

Article

Expectation Values and Variance Based on $\mathcal{L}^p$ -Norms

George Livadiotis

Southwest Research Institute, San Antonio, TX-78238, USA; E-Mail: glivadiotis@swri.edu;
Tel.: +1-210-522-3415

Received: 4 August 2012; in revised form: 13 November 2012 / Accepted: 14 November 2012 /
Published: 26 November 2012

Abstract: This analysis introduces a generalization of the basic statistical concepts of expectation values and variance for non-Euclidean metrics induced by $\mathcal{L}^p$ -norms. The non-Euclidean $\mathcal{L}^p$ means are defined by exploiting the fundamental property of minimizing the $\mathcal{L}^p$ deviations that compose the $\mathcal{L}^p$ variance. These $\mathcal{L}^p$ expectation values embody a generic formal scheme of means characterization. Having the p -norm as a free parameter, both the $\mathcal{L}^p$ -normed expectation values and their variance are flexible to analyze new phenomena that cannot be described under the notions of classical statistics based on Euclidean norms. The new statistical approach provides insights into regression theory and Statistical Physics. Several illuminating examples are examined.

Keywords: expectation values; variance; optimization; fitting methods; L_p norms; time series analysis

1. Introduction

2500 years after the Pythagorean discipline and the “self-imposed” intense study on the arithmetic, geometric and harmonic means, the power-means of the elements $\{y_i\}_{i=1}^N$, $y_i \in D_y \subseteq \mathbb{R}$, $\forall i = 1, \dots, N$, given by $M_p(\{y_i\}_{i=1}^N) = (\sum_{i=1}^N y_i^p / N)^{1/p}$, were introduced [1] as suitable generalization of the former Pythagorean ones, for $p = 1, 0, -1$, respectively. An even more general characterization, preceded of the power-means, concerns the Kolmogorov–Nagumo means [2,3], also known as φ -means, which are expressed in terms of a strictly monotonic function φ , as $M_\varphi(\{y_i\}_i^N) = \varphi^{-1}[\sum_{i=1}^N \varphi(y_i) / N]$.

Hardy *et al.* [4] showed that the φ -means are characterized by various fundamental properties of the ordinary arithmetic means. Later, Ben-Tal [5] showed that the φ -means are indeed ordinary arithmetic means defined on linear spaces with suitably chosen operations of addition and multiplication. The latter

justifies the φ -means alternative names of “quasilinear” [4] or “quasiarithmetic” [6]. The series $\{y_i\}_{i=1}^N$ can be rearranged, according to the y -values of its elements, to the set $\{y_k\}_{k=1}^W$, where each value y_k has possibility $\rho_k = \frac{\Delta N_k}{N}$, with ΔN_k being the number of elements $\{y_i\}_{i=1}^N$ that follow the equality $y_i = y_k$ (for examples, see [7]). However, the probability distribution $\{\rho_k\}_{k=1}^W$ can be constructed directly in association with $\{y_k\}_{k=1}^W$, and without the necessity of the series $\{y_i\}_{i=1}^N$. Then, the relation $\rho_k = \rho_k(\{y_k\}_{k=1}^W)$, $\forall k = 1, \dots, W$ can be derived. These weighted φ -means were introduced by [8], and expressed in terms of the probability distribution $\{\rho_k\}_{k=1}^W$, namely, $M_\varphi(\{y_k\}_{k=1}^W) = \varphi^{-1}[\sum_{k=1}^W \rho_k \varphi(y_k)]$, giving advance to the axiomatic theory of information functions [9].

The φ -means are found to have useful applications in a variety of topics, namely, in statistics [10], in decision theory [11], in signal processing [12], in thermostatics [13,14], *etc.* Within the framework of signal processing, [12] succeeded in specifying classes of signals associated with the quasiarithmetic mean of two variables. However, the main efforts to use the φ -means to this topic, are typically addressed in signal denoising processes. For example, by applying the power-means, the moving average shifts towards small signal values for small p and emphasizes large signal values for large p , thus highlighting the fluctuations of the preferred scaling. As soon as the nonlinear function φ is appropriately chosen, φ -means-based filters can efficiently reduce noise at a preferred scale of signal values. Another significant nonlinear filtering scheme, having interesting properties in signal denoising, is based on the moving median, instead of the moving φ -means technique [15,16]. Lately, the median value found to play a key-point role in signal processing optimization and block entropy analysis [7,17]. The median is based on the non-Euclidean norm $\mathcal{L}^1$ (Taxicab norm), and thus, it does not embody the characterization of φ -means, which are in their basis Euclidean, *i.e.*, they are induced by the Euclidean norm $\mathcal{L}^2$ [18].

In [7] a novel generalized characterization of means was introduced, namely, the non-Euclidean means, based on metrics induced by $\mathcal{L}^p$ norms, wherein the median is included as special case for $p = 1$ ($\mathcal{L}^1$), while the non-Euclidean φ -means can also be defined. (See also the work of [19], where the general clustering approaches is investigated using, among others, similar well-defined means based on non-Euclidean optimization.) In this way, the $\mathcal{L}^p$ expectation value of a given energy spectrum $\{\varepsilon_k\}_{k=1}^W$ is defined, representing the non-Euclidean adaptation of internal energy U_p . (This issue is mentioned in Section 3.1; it can be further examined in the framework of non-Euclidean-normed Statistical Mechanics.)

The Euclidean norm $\mathcal{L}^2$ is also known as “Pythagorean” norm. Hereafter, we prefer referring to the non-Pythagorean norms as non-Euclidean, inheriting the same characterization to Statistics. One may adopt the more explicit characterizations of “non-Euclidean-normed” statistics, for avoiding any confusion with the non-Euclidean metric of the (Euclidean-normed) Riemannian geometry.

The paper is organized as follows. In Section 2, the $\mathcal{L}^p$ means are defined, resulting from the minimization of the respective $\mathcal{L}^p$ deviations. Similarly, the more generic Φ -normed φ -means are also defined. The general scheme of means characterization, wherein the $\mathcal{L}^p$ means are embodied, is given. In Section 3, the concept of $\mathcal{L}^p$ -expectation values is thoroughly studied. The $\mathcal{L}^p$ means are expressed in terms of the $\hat{\mathcal{L}}_p$ operator, which helps to automatically retrieve the non-Euclidean representation of a formulation from its respective Euclidean one. The non-Euclidean expectation and variance operators are accordingly deduced. Several examples of Statistical Mechanics are examined: gas in thermal equilibrium, space plasmas out of thermal equilibrium, and multi-dimensional quantum

harmonic oscillator at thermal equilibrium. Section 4 introduces the $\mathcal{L}^p$ variance; its expression is justified and examined in detail. In Section 5, some further analytical and numerical examples are examined. It is shown that in the case of symmetric distributions of data, the whole set of $\mathcal{L}^p$ means degenerate to one single value, while for asymmetric ones, spectrum-like range of $\mathcal{L}^p$ means is generated. In addition, we deal with numerical data of the magnitude of Earth's total magnetic field. Finally, Section 6 draws the conclusions.

2. The Generalized Formal Scheme of Means Characterization

2.1. The Means Characterization Based on Optimization Methods

The non-Euclidean means μ_p , based on $\mathcal{L}^p$ norms, are defined as follows

$$\sum_{k=1}^W \rho_k |y_k - \mu_p|^{p-1} \text{sign}(y_k - \mu_p) = 0 \quad (1)$$

where the median μ_1 and the arithmetic mean μ_2 follow as special cases for the Taxicab $\mathcal{L}^1$ and Euclidean $\mathcal{L}^2$, respectively. Both the median μ_1 and arithmetic μ_2 means can be implicitly written in the form of Equation (1) as $\sum_{k=1}^W \rho_k \text{sign}(y_k - \mu_1) = 0$ and $\sum_{k=1}^W \rho_k |y_k - \mu_2| \text{sign}(y_k - \mu_2) = 0$ ($\Leftrightarrow \mu_2 = \sum_{k=1}^W \rho_k y_k$), respectively. These forms can be thought as the result of the corresponding optimizations of the Total Absolute Deviations $TAD(\{y_k\}_{k=1}^W; \alpha) = \sum_{k=1}^W \rho_k |y_k - \alpha|$ (also called Total Boundary Area - TBA), and the Total Square Deviations $TSD(\{y_k\}_{k=1}^W; \alpha)^2 = \sum_{k=1}^W \rho_k |y_k - \alpha|^2$, for describing the optimal values $\alpha^* = \mu_1$ and $\alpha^* = \mu_2$, respectively [7,18,20]. Similarly, the $\mathcal{L}^p$ mean μ_p emerges from minimizing the Total p -Deviations, i.e.,

$$TD_p(\{y_k\}_{k=1}^W; \alpha; p)^p = \sum_{k=1}^W \rho_k |y_k - \alpha|^p \quad (2)$$

The optimization leads to (i) the normal Equation (1), from which the optimal parameter $\alpha^* = \mu_p$ can be derived as an implicit expression of p , (ii) the total deviations minimum value $TD_{p,Min}(\{y_k\}_{k=1}^W; p)^p = \sum_{k=1}^W \rho_k |y_k - \alpha^*|^p$, which is the p th absolute moment at the $\mathcal{L}^p$ mean $\alpha^* = \mu_p$. Therefore, the Euclidean mean μ_2 , known as the one minimizing the (Euclidean) variance, is generalized to the $\mathcal{L}^p$ mean μ_p , which is the one minimizing the $\mathcal{L}^p$ variance, a quantity proportional to $TD_p(\{y_k\}_{k=1}^W; \alpha; p)^p$ (Sections 3.1,4).

