)}, & \text{if } p = 0, \\ \min(x, y), & \text{if } p = -\infty, \\ \max(x, y), & \text{if } p = \infty, \\ \left(\frac{y^{p+1} - x^{p+1}}{(p+1)(y-x)} \right)^{1/p} & \text{otherwise.} \end{cases} \quad (2.26)$$

For $x = y$, $L^p(x, x) = x$.

Note 2.46 Generalized logarithmic means are also called Stolarsky means, sometimes L^p is called L^{p+1} .

Note 2.47 The generalized logarithmic mean is symmetric. The limiting cases $x = 0$ depend on p , although $L^p(0, 0) = 0$.

Special cases

- The function $L^0(x, y)$ is called *identric* mean;
- $L^{-2}(x, y) = G(x, y)$, the geometric mean;
- L^{-1} is called the logarithmic mean;
- $L^{-1/2}$ is the power mean with $p = -1/2$;
- L^1 is the arithmetic mean;
- Only $L^{-1/2}$, L^{-2} and L^1 are quasi-arithmetic means.

Note 2.48 For each value of p the generalized logarithmic mean is strictly increasing in x, y , hence they are aggregation functions (Figs. 2.18, 2.19 and 2.20).

Note 2.49 Generalized logarithmic means can be extended for n arguments, see Sect. 2.4.9.

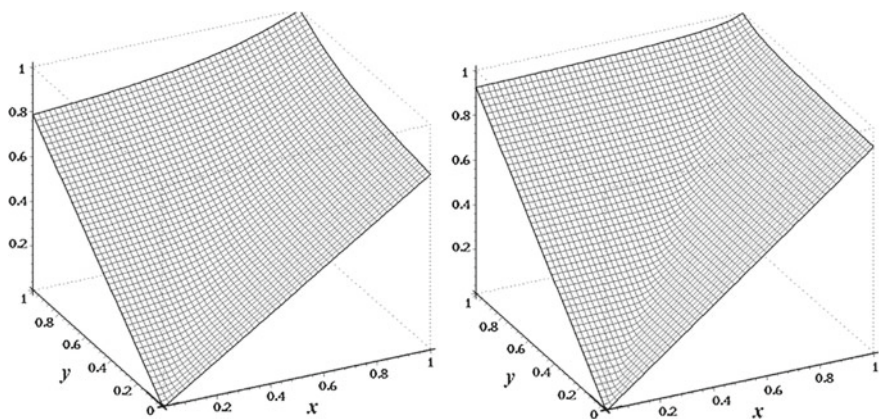


Fig. 2.18 3D plots of generalized logarithmic means L^{10} and L^{50}

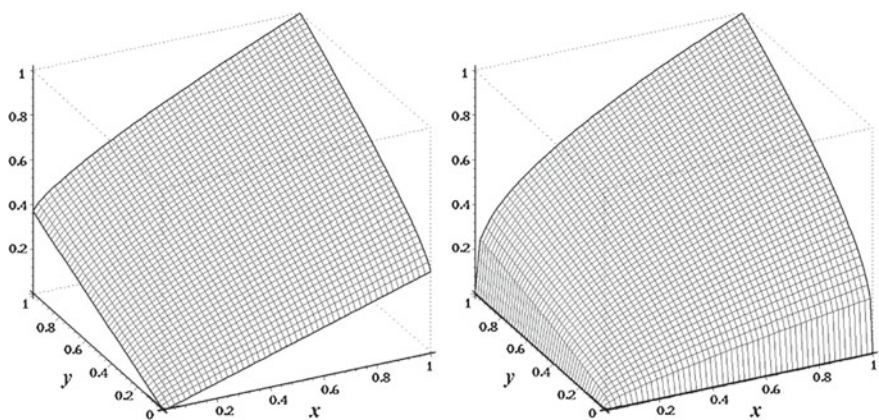


Fig. 2.19 3D plots of generalized logarithmic means L^0 (identric mean) and L^{-1} (logarithmic mean)

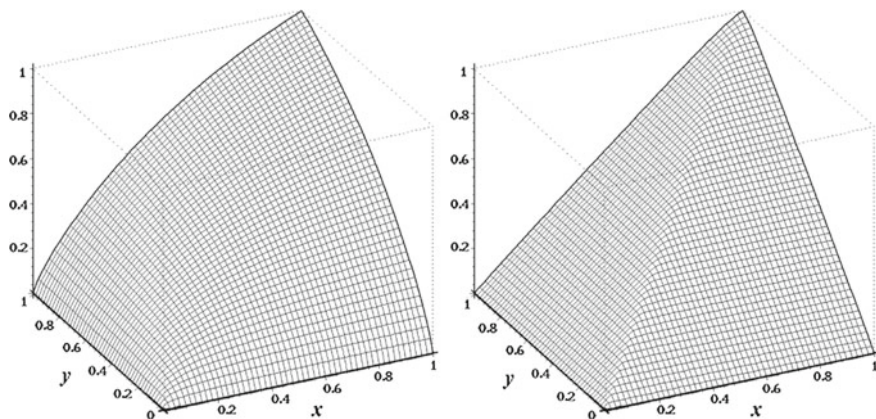


Fig. 2.20 3D plots of generalized logarithmic means L^{-5} and L^{-100}

2.4.5 Cauchy and Lagrangean Means

Definition 2.50 (*Cauchy mean*) Let us take two differentiable functions $g, h : \mathbb{I} \rightarrow \mathbb{R}$ such that $g' \neq 0$ and $\frac{g'}{h'}$ is invertible. Then the Cauchy mean is given for $x \neq y$ by

$$C^{g,h}(x, y) = \left(\frac{g'}{h'} \right)^{-1} \left(\frac{g(x) - g(y)}{h(x) - h(y)} \right). \quad (2.27)$$

For $x = y$ the definition is augmented with $C^{g,h}(x, x) = x$.

The Cauchy means are continuous, symmetric and strictly increasing. The special case of $h = Id$ is called the Lagrangean mean L^g . The generalized logarithmic means are Lagrangean means L^g with $g(t) = t^{p+1}$, $p \neq -1, 0$, $g(t) = \log(t)$ for $p = -1$, and $g(t) = t \cdot \log t$ for $p = 0$.

The Cauchy mean $C^{g,h}$ is a φ -transform of the Lagrangean mean $L^{g \circ h^{-1}}$ with $\varphi = h$. Two Lagrangean means with the generators g and h are the same mean if and only if the generators are related as $h' = ag' + b$, $a, b \in \mathbb{R}$, which is the same relation as for the generators of the quasi-arithmetic means. For this reason we can assume that g is convex.

Some Lagrangean (resp. Cauchy) means are quasi-arithmetic means (e.g., the arithmetic and geometric means), but some are not. For instance the harmonic mean is not Lagrangean, and the logarithmic mean is not quasi-arithmetic. The Cauchy mean $C^{g,h}$ is the quasi-arithmetic mean with the generator h if $h = g'/h'$. Homogeneous Lagrangean means are necessarily generalized logarithmic means. The Lagrangean mean generated by $g(t) = t^{p+1}$ is called Stolarsky mean. The Cauchy mean generated by two power functions $g(t) = t^p$, $h(t) = t^s$ is called the extended mean (sometimes also referred to as Stolarsky mean). For more details about these means refer to [Bul03, Mat11].

2.4.6 Mean of Bajraktarevic

Definition 2.51 (*Mean of Bajraktarevic*) Let $\mathbf{w}(t) = (w_1(t), \dots, w_n(t))$ be a vector of weight functions $w_i : \mathbb{I} \rightarrow [0, \infty[$, and let $g : \mathbb{I} \rightarrow [-\infty, \infty]$ be a strictly monotone function. The mean of Bajraktarevic is the function

$$f(\mathbf{x}) = g^{-1} \left(\frac{\sum_{i=1}^n w_i(x_i) g(x_i)}{\sum_{i=1}^n w_i(x_i)} \right). \quad (2.28)$$

2.4.7 Mixture Functions

The Bajraktarevic mean is also called a *mixture* function [MPR03] when $g(t) = t$. The function g is called the generating function of this mean. If $w_i(t) = w_i$ are constants for all $i = 1, \dots, n$, it reduces to the quasi-arithmetic mean. The special case of Gini mean $G^{p,q}$ is obtained by taking $w_i(t) = w_i t^q$ and $g(t) = t^{p-q}$ if $p > q$, or $g(t) = \log(t)$ if $p = q$.

The mean of Bajraktarevic is not generally an aggregation function because it fails the monotonicity condition. The following sufficient condition for monotonicity of mixture functions has been established in [MPR03].

Let weight functions $w_i(t) > 0$ be differentiable and monotone increasing, and $g(t) = t$. If $w'_i(t) \leq w_i(t)$ for all $t \in \mathbb{I}$ and all $i = 1, \dots, n$, then f in (2.28) is monotone increasing (i.e., an aggregation function).

The other conditions were established in [MS06, MSV08]. Let $\mathbb{I} = [0, 1]$. Then sufficient conditions for monotonicity are: $w(x) \geq w'(x)(1 - x)$ for all $x \in [0, 1]$, or, if we fix the dimension n of the domain, $\frac{w^2(x)}{(n-1)w(1)} + w(x) \geq w'(x)(1 - x)$, $x \in [0, 1]$, $n > 1$.