The generalized $\mathcal{L}^p$ φ -means $\mu_{(\varphi,p)}$ can be defined given the strictly monotonic function $\varphi : y_k \rightarrow z_k = \varphi(y_k), \forall k = 1, \dots, W$, i.e.,

$$\sum_{k=1}^W \rho_k |\varphi(y_k) - \varphi(\mu_{(\varphi,p)})|^{p-1} \text{sign}[\varphi(y_k) - \varphi(\mu_{(\varphi,p)})] = 0 \quad (3)$$

which is the normal equation derived from the Total (φ, p) -Deviations,

$$TD_{(\varphi,p)}(\{y_k\}_{k=1}^W; \alpha; p)^p = \sum_{k=1}^W \rho_k |\varphi(y_k) - \varphi(\alpha)|^p \quad (4)$$

Thereafter, we arrive at the classical φ -means $\mu_{(\varphi, p=2)} = M_\varphi$ by considering the Euclidean norm. Even further, by considering an arbitrary functional norm $\Phi : u \rightarrow \Phi(|u|)$ ($\Phi(0) = 0$), the Total (φ, Φ) -Deviations are formulated

$$\Phi \left[TD_{(\varphi, \Phi)}(\{y_k\}_{k=1}^W; \alpha) \right] = \sum_{k=1}^W \rho_k \Phi \left[|\varphi(y_k) - \varphi(\alpha)| \right] \quad (5)$$

leading to the normal equation

$$\sum_{k=1}^W \rho_k \Phi \left[|\varphi(y_k) - \varphi(\mu_{(\varphi, \Phi)})| \right] \text{sign}[\varphi(y_k) - \varphi(\mu_{(\varphi, \Phi)})] = 0 \quad (6)$$

from which we obtain the Φ -normed φ -means $\mu_{(\varphi, \Phi)}$. It is noted that the solution of Equation (5) is generally called M -estimator, a broad class of estimators, which are obtained by minimizing the sums of functions of data. An M -estimator can be defined to be a zero of an estimating function that often is the derivative of another statistical function. When this differentiation is possible, leading to Equation (6), the M -estimator is said to be of ψ -type. (For more on M -estimators theory, see [21,22].)

2.2. Formal Scheme of Means Characterization

The characterization of means based on optimization methods has already been found useful, providing insights into the optimization theory and its applications (e.g., see [7]). The most important application involves reestablishing the concept of expectation values. However, we have to be certain that the optimization characterization of means, as described in Section 2.1, can embody the general scheme of means characterization.

Aczél [23] suggested an axiomatic characterization of means, settled by five fundamental properties of the ordinary arithmetic means, and succeeded to reproduce the φ -means. In particular, any univalued, bivariable function $M(y_1, y_2)$, $y_1, y_2 \in D_y$ constitutes a general mean of y_1, y_2 , if the following preconditions are fulfilled: (i) Continuity; (ii) Strict monotonicity: if $y_1 < y_1' (>)$, then $M(y_1, y_2) < M(y_1', y_2) (>)$, and the same holds for $y_2 < y_2'$; (iii) Bisymmetry: $M(M(y_1, y_2), M(y_3, y_4)) = M(M(y_1, y_3), M(y_2, y_4))$; (iv) Reflexivity: $M(y, y) = y$; (v) Symmetry: $M(y_1, y_2) = M(y_2, y_1)$. Aczél also proved [24] that, if and only if the (i)–(iv) preconditions are fulfilled, then $M(y_1, y_2) = \varphi^{-1}[\lambda\varphi(y_1) + (1-\lambda)\varphi(y_2)]$, $\forall y_1, y_2 \in D_y$, $0 < \lambda < 1$, and φ : strictly monotonic function. For $\lambda = 1/2$, (v) is also fulfilled, hence, $M(y_1, y_2)$ coincides with the φ -mean of y_1, y_2 , i.e., $M(y_1, y_2) = M_\varphi(y_1, y_2)$.

Aczél also pointed [23] that the “internness” property, i.e., $\text{Min}(y_1, y_2) \leq M(y_1, y_2) \leq \text{Max}(y_1, y_2)$, follows from preconditions (i), (ii), (iv). However, it is evident that the internness, together with the continuity (i), leads to the reflexivity (iv). This remark implies that we can settle the internness as precondition, instead of the strict monotonicity and reflexivity. (Besides, the median does not follow strict monotonicity, while it follows the internness. In addition, internness ensures that the mean shall preserve the units of y -values under units-transformations.) Then, turning aside of preconditions (ii) and (iii) that entail the Euclidean character of φ -means, an alternative characterization of means can be given by the univalued, N -multivariable function $M(\{y_i\}_{i=1}^N)$, fulfilling the three preconditions: (i) Continuity; (ii) Internness: $\text{Min}(\{y_i\}_{i=1}^N) \leq M(\{y_i\}_{i=1}^N) \leq \text{Max}(\{y_i\}_{i=1}^N)$; (iii) Symmetry: For $y_i \rightarrow y_{j_i}, \forall i =$

$1, \dots, N$: $\{y_i\}_{i=1}^N = \{y_{j_i}\}_{i=1}^N$, then $M(\{y_i\}_{i=1}^N) = M(\{y_{j_i}\}_{i=1}^N)$. (This characterization considered also by [25] for $N = 2$.)

It can be easily verified that the non-Euclidean, Φ -normed, φ -means $\mu_{(\Phi, \varphi)}$ obey to the above characterization. Throughout, we deal with the $\mathcal{L}^p$ means μ_p . The uniqueness of μ_p means can be provided for any $p \geq 1$. (The restriction $p \geq 1$ is required for the triangle inequality of the norm's definition to hold.) In particular, the non-zero derivative

$$\frac{\partial p}{\partial \mu_p} = \frac{(p-1) \sum_{k=1}^W \rho_k |y_k - \mu_p|^{p-2}}{\sum_{k=1}^W \rho_k |y_k - \mu_p|^{p-1} \text{sign}(y_k - \mu_p) \ln(|y_k - \mu_p|)}, \forall p > 1 \quad (7)$$

and the fact that the derivatives of any order of $\mu_p(p)$, inductively, are analytically expressed in terms of the (univalued) Euclidean mean $\mu_p(p)|_{p=2} = \mu_2$, can ensure for the uniqueness of μ_p , for a given $p > 1$ (for $p = 1$, see [7]). Note that for the specific case where $\exists k = k' : y_{k'} = \mu_p$, then for $p < 2$, Equation (7) gives $\partial p / \partial \mu_p \simeq (p-1) / [(y_{k'} - \mu_p) \ln(|y_{k'} - \mu_p|)] \rightarrow \infty$, i.e., $\partial \mu_p / \partial p \rightarrow 0$.