Analogous results have been obtained for decreasing weighting functions using duality (with respect to the standard negation). Taking the dual weighting function $w^d(x) = w(1 - x)$, the resulting mixture function is the dual to f_w ; that is, $f_{w^d} = 1 - f_w$. Additionally, f_w is invariant to scaling of the weight functions (i.e., $f_{\alpha w} = f_w \forall \alpha \in \mathbb{R} \setminus \{0\}$).

In this context we also state a result which ensures that by making the weights of an averaging function dependent on the arguments, as in mixture functions, we do not change the averaging behaviour of that function.

Proposition 2.52 Let $f_{\mathbf{w}}$ be an averaging aggregation function with constant weighting vector \mathbf{w} . Then the function $f_{\mathbf{u}(\mathbf{x})}$ obtained from $f_{\mathbf{w}}$ by allowing the weights to depend on \mathbf{x} , and using normalisation of weights $w_i = \frac{u_i(\mathbf{x})}{\sum u_i(\mathbf{x})}$ (and assuming $u_i \geq 0$) is bounded by $\min(\mathbf{x}) \leq f_{\mathbf{u}(\mathbf{x})}(\mathbf{x}) \leq \max(\mathbf{x})$.

Proof For every fixed \mathbf{x} the weights $w_i(\mathbf{x}) = \frac{u_i(\mathbf{x})}{\sum u_i(\mathbf{x})}$ are fixed, non-negative and add to one. Then the value of $f_{\mathbf{w}}(\mathbf{x})$ with such weights is bounded by the minimum and maximum, because $f_{\mathbf{w}}$ is averaging (for every choice of \mathbf{w}). \square

2.4.8 Compound Means

In this section we will be talking about bivariate means, although extensions to trivariate and multivariate functions are possible.

Consider the arithmetic and geometric means M and G and take (strictly) positive arguments $x, y > 0$. Consider the sequence defined by $x_0 = x, y_0 = y$ and

$$x_n = G(x_{n-1}, y_{n-1}), \quad y_n = M(x_{n-1}, y_{n-1}), \quad n > 0. \quad (2.29)$$

Definition 2.53 (*Arithmetico-geometric means*) The mean $G \otimes M(x, y)$ is the common value of the limits $\lim_{n \rightarrow \infty} x_n = \lim_{n \rightarrow \infty} y_n$, and it is called the Arithmetico-geometric mean, or AGM.

The iterative process which leads to the sequences $x_n, y_n, n = 0, 1, \dots$ is called Gaussian iterations. A detailed analysis of the AGM is given in [BB87], see also [Bul03]. Among the properties of the AGM we list the following.

- $G \otimes M$ is symmetric;
- $G \otimes M(x, y) = G \otimes M(x_n, y_n)$ for all $n \geq 0$;
- $G \otimes M$ is (positively) homogeneous;
- $G \leq G \otimes M \leq M$.

One can extend the AGM to other means A and B using the following definition.

Definition 2.54 (*Compound means*) Let A and B be bivariate means. Then define the sequences for $x, y > 0$ by $x_0 = x, y_0 = y$ and

$$x_n = A(x_{n-1}, y_{n-1}), \quad y_n = B(x_{n-1}, y_{n-1}), \quad n > 0.$$

If both the limits $\lim_{n \rightarrow \infty} x_n$ and $\lim_{n \rightarrow \infty} y_n$ are equal, then this common value is called the compound mean $A \otimes B$.

Of course, the AGM is the compound mean with $A = G$ and $B = M$. Another interesting case is $H \otimes M = G$. On the other hand $\max \otimes \min$ does not exist (unless $x = y$). If both A and B are symmetric, then $A \otimes B = B \otimes A$ if the compound mean exists. The compound mean is not associative, because $A \otimes (B \otimes B) = A \otimes B$.

The following result establishes the existence of the compound means (see [Bul03], p. 415).

Proposition 2.55 *If two means A and B are continuous and satisfy $\min < A \leq B < \max$ (i.e., the means are strictly internal and comparable), then their compound means exist.*

In particular the compounds of the power means exist.

2.4.9 Extending Bivariate Means to More Than Two Arguments

Some of the means, like the logarithmic mean, are defined for two arguments only. An interesting question is how to extend these definitions to more than two arguments. There are a few methods here, and two are presented in the sequel.

Let us make a change of variables and set $x_i = e^{u_i}$. Recall that

$$e^u = 1 + u + \frac{u^2}{2!} + \frac{u^3}{3!} \cdots$$

We easily obtain for the arithmetic and geometric means respectively

$$M(x_1, x_2) = 1 + \frac{u_1 + u_2}{2} + \frac{u_1^2 + u_2^2}{2 \cdot 2!} + \frac{u_1^3 + u_2^3}{2 \cdot 3!} + \cdots,$$

$$G(x_1, x_2) = \sqrt{e^{u_1} e^{u_2}} = 1 + \frac{u_1 + u_2}{2} + \frac{(u_1 + u_2)^2}{2^2 \cdot 2!} + \frac{(u_1 + u_2)^3}{2^3 \cdot 3!} + \cdots.$$

For the logarithmic mean L^{-1} we obtain

$$\begin{aligned} L^{-1}(x_1, x_2) = \frac{e^{u_1} - e^{u_2}}{u_1 - u_2} &= 1 + \frac{u_1 + u_2}{2} + \frac{u_1^2 + u_1 u_2 + u_2^2}{3 \cdot 2!} \\ &+ \frac{u_1^3 + u_1^2 u_2 + u_1 u_2^2 + u_2^3}{4 \cdot 3!} + \cdots. \end{aligned}$$

The generic form of the polynomial term for $m > 1$ is

$$B_m u_1^m + B_{m-1} u_1^{m-1} u_2 + B_{m-2} u_1^{m-2} u_2^2 + \cdots + B_0 u_2^m$$

divided by the sum of its coefficients $B_m + B_{m+1} + \cdots + B_0$. These coefficients characterize each mean completely. For the arithmetic mean we have $B_0 = B_m = 1$ and the rest are zeros. For the geometric mean $B_i = \binom{m}{i}$, the binomial coefficients, and for the logarithmic mean all $B_i = 1$.

It is proposed to generalize the logarithmic mean by preserving the simple structure of the coefficients B_i [Mus02], so that

$$\begin{aligned}
 L^{-1}(x_1, \dots, x_n) = & 1 + \frac{u_1 + \dots + u_n}{n} \\
 & + \frac{u_1^2 + u_1 u_2 + \dots + u_1 u_n + u_2^2 + \dots + u_n^2}{\binom{n+1}{2} \cdot 2!} + \dots \\
 & + \frac{u_1^m + u_1^{m-1} u_2 + \dots + u_1^{m-1} u_n + \dots + u_n^m}{\binom{n+m-1}{m} \cdot m!} + \dots
 \end{aligned} \tag{2.30}$$

In this series the polynomial term has expression

$$P(n, m) = \sum_{i_1 + i_2 + \dots + i_n = m} u_1^{i_1} u_2^{i_2} \dots u_n^{i_n}$$

so that all coefficients B are equal to 1. The numbers $\binom{n+m-1}{m}$ correspond to the number of summands. Mustonen [Mus02] transformed (2.30) to a closed form

$$L^{-1}(x_1, \dots, x_n) = (n-1)! \sum_{i=1}^n \frac{x_i}{\prod_{j \neq i} \log(x_i/x_j)} \tag{2.31}$$

when all x_i are mutually different positive numbers (see also [Mer04] for a proof). This extension is the same as in [Neu94]. Furthermore, by comparing the above expressions to the divided differences, Mustonen [Mus02] provides a recursive formula

$$L^{-1}(x_1, \dots, x_n) = (n-1) \frac{L^{-1}(x_2, \dots, x_n) - L^{-1}(x_1, \dots, x_{n-1})}{\log x_n/x_1} \tag{2.32}$$

for $n = 2, 3, \dots$

There are alternative formulas for extending the logarithmic mean, for example [Pit85]:

$$L^{-1}(x_1, \dots, x_n) = \left((n-1) \sum_{i=1}^n \frac{x_i^{n-2} \log x_i}{\prod_{j \neq i} \log(x_i/x_j)} \right)^{-1}.$$

The extension based on the method of Mustonen was also applied to the generalized logarithmic mean L^p in (2.26) (and even to a more general case of Stolarsky-Tobey means) by using

$$L^p(x, y) = \left(\frac{L^{-1}(x^{p+1}, y^{p+1})}{L^{-1}(x, y)} \right)^{\frac{1}{p}},$$

and then writing the left-hand side as the ratio of n -variate logarithmic means.

A different approach, which also includes a way of assigning the weights to the inputs, was recently presented in [Duj15] and then extended in [BD15, DB15]. It is based on multiple invocations of the bivariate function in order to approximate the weighted n -variate idempotent function with any desired accuracy. This is similar to recursive functions (Definition 1.41) but invocations of the bivariate function happen in a different order which allows one to introduce a weighting vector.

Suppose we want to approximate (or construct for this matter) a weighted n -variate idempotent function with the weighting vector \mathbf{w} , by using only the unweighted bivariate idempotent function. For this let us fix a number of levels $L \geq 1$ and create an auxiliary vector of arguments \mathbf{X} of size 2^L , in which the inputs from \mathbf{x} are repeated according to the weights. That is, x_1 is taken k_1 times, x_2 is taken k_2 times and so on, and $\frac{k_1}{2^L} \approx w_1$, $\frac{k_2}{2^L} \approx w_2$, ..., and $\sum k_i = 2^L$. One way of doing so is to take $k_i = \lfloor w_i 2^L + \frac{1}{2} \rfloor$, $i = 1, \dots, n-1$ and $k_n = 2^L - k_1 - k_2 - \dots - k_{n-1}$.