3. The Concept of $\mathcal{L}^p$ -Expectation Values

3.1. The Non-Euclidean Norm Operator $\hat{\mathcal{L}}_p$

The non-Euclidean $\mathcal{L}^p$ -expectation value $\langle y \rangle_p$ is implicitly given by

$$\sum_{k=1}^W \rho_k |y_k - \langle y \rangle_p|^{p-1} \text{sign}(y_k - \langle y \rangle_p) = 0 \quad (8)$$

and it is apparent that the most of the fundamental properties of the Euclidean expectation value are not inherited by the non-Euclidean expectation values. In particular, we distinguish among others, the following two Euclidean properties: (i) $\langle y \rangle_2 = \sum_{k=1}^W \rho_k y_k$ (by definition), and (ii) $\frac{\partial}{\partial \beta} \langle y \rangle_2 = \sum_{k=1}^W \frac{\partial \rho_k}{\partial \beta} y_k$, for a given parameter β , for which $\rho_k = \rho_k(\{y_{k'}\}_{k'=1}^W; \beta)$, $\forall k = 1, \dots, W$, and $\langle y \rangle_p = \langle y \rangle_p(\{y_k\}_{k=1}^W; \beta)$. Now, we examine whether the above two properties can be fulfilled even for the case of non-Euclidean $\mathcal{L}^p$ -expectation value $\langle y \rangle_p$, and a suitably transformed data set $\{y_k^{(p)}\}_{k=1}^W$, namely, (i) $\langle y \rangle_p = \sum_{k=1}^W \rho_k y_k^{(p)}$, and (ii) $\frac{\partial}{\partial \beta} \langle y \rangle_p = \sum_{k=1}^W \frac{\partial \rho_k}{\partial \beta} y_k^{(p)}$. Indeed, this is true for the specific transformation $\hat{\mathcal{L}}_p : y_k \rightarrow y_k^{(p)} \equiv \hat{\mathcal{L}}_p(y_k)$, $\forall k = 1, \dots, W$, where $\hat{\mathcal{L}}_p$ denotes the non-Euclidean norm operator, i.e.,

$$\hat{\mathcal{L}}_p(y_k) = \frac{|y_k - \langle y \rangle_p|^{p-1} \text{sign}(y_k - \langle y \rangle_p)}{(p-1)\phi_p} + \langle y \rangle_p \quad (9)$$

where $\phi_p \equiv \sum_{k=1}^W \rho_k |y_k - \langle y \rangle_p|^{p-2}$. Finally, $\hat{\mathcal{L}}_p$ exhibits the following properties:

(i) The non-Euclidean mean of $\{y_k\}_{k=1}^W$ is the Euclidean mean of $\{\hat{\mathcal{L}}_p(y_k)\}_{k=1}^W$

$$\langle y \rangle_p = \langle \hat{\mathcal{L}}_p(y) \rangle_2 = \sum_{k=1}^W \rho_k \hat{\mathcal{L}}_p(y_k) \quad (10)$$

(ii) Zero-mean of $\{\frac{\partial}{\partial \beta} \hat{\mathcal{L}}_p(y_k)\}_{k=1}^W$,

$$0 = \sum_{k=1}^W \rho_k \frac{\partial}{\partial \beta} \hat{\mathcal{L}}_p(y_k), \text{ or, } \frac{\partial}{\partial \beta} \langle y \rangle_p = \sum_{k=1}^W \frac{\partial \rho_k}{\partial \beta} \hat{\mathcal{L}}_p(y_k) \quad (11)$$

(iii) Norm-derivative (Equation (7)):

$$\frac{\partial \langle y \rangle_p}{\partial p} = \sum_{k=1}^W \frac{\partial \rho_k}{\partial p} \hat{\mathcal{L}}_p(y_k) + \sum_{k=1}^W \rho_k \hat{\mathcal{L}}_p(y_k - \langle y \rangle_p) \ln |y_k - \langle y \rangle_p| \quad (12)$$

$$\Rightarrow \frac{\partial \langle y \rangle_p}{\partial p} = \sum_{k=1}^W \rho_k \hat{\mathcal{L}}_p(y_k - \langle y \rangle_p) \ln |y_k - \langle y \rangle_p|, \text{ if } \frac{\partial \rho_k}{\partial p} = 0 \quad (13)$$

(iv) In the Euclidean case, $\hat{\mathcal{L}}_p$ degenerates to the identity operator $\hat{\mathcal{L}}_{p=2} = \hat{1}$.

(v) Linear operations: $\hat{\mathcal{L}}_p(\lambda y_k + c) = \lambda \hat{\mathcal{L}}_p(y_k) + c, \forall \lambda, c \in \mathfrak{R}$.

Hence, $\langle y - \langle y \rangle_p \rangle_p = \langle \hat{\mathcal{L}}_p(y - \langle y \rangle_p) \rangle_2 = 0$, which reads Equation (8).

(vi) Non-additivity of $\hat{\mathcal{L}}_p$: $\hat{\mathcal{L}}_p(y_k + z_k) \neq \hat{\mathcal{L}}_p(y_k) + \hat{\mathcal{L}}_p(z_k)$.

(vii) Inverse of non-Euclidean norm operator, $\hat{\mathcal{L}}_p^{-1}, \hat{\mathcal{L}}_p^{-1} \hat{\mathcal{L}}_p = \hat{\mathcal{L}}_p \hat{\mathcal{L}}_p^{-1} = \hat{1}$,

$$\hat{\mathcal{L}}_p^{-1}(y_k) = \frac{|y_k - \langle y \rangle_p|^{\frac{1}{p-1}} \text{sign}(y_k - \langle y \rangle_p)}{[(p-1)\phi_p]^{\frac{1}{1-p}}} + \langle y \rangle_p \quad (14)$$

As we mentioned in property (v), Equation (8) can be rewritten in the form $\langle \hat{\mathcal{L}}_p(y - \langle y \rangle_p) \rangle_2 = 0$, or $C \cdot \sum_{k=1}^W \rho_k |y_k - \langle y \rangle_p|^{p-1} \text{sign}(y_k - \langle y \rangle_p) = 0$, with $C = \frac{1}{(p-1)\phi_p}$. Obviously, this leads to Equation (8), for any value of the scalar $C \neq 0$. However, property (ii) is fulfilled if and only if C has the specific expression $C = \frac{1}{(p-1)\phi_p}$ (see further below). Property (ii) is important when we incorporate the non-Euclidean-normed Statistics in Statistical Mechanics: (a) The Canonical probability distribution can be automatically derived and explicitly expressed. (If the scalar C were expressed by any other formulation, after the extremization of entropy in the Canonical Ensemble, we would not be able to solve in terms of the probability, namely, to express explicitly the probability in terms of the energy.) (b) The basic relation that connects Statistical Mechanics with Thermodynamics has to remain the same with the classical case. Namely, the classical relation between the derivative of the partition function $\mathcal{Z}_p$ and the mean energy (internal energy) U_p has to remain invariant, independently of the p -norm. In other words, if $\hat{\mathcal{L}}_p(\varepsilon_k) = C|\varepsilon_k - U_p|^{p-1} \text{sign}(\varepsilon_k - U_p) + U_p$, then the specific expression of the scalar $C = \frac{1}{(p-1)\phi_p}$, yields the following scheme:

$$U_p = -\frac{\partial \ln \mathcal{Z}_p}{\partial \beta} \Leftrightarrow C = \frac{1}{(p-1)\phi_p} \Leftrightarrow \sum_{k=1}^W \rho_k \frac{\partial}{\partial \beta} \hat{\mathcal{L}}_p(\varepsilon_k) = 0 \quad (15)$$

where the non-Euclidean $\mathcal{L}^p$ expectation value of energy states $\{\varepsilon_k\}_{k=1}^W$ yields the internal energy $\langle \varepsilon \rangle_p = U_p$, while the $\mathcal{L}^p$ Canonical partition function $\mathcal{Z}_p$ (in Boltzmann–Gibbs Statistical Mechanics) is given by $\mathcal{Z}_p(\{\varepsilon_k\}_{k=1}^W) = \sum_{k=1}^W \exp[-\beta \hat{\mathcal{L}}_p(\varepsilon_k)]$. Indeed, we have

$$-\frac{\partial \ln \mathcal{Z}_p}{\partial \beta} = \frac{\partial(C\beta)}{\partial \beta} \sum_{k=1}^W \rho_k |\varepsilon_k - U_p|^{p-1} \text{sign}(\varepsilon_k - U_p) + U_p + \left[1 - C(p-1) \sum_{k=1}^W \rho_k |\varepsilon_k - U_p|^{p-2} \right] \beta \frac{\partial U_p}{\partial \beta} = U_p + [1 - C(p-1)\phi_p] \beta \frac{\partial U_p}{\partial \beta},$$

which equals U_p , if and only if $C = \frac{1}{(p-1)\phi_p}$. On the other hand, we have

$$\sum_{k=1}^W p_k \frac{\partial}{\partial \beta} \hat{\mathcal{L}}_p(\varepsilon_k) = \frac{\partial C}{\partial \beta} \sum_{k=1}^W p_k |\varepsilon_k - U_p|^{p-1} \text{sign}(\varepsilon_k - U_p) + \frac{\partial U_p}{\partial \beta} \times$$

$$\left[1 - C(p-1) \sum_{k=1}^W p_k |\varepsilon_k - U_p|^{p-2} \right] = \frac{\partial U_p}{\partial \beta} [1 - C(p-1)\phi_p],$$

which equals zero, leading to Equation (11), if and only if $C = \frac{1}{(p-1)\phi_p}$.