This process of replicating each argument x_i a suitable number of times consistent with the weight w_i was presented earlier in [CMY04].

Next, let us build a binary tree with L levels presented in Fig. 2.21, where at each node a value is produced by aggregating the values of two children nodes with the given bivariate averaging function f (denoted by B in this figure). We start from the leaves of the tree which contain the elements of the vector \mathbf{X} . The value at the root node will be the desired output of the n -variate weighted function.

Of course, the output value will depend on the order in which the arguments x_i are processed, and to remove such dependency one can reorder \mathbf{x} (say, in the decreasing order).

It is not difficult to check that the resulting function F is (a) idempotent, and (b) monotone increasing, as long as the bivariate function f is idempotent and monotone increasing. Furthermore, if f is homogeneous and/or shift-invariant, so is the resulting function F , and if f is bounded by two idempotent functions $A \leq f \leq B$, then so will be F . F will have the same absorbing element as f (if any). If f is a

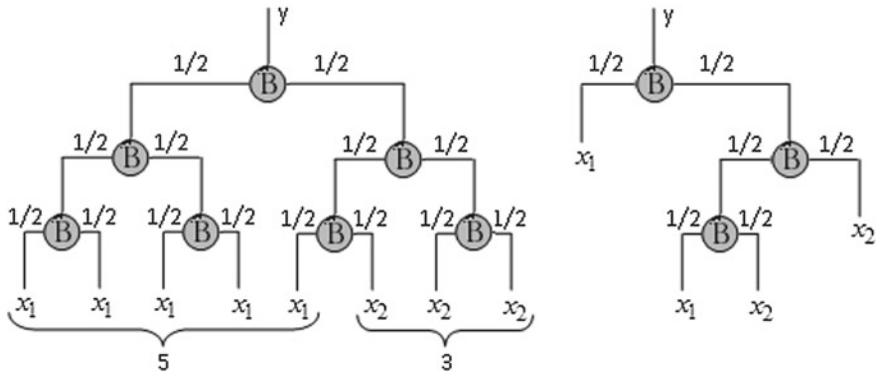


Fig. 2.21 Representation of a weighted arithmetic mean in a binary tree construction. The tree on the right is pruned by using idempotency

quasi-arithmetic mean, then F will be the corresponding n -variate quasi-arithmetic mean. Duality and φ -transforms are also preserved in the binary tree construction. If we select equal weights, so that the numbers k_i are approximately equal (they will actually be equal if n is a power of two), one gets the approximately unweighed extension of the given bivariate function f , and the accuracy of matching the weights is controlled by the number of levels L . For details of the proofs see [BD15, DB15].

A straightforward algorithm for doing so, which starts from the vector \mathbf{X} computed as before, is very simple:

1. Compute $k_i = \lfloor w_i 2^L + \frac{1}{2} \rfloor, i = 1, \dots, n-1$ and $k_n = 2^L - k_1 - k_2 - \dots - k_{n-1}$. create the array $X := (x_1, \dots, x_1, \dots, x_n, \dots, x_n)$ by taking k_1 copies of x_1, k_2 copies of x_2 and so on;
2. $N := 2^L$;
3. Repeat L times:
 - (a) $N := N/2$;
 - (b) For $i = 1 \dots N$ do $X[i] := f(X[2i-1], X[2i])$;
4. Return $X[1]$.

One practical disadvantage of this algorithm is that its computational complexity is $O(2^L)$ in terms of the number of invocations of f . This makes it much slower than an explicit formula like (2.31). However it is possible to appropriately prune the binary tree by relying on idempotency of f . Indeed no invocation of f is necessary if both of its arguments are equal. Such a pruning was presented in [Duj15] in the special case of $n = 2$ where the aim was to construct a weighed bivariate mean from an unweighed one. Below we present a different (and more general) algorithm for the n -variate case whose complexity is $O(L(n-1))$, following [BD15]. This

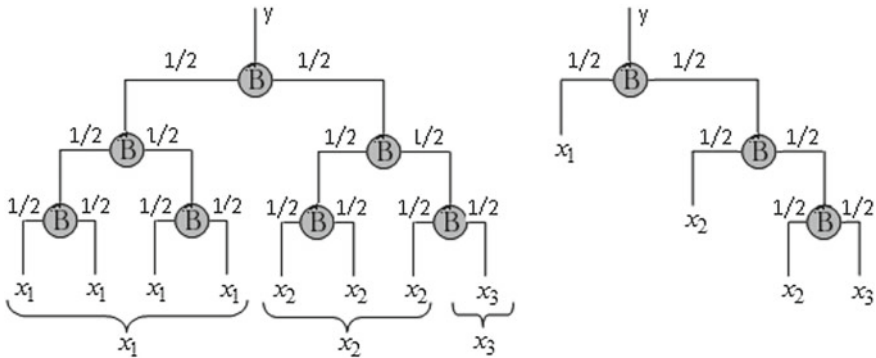


Fig. 2.22 Representation of a weighted 3-variate mean in a binary tree construction. The tree on the right is pruned by using idempotency. The weights $\mathbf{w} = (\frac{1}{2}, \frac{3}{8}, \frac{1}{8})$ are matched exactly

complexity is the upper bound, as at each level of the binary tree one can get at most $n - 1$ nodes with different values of the child nodes, so that pruning is impossible and f must be executed (Fig. 2.22).

The algorithm is a recursive depth-first traversal of the binary tree. A branch is pruned if it is clear that all its leaves have exactly the same value, and by idempotency this is the value of the root node of that branch.

Pruned tree aggregation (PTA) algorithm

function $node(m, N, k, x)$

1. If $N[k] \geq 2^m$ then do:
 - (a) $N[k] := N[k] - 2^m$;
 - (b) $y := x[k]$;
 - (c) If $N[k] = 0$ then $k := k + 1$;
 - (d) return y ;
- else
2. return $f(node(m - 1, N, k, x), node(m - 1, N, k, x))$.

function $PTA(n, w, x, L)$

1. create the array $N = (k_1, k_2, \dots, k_n)$ by using
 $k_i := \lfloor w_i 2^L + \frac{1}{2} \rfloor, i = 1, \dots, n - 1$ and $k_n := 2^L - k_1 - k_2 - \dots - k_{n-1}$;
2. $k := 1$;
3. return $node(L, N, k, x)$.

To see the complexity of this algorithm note that f is never executed if its arguments are the same, i.e., all the branches of the binary tree that can be pruned are pruned. Also note that both N and k are input-output parameters, so that the two arguments of f at step 2 are different as N and k change from one invocation of the function $node$ to another. The C++ code implementing this algorithm is listed in Fig. 2.23.

The PTA algorithm is numerically efficient, and its execution time is comparable to, or could be even smaller than that of analytical formulas like (2.31). In addition, it is a universal method applicable to any bivariate mean, and it allows a very intuitive introduction of weights as multiplicities of the inputs.

```

double node(double x[], long int N[], long int C, int & k,
            double(*F)(double,double))
{
    /* recursive function in the binary tree processing
    Parameters: x - input vector, N vector of multiplicities of x
    m current level of recursion counted from L (root node) to 0 (leaves)
    k - input-output parameter, the current index of x being processed */
    if(N[k]>= C) { /* we use idempotency here to prune the tree */
        N[k] -= C;
        if(N[k]<=0) return x[k++]; else return x[k];
    }
    C /= 2;
    /* tree not pruned, process the children nodes */
    return F( node(x,N,C,k,F), node(x,N,C,k,F) );
}

double nvariatef(double x[], double w[], int n,
                double(*F)(double,double), int L)
/* Function F is the symmetric base aggregator.
w[ ] = array of weights of inputs x[ ], n is the dimension of x and w.
the weights must add to one and be non-negative.
L = number of binary tree levels
Run time = O[(n-1)L] */
{
    long int t=0;
    int k=0;
    long int C=1<<L; /* C=2^m */
    long int N[n]; /* multiplicities of x based on the weights */
    for(int i=0;i<n-1;i++) {
        N[i]=w[i]*C+0.5;
        t+=N[i]; }
    N[n-1]=C-t;
    return node(x,N,C,k,F);
}

...
/* calling the function with f=HR*/
double binaryf(double x, double y) { return (x+y+sqrt(x*y))/3;}
double x[4]={0.2,0.2,0.4,0.8};
double w[4]={0.1,0.2,0.3,0.4};
int n=4, L=4;
double y=nvariatef(x,w,n,&binaryf,L);