Given the set of states $\{y_k\}_{k=1}^W$ of a physical quantity, which are included in the Euclidean representation of a formulation $F[\{y_k\}_{k=1}^W]$, the respective non-Euclidean representation can be often retrieved automatically by replacing y_k with $\hat{\mathcal{L}}_p(y_k)$, $\forall k = 1, \dots, W$, i.e., $F_p[\{y_k\}_{k=1}^W] = F[\{\hat{\mathcal{L}}_p(y_k)\}_{k=1}^W]$. As an example, the $\mathcal{L}^p$ Canonical partition function in Boltzmann–Gibbs Statistical Mechanics is given by

$$\mathcal{Z}(\{\varepsilon_k\}_{k=1}^W) \equiv \sum_{k=1}^W \exp(-\beta \varepsilon_k) \Rightarrow \mathcal{Z}_p(\{\varepsilon_k\}_{k=1}^W) = \mathcal{Z}(\{\hat{\mathcal{L}}_p(\varepsilon_k)\}_{k=1}^W),$$

where $\{\varepsilon_k\}_{k=1}^W$ is a discrete energy spectrum.

Consequently, according to the above considerations, the non-Euclidean $\mathcal{L}^p$ -expectation estimator $\hat{E}_p$, acting on a random variable Y , is given by

$$\hat{E}_p(Y) = \hat{E}_2 \left[\hat{\mathcal{L}}_p(Y) \right] \quad (16)$$

where $\hat{E}_2 (\equiv \hat{E})$ is the classical (Euclidean) expectation estimator.

On the other hand, the non-Euclidean $\mathcal{L}^p$ -variance estimator $\hat{\sigma}_p^2$ has to result to the Total p -Deviations of Equation (2), or at least, to be proportional to that, such that its minimization leads to Equation (1). This can be achieved by setting,

$$\hat{\sigma}_p^2 = \hat{E}_2 \left[(Y - \langle Y \rangle_p) \hat{\mathcal{L}}_p(Y - \langle Y \rangle_p) \right] \quad (17)$$

(Section 4 revisits the concept of non-Euclidean $\mathcal{L}^p$ -variance, providing a convincing and consistent justification of Equation (17).) Finally, given a set of random variables $\{Y_i\}_{i=1}^n$, the non-Euclidean covariance estimator $(\hat{\sigma}_p^2)_{ij}$ is given by

$$(\hat{\sigma}_p^2)_{ij} = \hat{E}_2 \left[(Y_i - \langle Y_i \rangle_p) \hat{\mathcal{L}}_p(Y_j - \langle Y_j \rangle_p) \right] \quad (18)$$

3.2. The Non-Euclidean $\mathcal{L}^p$ -Mean Estimator and Its Expectation Value

Two basic theorems of the ordinary expectation values are inherited to the non-Euclidean ones: (1) The $\mathcal{L}^p$ -expectation value of the $\mathcal{L}^p$ -mean estimator $\hat{\mu}_{p,N}(\{y_j\}_{j=1}^N; p)$ is equal to the $\mathcal{L}^p$ -expectation value μ_p of any of the independent and identically distributed random variables $\{y_j\}_{j=1}^N$. (2) The $\mathcal{L}^p$ -mean estimator $\hat{\mu}_{p,N}$ converges to its $\mathcal{L}^p$ -expectation value, $\langle \hat{\mu}_{p,N} \rangle_p = \hat{E}_p(\hat{\mu}_{p,N}) = \mu_p$, as $N \rightarrow \infty$.

Given the sampling $\{y_i\}_{i=1}^N$, the non-Euclidean, $\mathcal{L}^p$ -mean estimator $\hat{\mu}_{p,N} = \hat{\mu}_{p,N}(\{y_j\}_{j=1}^N; p)$ is implicitly expressed by

$$\sum_{i=1}^N |y_i - \hat{\mu}_{p,N}(\{y_j\}_{j=1}^N; p)|^{p-1} \text{sign} [y_i - \hat{\mu}_{p,N}(\{y_j\}_{j=1}^N; p)] = 0 \quad (19)$$

Then, the non-Euclidean $\mathcal{L}^p$ -expectation value of $\hat{\mu}_{p,N}$, namely, $\langle \hat{\mu}_{p,N} \rangle_p \equiv \hat{E}_p(\hat{\mu}_{p,N}(\{y_j\}_{j=1}^N; p))$, is implicitly given by

$$\int \cdots \int_{\{y_j \in D_y\}_{j=1}^N} |\hat{\mu}_{p,N}(\{y_j\}_{j=1}^N; p) - \langle \hat{\mu}_{p,N} \rangle_p|^{p-1} \text{sign} [\hat{\mu}_{p,N}(\{y_j\}_{j=1}^N; p) - \langle \hat{\mu}_{p,N} \rangle_p] \times \mathfrak{L}(\{y_j\}_{j=1}^N) dy_1 \cdots dy_N = 0 \quad (20)$$

where $\mathfrak{L}(\{y_j\}_{j=1}^N)$ is the normalized joint probability density, so that

$$\int \cdots \int_{\{y_j \in D_y\}_{j=1}^N} \mathfrak{L}(\{y_j\}_{j=1}^N) dy_1 \cdots dy_N = 1 \quad (21)$$

Consider the sampling $\{y_i\}_{i=1}^N$, $y_i \in D_y \subseteq \mathbb{R}, \forall i = 1, \dots, N$, of the symmetrically distributed random variables $\{Y_i\}_{i=1}^N$. Namely, the joint distribution density has the property $\mathfrak{L}(\{y_j\}_{j=1}^N) = \mathfrak{L}(y_1 \dots y_k \dots y_i \dots y_N) = \mathfrak{L}(y_1 \dots y_i \dots y_k \dots y_N), \forall i, k (\neq i) = 1, \dots, N$.

Lemma 1: The symmetrically distributed random variables $\{Y_i\}_{i=1}^N$ are characterized by the same non-Euclidean $\mathcal{L}^p$ -expectation value, namely, $\langle Y_i \rangle_p = \hat{E}_p(Y_i) = \mu_p \in \mathbb{R}, \forall i = 1, \dots, N$, implicitly given by

$$\begin{aligned} \int \cdots \int_{\{y_j \in D_y\}_{j=1}^N} |y_i - \mu_p|^{p-1} \text{sign}(y_i - \mu_p) \mathfrak{L}(\{y_j\}_{j=1}^N) dy_1 \cdots dy_N = \\ = \int_{y_i \in D_y} |y_i - \mu_p|^{p-1} \text{sign}(y_i - \mu_p) \mathfrak{L}_y(y_i) dy_i = 0 \end{aligned} \quad (22)$$

where $\mathfrak{L}_{y_i}(u) \equiv \mathfrak{L}_y(u), \forall i = 1, \dots, N$, is the marginal distribution density, which is identical for all the random variables $\{Y_i\}_{i=1}^N$.

Lemma 2: Let the auxiliary functionals $\{G_i\}_{i=1}^N$, with $G_i = G_i(\{y_j\}_{j=1}^N; p) \equiv y_i - \hat{\mu}_{p,N}(\{y_j\}_{j=1}^N; p), \forall i = 1, \dots, N$. Then, their $\mathcal{L}^p$ -expectation values are zero, namely, $\langle G_i \rangle_p = \hat{E}_p(G_i) = 0, \forall i = 1, \dots, N$.