```

Fig. 2.23 A C++ implementation of the pruned binary tree algorithm

References

- [Acz48] J. Aczél, On mean values. *Bull. Am. Math. Soc.* **54**, 392–400 (1948)
- [Bel00] G. Beliakov, Shape preserving approximation using least squares splines. *Approx. Theory Appl.* **16**, 80–98 (2000)
- [Bel03] G. Beliakov, How to build aggregation operators from data? *Int. J. Intell. Syst.* **18**, 903–923 (2003)
- [Bel05] G. Beliakov, Learning weights in the generalized OWA operators. *Fuzzy Optim. Decis. Making* **4**, 119–130 (2005)
- [BCJ10] G. Beliakov, T. Calvo, S. James, On Lipschitz properties of generated aggregation functions. *Fuzzy Sets Syst.* **161**, 1437–1447 (2010)
- [BD15] G. Beliakov, J.J. Dujmovic, Extension of bivariate means to weighted means of several arguments by using binary trees, submitted (2015)
- [BJ13] G. Beliakov, S. James, Stability of weighted penalty-based aggregation functions. *Fuzzy Sets Syst.* **226**, 1–18 (2013)
- [BJGRM15] G. Beliakov, S. James, D. Gómez, J.T. Rodríguez, J. Montero, Learning stable weights for data of varying dimension. In: *Proceedings of the AGOP Conference*, Katowice, Poland (2015)
- [BMV04] G. Beliakov, R. Mesiar, L. Valaskova, Fitting generated aggregation operators to empirical data. *Int. J. Uncertainty, Fuzziness Knowl. Based Syst.* **12**, 219–236 (2004)
- [BPC07] G. Beliakov, A. Pradera, T. Calvo, *Aggregation Functions: A Guide for Practitioners*, vol. 221, *Studies in Fuzziness and Soft Computing* (Springer, Berlin, 2007)
- [Bon50] C. Bonferroni, Sulle medie multiple di potenze. *Bollettino Matematica Italiana* **5**, 267–270 (1950)
- [BB87] J.M. Borwein, P.B. Borwein, *PI and the AGM: A Study in Analytic Number Theory and Computational Complexity* (Wiley, New York, 1987)
- [Bul03] P.S. Bullen, *Handbook of Means and Their Inequalities* (Kluwer, Dordrecht, 2003)
- [CKKM02] T. Calvo, A. Kolesárová, M. Komorníková, R. Mesiar, Aggregation operators: properties, classes and construction methods. In: *Aggregation Operators*. *New Trends and Applications*, ed. by T. Calvo, G. Mayor, R. Mesiar (Physica—Verlag, Heidelberg, 2002), pp. 3–104
- [CM99] T. Calvo, G. Mayor, Remarks on two types aggregation functions. *Tatra Mount. Math. Publ.* **16**, 235–254 (1999)
- [CMY04] T. Calvo, R. Mesiar, R.R. Yager, Quantitative weights and aggregation. *IEEE Trans. Fuzzy Syst.* **12**, 62–69 (2004)
- [Cal+00] T. Calvo et al., Generation on weighting triangles associated with aggregation functions. *Int. J. Uncertainty Fuzziness Knowl. Based Syst.* **8**, 417–451 (2000)
- [CM09] M. Cardin, M. Manzi, Supermodular and ultramodular aggregation evaluators. In: *Proceedings of the 5th International Summer School of Aggregation Operators AGOP'09*, July 6–11. Palma de Mallorca, Spain, 2009
- [Cau21] A.L. Cauchy, Cours d'analyse de l'Ecole Royale Polytechnique. *Analyse algébrique* **1** (1821)
- [Chi29] O. Chisini, Sul concetto di media. *Periodico di Matematiche* **4**, 106–116 (1929)
- [DP85] D. Dubois, H. Prade, A review of fuzzy set aggregation connectives. *Inf. Sci.* **36**, 85–121 (1985)
- [DP04] D. Dubois, H. Prade, On the use of aggregation operations in information fusion processes. *Fuzzy Sets Syst.* **142**, 143–161 (2004)
- [Duj15] J.J. Dujmovic, An efficient algorithm for general weighted aggregation. In: *Proceedings of the AGOP Conference*, Katowice, Poland (2015)
- [Duj73] J.J. Dujmovic, Two integrals related to means. *Univ. Beograd. Publ. Elektrotehn. Fak.* (1973)
- [Duj74] J.J. Dujmovic, Weighted conjunctive and disjunctive means and their application in system evaluation. *Univ. Beograd. Publ. Elektrotehn. Fak.* 147–158 (1974)