Theorem 1: Consider the sampling $\{y_i\}_{i=1}^N, y_i \in D_y \subseteq \mathbb{R}, \forall i = 1, \dots, N$, of the symmetrically distributed random variables $\{Y_i\}_{i=1}^N$. According to *Lemma 1*, the random variables are characterized by the same non-Euclidean $\mathcal{L}^p$ -expectation value, namely, $\langle Y_i \rangle_p = \hat{E}_p(Y_i) = \mu_p \in \mathbb{R}, \forall i = 1, \dots, N$, which is implicitly expressed by Equation (22). Then, the $\mathcal{L}^p$ -expectation value of the $\mathcal{L}^p$ -mean estimator $\hat{\mu}_{p,N}(\{y_j\}_{j=1}^N; p)$ is equal to μ_p , i.e., $\langle \hat{\mu}_{p,N} \rangle_p = \hat{E}_p(\hat{\mu}_{p,N}(\{y_j\}_{j=1}^N; p)) = \mu_p$:

$$\begin{aligned} \int \cdots \int_{\{y_j \in D_y\}_{j=1}^N} |\hat{\mu}_{p,N}(\{y_j\}_{j=1}^N; p) - \mu_p|^{p-1} \text{sign} [\hat{\mu}_{p,N}(\{y_j\}_{j=1}^N; p) - \mu_p] \\ \times \mathfrak{L}(\{y_j\}_{j=1}^N) dy_1 \cdots dy_N = 0 \end{aligned} \quad (23)$$

Theorem 2: Consider the sampling $\{y_i\}_{i=1}^N, y_i \in D_y \subseteq \mathbb{R}, \forall i = 1, \dots, N$, of the independent and identically distributed random variables $\{Y_i\}_{i=1}^N$. The $\mathcal{L}^p$ -mean estimator $\hat{\mu}_{p,N}(\{y_j\}_{j=1}^N; p)$ converges to its $\mathcal{L}^p$ -expectation value, $\langle \hat{\mu}_{p,N} \rangle_p = \hat{E}_p[\hat{\mu}_{p,N}(\{y_j\}_{j=1}^N; p)] = \mu_p$, as $N \rightarrow \infty$, namely, $\lim_{N \rightarrow \infty} \hat{\mu}_{p,N} = \langle \hat{\mu}_{p,N} \rangle_p = \mu_p$. *Remark:* The independent and identically distributed random variables are also symmetrically distributed. Then, from *Theorem 1* we have $\langle \hat{\mu}_{p,N} \rangle_p = \hat{E}_p[\hat{\mu}_{p,N}(\{y_j\}_{j=1}^N; p)] = \langle Y_i \rangle_p = \hat{E}_p(Y_i) = \mu_p$. The $\mathcal{L}^p$ -expectation value $\langle Y_i \rangle_p = \mu_p$ is calculated given the marginal distribution density $\mathfrak{L}_y(y_i), \forall i = 1, \dots, N$. However, the expression of this distribution is in the generic case unknown, and thus, we estimate μ_p by means of $\hat{\mu}_{p,N}$ for $N \gg 1$.

3.3. Examples

In the following three examples from Statistical Mechanics, we examine the systems of (1) gas in thermal equilibrium, (2) space plasmas out of thermal equilibrium, and (3) multi-dimensional quantum harmonic oscillator at thermal equilibrium. The non-Euclidean-normed internal energy U_p is derived by utilizing the classical Euclidean probability distribution of Canonical Ensemble.

3.3.1. Gas at Thermal Equilibrium

For the continuous energy spectrum $\varepsilon \in [0, \infty)$ with distribution $p(\varepsilon)$ and degeneracy $g(\varepsilon)$, the $\mathcal{L}^p$ internal energy $\langle \varepsilon \rangle_p$ is given by

$$\int_0^\infty p(\varepsilon) g(\varepsilon) |\varepsilon - \langle \varepsilon \rangle_p|^{p-1} \text{sign}(\varepsilon - \langle \varepsilon \rangle_p) d\varepsilon = 0 \quad (24)$$

At classical thermal equilibrium, the energy distribution is given by the Boltzmann–Gibbs distribution $p(\varepsilon) \propto e^{-\frac{\varepsilon}{k_B T}}$. The degeneracy is $g(\varepsilon) \propto \varepsilon^{\frac{f}{2}-1}$, where f denotes the degrees of freedom. Hence, the internal energy, $\langle \varepsilon \rangle_p / (k_B T)$, is implicitly expressed in terms of p , f , as follows

$$\int_0^\infty e^{-x} x^{\frac{f}{2}-1} \left| x - \frac{\langle \varepsilon \rangle_p}{k_B T} \right|^{p-1} \text{sign} \left(x - \frac{\langle \varepsilon \rangle_p}{k_B T} \right) dx = 0 \quad (25)$$

where we set $x \equiv \varepsilon / (k_B T)$. In the Euclidean case, the internal energy is $\langle \varepsilon \rangle_2 = (f/2)k_B T$. In the non-Euclidean case, this is written as $\langle \varepsilon \rangle_p = (f_p/2)k_B T$, where f_p represents the reflected degrees of freedom; for sufficiently high number of degrees we find $f_p \simeq f + 0.66 \cdot (p - 2)$.

3.3.2. Plasma Out of Thermal Equilibrium

Classical systems are said to be in thermal equilibrium – the concept that any flow of heat (thermal conduction, thermal radiation) is in balance. However, thermal equilibrium is not the only possible state that is stationary (*i.e.*, the phase space distribution does not explicitly depend on time). For example, space plasmas are systems residing in stationary states but out of thermal equilibrium. For these systems, the energy distribution is well-described by the empirical kappa distribution (see [26] and references therein). Moreover, the kappa distribution was shown to be connected [26] with the solid background of non-extensive Statistical Mechanics [27], and represents the probability distribution that maximizes entropy in the Canonical Ensemble. The kappa distribution is the generalization of the classical Boltzmann–Gibbs exponential distribution that describes systems only at thermal equilibrium. The temperature and the kappa index that govern these distributions are the two independent controlling parameters of non-equilibrium systems. The invariant form of the kappa distribution, in which the temperature T , the kappa index κ_0 and the total degrees of freedom f are all independent variables [28], is given by

$$P(\varepsilon; T; \kappa_0; f) \propto \left(1 + \frac{1}{\kappa_0} \cdot \frac{\varepsilon}{k_B T} \right)^{-\kappa_0 - 1 - \frac{f}{2}} \quad (26)$$

where the kappa index ($\kappa_0 > 0$) determines a measure of how far the system is from the thermal equilibrium [29]. The kappa distribution recovers the Boltzmannian exponential distribution for

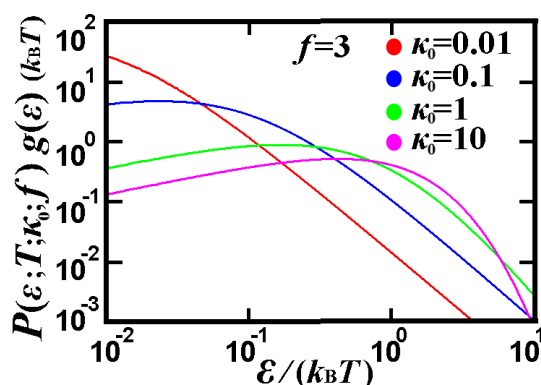
$\kappa_0 \rightarrow \infty$, which is the value of the kappa index characterizing thermal equilibrium. The smallest possible value of the kappa index is $\kappa_0 \rightarrow 0$, and determines the furthest stationary state from thermal equilibrium [30].

In Figure 1 the kappa distribution $P(\varepsilon; T; \kappa_0; f)g(\varepsilon) \times (k_B T)$ is depicted in terms of $\varepsilon/(k_B T)$ for $\kappa_0 = 0.01, 0.1, 1, 10$. The internal energy $\langle \varepsilon \rangle_p$ is given by

$$\int_0^\infty \left(1 + \frac{x}{\kappa_0}\right)^{-\kappa_0-1-\frac{f}{2}} x^{\frac{f}{2}-1} \left|x - \frac{\langle \varepsilon \rangle_p}{k_B T}\right|^{p-1} \text{sign}\left(x - \frac{\langle \varepsilon \rangle_p}{k_B T}\right) = 0 \quad (27)$$

(Again, we set $x \equiv \varepsilon/(k_B T)$ and the degeneracy is $g(\varepsilon) \propto \varepsilon^{\frac{f}{2}-1}$.) In Figure 2a we depict the internal energy, $\langle \varepsilon \rangle_p/(1/2 k_B T)$ with respect to the degrees of freedom f , for $p = 1.5, 2, 2.5$ and for kappa indices $\kappa_0 = 1.5$ and $\kappa_0 = 100$ (this practically equals the Boltzmannian exponential distribution at thermal equilibrium). Figure 2b plots the internal energy over the degrees of freedom, $\langle \varepsilon \rangle_p/(f/2 k_B T)$, for the same p and κ_0 . Figure 2c shows the dependence on the kappa index (for $f = 3$). Interestingly, the internal energy is kappa-dependent for any of the non-Euclidean norms, *i.e.*, $\langle \varepsilon \rangle_p = f_p/2 k_B T$, with $f_p = f_p(\kappa_0; p)$. (Hence, T does not well-define the kinetic temperature for $p \neq 2$ [26,30].) Note that the integral converges only for $\kappa_0 > p - 2$, which generalizes the inequality $\kappa_0 > 0$ of the Euclidean case.

Figure 1. The kappa distribution of energy, $P(\varepsilon; T; \kappa_0; f)g(\varepsilon) \times (k_B T)$, depicted in a log-log scale for $\kappa_0 = 0.01$ (red), 0.1 (blue), 1 (green), 10 (magenta), and for $f = 3$.



3.3.3. D-Dimensional Quantum Harmonic Oscillator in Thermal Equilibrium

For the energy states $\{\varepsilon_n\}_{n=1}^W$ with degeneracy $\{g_n\}_{n=1}^W$ that are associated with the discrete energy distribution $\{\rho_n\}_{n=1}^W$, the $\mathcal{L}^p$ internal energy $\langle \varepsilon \rangle_p$ is

$$\sum_{n=1}^W \rho_n g_n |\varepsilon_n - \langle \varepsilon \rangle_p|^{p-1} \text{sign}(\varepsilon_n - \langle \varepsilon \rangle_p) = 0 \quad (28)$$

with the Boltzmann's energy distribution $\rho_n \propto e^{-\frac{\varepsilon_n}{k_B T}}$ at thermal equilibrium. The D -dimensional quantum harmonic oscillator has discrete energy spectrum $\varepsilon_n = \hbar\omega(n + \frac{1}{2})$ with degeneracy $g_n = \binom{n+D-1}{n} \propto (n+D-1)!/n!$, $\forall n = 0, 1, 2, \dots$. Then, the internal energy $\langle \varepsilon \rangle_p$ is implicitly given by

$$\sum_{n=0}^{\infty} e^{-\frac{\hbar\omega}{k_B T} n} \cdot \left|n + \frac{1}{2} - \frac{\langle \varepsilon \rangle_p}{\hbar\omega}\right|^{p-1} \text{sign}\left(n + \frac{1}{2} - \frac{\langle \varepsilon \rangle_p}{\hbar\omega}\right) \cdot \frac{(n+D-1)!}{n!} = 0 \quad (29)$$

In Figure 3a, we depict $\langle \varepsilon \rangle_p / (\hbar \omega)$ with respect to $k_B T / (\hbar \omega)$, for $p = 1.8, 2, 2.5$ and $D = 2$. The relevant heat capacity, $c_{V,p} = (\partial \langle \varepsilon \rangle_p / \partial T)_V$, is shown in Figure 3b. For high temperatures, the Dulong–Petit limit is generalized to $\langle \varepsilon \rangle_p \simeq \frac{1}{2} f_p \cdot k_B T + \alpha_p \cdot \hbar \omega$, and heat capacity $c_{V,p} \simeq \frac{1}{2} f_p \cdot k_B$, where the reflected degrees of freedom are $f_p \equiv 2D + 0.6 \cdot (p - 2)$; the involved constant is $\alpha_p \simeq 0.015D(p - 2)$. Each dimension of the Euclidean quantum oscillator has two degrees of freedom so that $f_2 = 2D$; the non-Euclidean quantum oscillator deviates proportionally to the factor $f_p - f_2 \propto p - 2$. In Figure 3c,d we depict $\langle \varepsilon \rangle_p$ and $c_{V,p}$ as a function of the dimensionality D , for $p = 1.8, 2, 2.5$ and $k_B T / (\hbar \omega) = 1$. While for the Euclidean norm the heat capacity gives $c_{V,p} / k_B = D$, for the sub- and super-Euclidean norms we have $c_{V,p} / k_B < D$ and $> D$, respectively. Finally, we remark that the heat capacity appears to have periodic cusps that become smaller in higher temperatures.

Figure 2. The dependence of the internal energy on the degrees of freedom f and the kappa index κ_0 . **(a)** Plot of $\langle \varepsilon \rangle_p / (1/2 k_B T)$ vs. the degrees of freedom f , for $p = 1.5$ (red), 2 (blue), 2.5 (green), and $\kappa_0 = 1.5$ (solid), and 100 (dash). **(b)** Plot of $\langle \varepsilon \rangle_p / (f/2 k_B T)$, vs. f , for the same p and κ_0 . **(c)** The same plot as (b), but vs. κ_0 and for $f = 3$. We observe larger deviation from the Euclidean norm for smaller kappa indices. Note that the graph is restricted to $\kappa_0 > p - 2$ (see text).

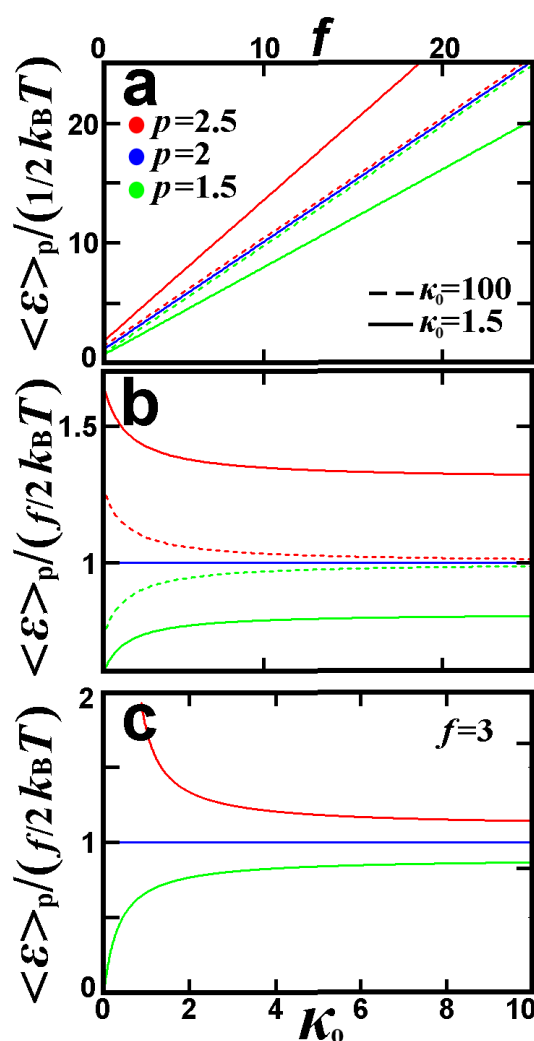
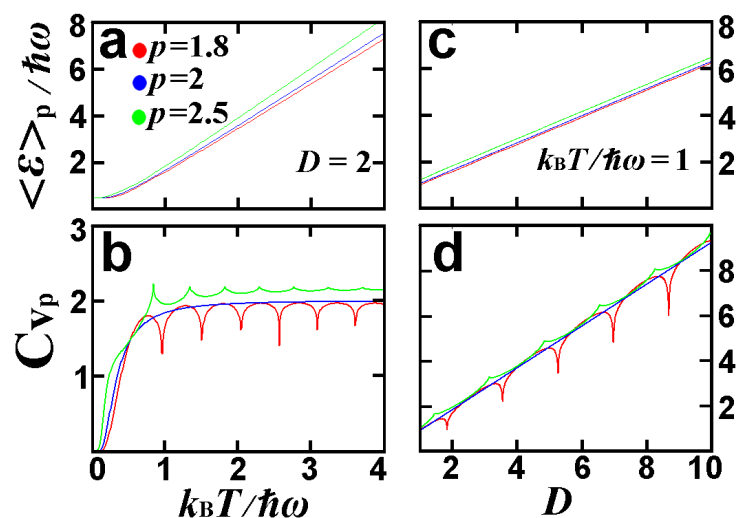


Figure 3. The internal energy and heat capacity of the D -dimensional quantum harmonic oscillator for non-Euclidean norms ($\forall p$) at thermal equilibrium ($\kappa_0 \rightarrow \infty$). (a), (b) The internal energy $\langle \varepsilon \rangle_p / (\hbar\omega)$ and heat capacity $c_{V,p}$, are respectively plotted vs. the temperature $k_B T / (\hbar\omega)$, for $p = 1.8$ (red), 2 (blue), 2.5 (green), and $D = 2$. (c), (d) Similar to (a) and (b), but plotted as functions of the dimensionality D , for $k_B T / (\hbar\omega) = 1$.