- [Duj05] J.J. Dujmovic, Seven flavors of andness/orness. In: *EUROFUSE* Belgrade, 2005
- [Duj07] J.J. Dujmovic, Continuous preference logic for system evaluation. *IEEE Trans. Fuzzy Syst.* **15**, 1082–1099 (2007)
- [DB15] J.J. Dujmovic, G. Beliakov, Idempotent weighted aggregation based on binary aggregation trees, submitted (2015)
- [Eis71] C. Eisenhart, The development of the concept of the best mean of a set of measurements from antiquity to the present day. In: *American Statistical Association Presidential Address*, <http://www.york.ac.uk/depts/maths/histstat/eisenhart.pdf> (1971)
- [FSM03] J.M. Fernández Salido, S. Murakami, Extending Yager's orness concept for the OWA aggregators to other mean operators. *Fuzzy Sets Syst.* **139**, 515–542 (2003)
- [FM97] J. Fodor, J.-L. Marichal, On nonstrict means. *Aequationes Mathematicae* **54**, 308–327 (1997)
- [FR94] J. Fodor, M. Roubens, *Fuzzy Preference Modelling and Multicriteria Decision Support* (Kluwer, Dordrecht, 1994)
- [Gin58] C. Gini, *Le Medie*. Milan (Russian translation, Srednie Velichiny, Statistica, Moscow, 1970): Unione Tipografico-Editorial Torinese, 1958
- [GMRR12] D. Gómez, J. Montero, J.T. Rodríguez, K. Rojas, Stability in aggregation operators. In: *Proceedings of IPMU*, Catania, Italy, 2012
- [GRMRB14] D. Gómez, K. Rojas, J. Montero, J.T. Rodríguez, G. Beliakov, Consistency and stability in aggregation operators: an application to missing data problems. *Int. J. Comput. Intell. Syst.* **7**, 595–604 (2014)
- [KSP14] A. Kishor, A.K. Singh, N.R. Pal, Orness measure of OWA operators: a new approach. *IEEE Trans. Fuzzy Syst.* **22**, 1039–1045 (2014)
- [Kom01] M. Komorníková, Aggregation operators and additive generators. *Int. J. Uncertainty Fuzziness Knowl Based Syst.* **9**, 205–215 (2001)
- [Liu06] X. Liu, An orness measure for quasi-arithmetic means. *IEEE Trans. Fuzzy Syst.* **14**, 837–848 (2006)
- [MPR03] R.A. Marques Pereira, R.A. Ribeiro, Aggregation with generalized mixture operators using weighting functions. *Fuzzy Sets Syst.* **137**, 43–58 (2003)
- [Mat11] J. Matkowski, A mean-value theorem and its applications. *J. Math. Anal. Appl.* **373**, 227–234 (2011)
- [MC97] G. Mayor, T. Calvo, Extended aggregation functions. In: *IFSA'97*. vol. 1, Prague, 1997, pp. 281–285
- [Mer04] J. Merikoski, Extending means of two variables to several variables. *J. Inequal. Pure Appl. Math.* **5**, Article 65 (2004)
- [MS06] R. Mesiar, J. Spirková, Weighted means and weighting functions. *Kybernetik* **42**, 151–160 (2006)
- [MSV08] R. Mesiar, J. Spirková, L. Vavříková, Weighted aggregation operators based on minimization. *Inf. Sci.* **178**, 1133–1140 (2008)
- [Mus02] S. Mustonen, *Logarithmic Mean for Several Arguments*, <http://www.survo.fi/papers/logmean.pdf> (2002)
- [Neu94] E. Neuman, The weighted logarithmic mean. *J. Math. Anal. Appl.* **188**, 885–900 (1994)
- [Pit85] A.O. Pittenger, The logarithmic mean in n variables. *Am. Math. Monthly* **92**, 99–104 (1985)
- [PTC02] A. Pradera, E. Trillas, T. Calvo, A general class of triangular norm-based aggregation operators: quasi-linear T-S operators. *Int. J. Approx. Reasoning* **30**, 57–72 (2002)
- [RA14] J.T. Rickard, J. Aisbett, New classes of threshold aggregation functions based upon the Tsallis q -exponential with applications to perceptual computing. *IEEE Trans. Fuzzy Syst.* **22**, 672–684 (2014)
- [RGMR13] K. Rojas, D. Gómez, J. Montero, J.T. Rodríguez, Strictly stable families of aggregation operators. *Fuzzy Sets Syst.* **228**, 44–63 (2013)
- [Rub71] E. Rubin, Quantitative commentary on thucydides. *Am. Stat.* **25**, 52–54 (1971)

- [Tor02] V. Torra, Learning weights for the quasi-weighted means. *IEEE Trans. Fuzzy Systems* **10**, 653–666 (2002)
- [TY05] L. Troiano, R.R. Yager, A measure of dispersion for OWA operators. In: *Proceedings of the 11th IFSA World Congress*, ed. by Y. Liu, G. Chen, M. Ying (Tsinghua University Press, Springer, Beijing China, 2005), pp. 82–87
- [Yag88] R.R. Yager, On ordered weighted averaging aggregation operators in multicriteria decision making. *IEEE Trans. Syst. Man Cybern.* **18**, 183–190 (1988)
- [Yag91] R.R. Yager, Connectives and quantifiers in fuzzy sets. *Fuzzy Sets Syst.* **40**, 39–76 (1991)
- [Yag95] R.R. Yager, Measures of entropy and fuzziness related to aggregation operators. *Inf. Sci.* **82**, 147–166 (1995)

<http://www.springer.com/978-3-319-24751-9>

A Practical Guide to Averaging Functions

Beliakov, G.; Bustince Sola, H.; Calvo Sánchez, T.

2016, XIX, 352 p. 44 illus. in color., Hardcover

ISBN: 978-3-319-24751-9