4. The Non-Euclidean $\mathcal{L}^p$ -Variance of the $\mathcal{L}^p$ -Expectation Value

4.1. Preliminaries: Formulations

We study the $\mathcal{L}^p$ -variance, defined either in the discrete description $\{y_k\}_{k=1}^W$, $y_k \in D_y \subseteq \mathbb{R}, \forall k = 1, \dots, W$, with probability distribution $\{\rho_k\}_{k=1}^W$,

$$\sigma_p^2 = \frac{\sum_{k=1}^W \rho_k |y_k - \langle y \rangle_p|^p}{(p-1) \sum_{k=1}^W \rho_k |y_k - \langle y \rangle_p|^{p-2}} \quad (30)$$

with $\langle y \rangle_p$ implicitly given by Equation (8), or in the continuous description $y \in D_y \subseteq \mathbb{R}$, with probability distribution $\rho(y)$,

$$\sigma_p^2 = \frac{\int_{y \in D_y} \rho(y) |y - \langle y \rangle_p|^p dy}{(p-1) \int_{y \in D_y} \rho(y) |y - \langle y \rangle_p|^{p-2} dy} \quad (31)$$

with $\langle y \rangle_p$ implicitly given by

$$\int_{y \in D_y} \rho(y) |y - \langle y \rangle_p|^{p-1} \text{sign}(y - \langle y \rangle_p) dy = 0 \quad (32)$$

4.2. Examples

4.2.1. Example 1: Gaussian distribution

Consider the Gaussian probability distribution density

$$f_G(y; \mu, \sigma) = \frac{1}{\sqrt{2\pi}} \frac{1}{\sigma} e^{-\frac{1}{2\sigma^2}(y-\mu)^2} \quad (33)$$

Then, the $\mathcal{L}^p$ variance is given by

$$\begin{aligned}\sigma_p^2 &= \frac{\int_{-\infty}^{+\infty} |y - \mu|^p e^{-\frac{1}{2\sigma^2}(y-\mu)^2} dy}{(p-1) \int_{-\infty}^{+\infty} |y - \mu|^{p-2} e^{-\frac{1}{2\sigma^2}(y-\mu)^2} dy} = \frac{\sigma^2}{p-1} \frac{\int_{-\infty}^{+\infty} |u|^p e^{-\frac{1}{2}u^2} du}{\int_{-\infty}^{+\infty} |u|^{p-2} e^{-\frac{1}{2}u^2} du} = \\ &= \frac{\sigma^2}{p-1} \frac{\int_0^{+\infty} u^p e^{-\frac{1}{2}u^2} du}{\int_0^{+\infty} u^{p-2} e^{-\frac{1}{2}u^2} du} = \sigma^2 \frac{2}{p-1} \frac{\int_0^{+\infty} z^{\frac{p+1}{2}-1} e^{-z} dz}{\int_0^{+\infty} z^{\frac{p-1}{2}-1} e^{-z} dz} = \sigma^2 \frac{2}{p-1} \frac{\Gamma(\frac{p+1}{2})}{\Gamma(\frac{p-1}{2})} = \sigma^2\end{aligned}\quad (34)$$

where we have set $u \equiv \frac{y-\mu}{\sigma}$, and $z \equiv \frac{u^2}{2}$. Thus, the $\mathcal{L}^p$ variance yields the well-known Euclidean variance σ^2 of the Gaussian distribution.

4.2.2. Example 2: Generalized Gaussian distribution

Consider the Generalized Gaussian probability distribution density,

$$f_{GG}(y; \mu, \sigma; Q, p) = C_{Q,p} \frac{1}{\sigma} e^{-\eta_{Q,p} |\frac{y-\mu}{\sigma}|^Q} \quad (35)$$

where $Q \geq 0$ is the shape parameter, $C_{Q,p}$ is the normalization constant, and $\eta_{Q,p}$ is calculated by setting $\sigma_p^2 \equiv \sigma^2$,

$$\begin{aligned}\sigma_p^2 &= \frac{\int_{-\infty}^{+\infty} |y - \mu|^p e^{-\eta_{Q,p} |\frac{y-\mu}{\sigma}|^Q} dy}{(p-1) \int_{-\infty}^{+\infty} |y - \mu|^{p-2} e^{-\eta_{Q,p} |\frac{y-\mu}{\sigma}|^Q} dy} = \frac{\sigma^2}{p-1} \frac{\int_{-\infty}^{+\infty} |u|^p e^{-\eta_{Q,p} |u|^Q} du}{\int_{-\infty}^{+\infty} |u|^{p-2} e^{-\eta_{Q,p} |u|^Q} du} = \\ &= \frac{\sigma^2}{p-1} \frac{\int_0^{+\infty} u^p e^{-\eta_{Q,p} u^Q} du}{\int_0^{+\infty} u^{p-2} e^{-\eta_{Q,p} u^Q} du} = \sigma^2 \frac{\eta_{Q,p}^{-\frac{2}{Q}}}{p-1} \frac{\int_0^{+\infty} z^{\frac{p+1}{Q}-1} e^{-z} dz}{\int_0^{+\infty} z^{\frac{p-1}{Q}-1} e^{-z} dz} = \sigma^2 \frac{\eta_{Q,p}^{-\frac{2}{Q}}}{p-1} \frac{\Gamma(\frac{p+1}{Q})}{\Gamma(\frac{p-1}{Q})} = \sigma^2\end{aligned}\quad (36)$$

where we have set $u \equiv \frac{y-\mu}{\sigma}$, $z \equiv \eta_{Q,p} u^Q$, and

$$\eta_{Q,p} \equiv \left[\frac{1}{p-1} \frac{\Gamma(\frac{p+1}{Q})}{\Gamma(\frac{p-1}{Q})} \right]^{\frac{Q}{2}}$$

Of particular interest is the case where the shape parameter equals the norm, $Q = p$, namely,

$$f_{GG}(y; \mu, \sigma; p) = C_p \frac{1}{\sigma} e^{-\eta_p |\frac{y-\mu}{\sigma}|^p} \quad (37)$$

with

$$\eta_p = \left[\frac{1}{p-1} \frac{\Gamma(1 + \frac{1}{p})}{\Gamma(1 - \frac{1}{p})} \right]^{\frac{p}{2}} = \left[\frac{\sin(\frac{\pi}{p}) \Gamma(\frac{1}{p})^2}{\pi p(p-1)} \right]^{\frac{p}{2}}$$

(Note that $\langle y \rangle_p = \mu$ for both the examples. See Section 5.1.)

4.3. Justification of the $\mathcal{L}^p$ -Variance Expression

Let the set of $\{Y_i\}_{i=1}^N$ independent and identically distributed random variables with sampling values $\{y_i\}_{i=1}^N$, $y_i \in D_y \subseteq \mathbb{R}$, $\forall i = 1, \dots, N$.

First, we consider that each variable is distributed as $Y_i \sim \mathcal{N}(\mu, \sigma)$, namely, by the Gaussian probability distribution density

$$f_G(y_i; \mu, \sigma) = \frac{1}{\sqrt{2\pi}} \frac{1}{\sigma} e^{-\frac{1}{2\sigma^2}(y_i - \mu)^2} \quad (38)$$

$\forall i = 1, \dots, N$. Then, the likelihood function is constructed

$$\begin{aligned} \mathfrak{L}_y(\{y_i\}_{i=1}^N; \mu, \sigma) &= \prod_{i=1}^N f_G(y_i; \mu, \sigma) = (\sqrt{2\pi}\sigma)^{-N} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^N (y_i - \mu)^2} \\ \Rightarrow \ln[\mathfrak{L}_y(\{y_i\}_{i=1}^N; \mu, \sigma)] &= -\frac{N}{2} \ln(2\pi) - N \ln \sigma - \frac{1}{2\sigma^2} \sum_{i=1}^N (y_i - \mu)^2 \end{aligned} \quad (39)$$

Then, $\frac{\partial}{\partial \mu} \ln[\mathfrak{L}_y(\{y_i\}_{i=1}^N; \mu; \sigma)] = 0$ leads to

$$\sum_{i=1}^N (y_i - \mu) = 0 \Rightarrow \mu = \frac{1}{N} \sum_{i=1}^N y_i \quad (40)$$

while $\frac{\partial}{\partial \sigma} \ln[\mathfrak{L}_y(\{y_i\}_{i=1}^N; \mu; \sigma)] = 0$ leads to

$$-N \frac{1}{\sigma} + \frac{1}{\sigma^3} \sum_{i=1}^N (y_i - \mu)^2 = 0 \Rightarrow \sigma^2 = \frac{1}{N} \sum_{i=1}^N (y_i - \mu)^2 \quad (41)$$

Thus, for the normally distributed $\{y_i\}_{i=1}^N$, the best estimators for the true value of the parameters μ and σ^2 are the Euclidean estimators of the mean and variance, that is $\hat{\mu}_{2,N}(\{y_i\}_{i=1}^N) = \frac{1}{N} \sum_{i=1}^N y_i$ and $\hat{\sigma}_{2,N}^2(\{y_i\}_{i=1}^N) = \frac{1}{N} \sum_{i=1}^N (y_i - \mu)^2$, respectively. What is important here is not the variance σ^2 of the theoretical distribution density $f_G(y_i; \mu; \sigma)$, but the variance $\widehat{S}_{2,N}^2$ of the Euclidean mean estimator $\hat{\mu}_{2,N}(\{y_i\}_{i=1}^N)$ [31], given by

$$\frac{1}{\widehat{S}_{2,N}^2} = -\frac{\partial^2}{\partial \mu^2} \ln[\mathfrak{L}_y(\{y_i\}_{i=1}^N; \mu; \sigma)] = N \frac{1}{\sigma^2} = N \frac{1}{\widehat{\sigma}_{2,N}^2} \quad (42)$$

so that the value of μ , estimated by $\hat{\mu}_{2,N}$, has a variance $\widehat{S}_{2,N}^2$, given by

$$\widehat{S}_{2,N}^2 = \frac{1}{N} \widehat{\sigma}_{2,N}^2 \quad (43)$$

Now consider that each variable is distributed as $Y_i \sim \mathcal{GG}(\mu, \sigma; Q, p)$, namely, by the General Gaussian probability distribution density of Equation (35),

$$f_{GG}(y_i; \mu, \sigma; p) = C_p \frac{1}{\sigma} e^{-\eta_p \left| \frac{y_i - \mu}{\sigma} \right|^p} \quad (44)$$

$\forall i = 1, \dots, N$. Then, the likelihood function is constructed

$$\begin{aligned} \mathfrak{L}_y(\{y_i\}_{i=1}^N; \mu, \sigma; p) &= \prod_{i=1}^N f_{GG}(y_i; \mu, \sigma; p) = C_p^N \sigma^{-N} e^{-\frac{\eta_p}{\sigma^p} \sum_{i=1}^N |y_i - \mu|^p} \\ \Rightarrow \ln[\mathfrak{L}_y(\{y_i\}_{i=1}^N; \mu, \sigma; p)] &= N \ln C_p - N \ln \sigma - \frac{\eta_p}{\sigma^p} \sum_{i=1}^N |y_i - \mu|^p \end{aligned} \quad (45)$$

Then, $\frac{\partial}{\partial \mu} \ln[\mathcal{L}_y(\{y_i\}_{i=1}^N; \mu, \sigma; p)] = 0$ leads to

$$\sum_{i=1}^N |y_i - \mu|^{p-1} \text{sign}(y_i - \mu) = 0 \quad (46)$$

Thus, for $\{y_i\}_{i=1}^N$ distributed by the General Gaussian, the best estimator for the true value of the parameter μ is the non-Euclidean $\mathcal{L}^p$ -mean estimator $\hat{\mu}_{p,N}(\{y_i\}_{i=1}^N; p)$, defined in Equation (19).

In addition, $\frac{\partial}{\partial \sigma} \ln[\mathcal{L}_y(\{y_i\}_{i=1}^N; \mu, \sigma; p)] = 0$ leads to

$$-N \frac{1}{\sigma} + \frac{\eta_p p}{\sigma^{p+1}} \sum_{i=1}^N |y_i - \mu|^p = 0 \Rightarrow \frac{\sigma^p}{\eta_p p} = \frac{1}{N} \sum_{i=1}^N |y_i - \mu|^p \quad (47)$$

and thus, the variance $\widehat{S}_{p,N}^2$ of the estimator $\hat{\mu}_{p,N}(\{y_i\}_{i=1}^N; p)$ is given by

$$\frac{1}{\widehat{S}_{p,N}^2} = -\frac{\partial^2}{\partial \mu^2} \ln[\mathcal{L}_y(\{y_i\}_{i=1}^N; \mu, \sigma; p)] = \frac{\eta_p p}{\sigma^p} (p-1) \sum_{i=1}^N |y_i - \mu|^{p-2} = N \frac{(p-1) \sum_{i=1}^N |y_i - \mu|^{p-2}}{\sum_{i=1}^N |y_i - \mu|^p} \equiv N \frac{1}{\widehat{\sigma}_{p,N}^2}$$

or

$$\widehat{S}_{p,N}^2 = \frac{1}{N} \widehat{\sigma}_{p,N}^2 \quad (48)$$

which is similar to Equation (43) after replacing the Euclidean variance, $\widehat{\sigma}_{2,N}^2$, with the non-Euclidean one, $\widehat{\sigma}_{p,N}^2$. Finally, the $\mathcal{L}^p$ variance (e.g., as given in Equation (30)) can be written in the form of Equation (17), i.e.,

$$\begin{aligned} \sigma_p^2 &= \frac{\sum_{k=1}^W \rho_k |y_k - \langle y \rangle_p|^p}{(p-1) \sum_{k=1}^W \rho_k |y_k - \langle y \rangle_p|^{p-2}} = \sum_{k=1}^W \rho_k (y_k - \langle y \rangle_p) \frac{|y_k - \langle y \rangle_p|^{p-1} \text{sign}(y_k - \langle y \rangle_p)}{(p-1) \phi_p} \\ &= \sum_{k=1}^W \rho_k (y_k - \langle y \rangle_p) \hat{\mathcal{L}}_p(y_k - \langle y \rangle_p) = \left\langle (y - \langle y \rangle_p) \hat{\mathcal{L}}_p(y - \langle y \rangle_p) \right\rangle_2. \end{aligned}$$

Note that the $\mathcal{L}^p$ mean $\langle y \rangle_p$ (Equations (1) and (8)) is derived from the minimization of $\sigma_p^2(\alpha) = \frac{1}{(p-1)\phi_p} TD_p(\{y_k\}_{k=1}^W; \alpha; p)^p$, which is proportional to the total deviations TD_p^p (Equation (2)).

Moreover, total deviations $TD_p(\{y_k\}_{k=1}^W; \alpha; p)$ can be expanded at the minimum $\alpha = \alpha^*$,

$$TD_p(\{y_k\}_{k=1}^W; \alpha; p)^p = A_0(p) + A_1(p)(\alpha - \alpha^*) + A_2(p)(\alpha - \alpha^*)^2 + O[(\alpha - \alpha^*)^3] \quad (49)$$

where $A_0(p) = TD_{p,Min}(\{y_k\}_{k=1}^W; p)^p = \sum_{k=1}^W \rho_k |y_k - \alpha^*(p)|^p$ is the Total p -Deviations minimum value, $A_1(p) = -p \sum_{k=1}^W \rho_k |y_k - \alpha^*(p)|^{p-1} \text{sign}(y_k - \alpha^*(p)) = 0$ is the Total p -Deviation first derivative (in terms of the parameter α), which is equal to zero, while the curvature factor $A_2(p)$ is given by

$$A_2(p) = \frac{p(p-1)}{2} \sum_{k=1}^W \rho_k |y_k - \alpha^*(p)|^{p-2} \quad (50)$$

As it is shown in [7], the variance σ_p^2 that is proportional to the square error of the optimal value $\alpha^*(p)$ is inversely proportional to $A_2(p)$,

$$\sigma_p^2 \propto \frac{1}{(p-1) \sum_{k=1}^W \rho_k |y_k - \alpha^*(p)|^{p-2}} \quad (51)$$

On the other hand, it is expected that the variance σ_p^2 will be proportional also to $A_0(p)$, namely,

$$\sigma_p^2 \propto \sum_{k=1}^W \rho_k |y_k - \alpha^*(p)|^p \quad (52)$$

which completes the proportionality of Equation (51), leading to Equation (30).

There are two more ways to show the specific expression of $\mathcal{L}^p$ -variance: The proportionality factor C connects, either the variance with the Total Deviations, *i.e.*, $\sigma_p^2 = C \cdot TD_p^p$, or the energy states $\{\varepsilon_k\}_{k=1}^W$ with the “ $\mathcal{L}^p$ energy states” $\{\hat{\mathcal{L}}_p(\varepsilon_k)\}_{k=1}^W$, *i.e.*, $\hat{\mathcal{L}}_p(\varepsilon_k - U_p) = C|\varepsilon_k - U_p|^{p-1} \text{sign}(\varepsilon_k - U_p)$. The property (ii) of the non-Euclidean norm operator $\hat{\mathcal{L}}_p$ holds if and only if $C = \frac{1}{(p-1)\phi_p}$. This can be shown in the following two ways: (1) Due to the property (ii) of the non-Euclidean norm operator, the respective Canonical probability distribution can be automatically derived. Indeed, the internal energy constraint $\langle \varepsilon \rangle_p = \sum_{k=1}^W \rho_k \hat{\mathcal{L}}_p(\varepsilon_k)$ leads to the derivative $\frac{\partial}{\partial \rho_j} \sum_{k=1}^W \rho_k \hat{\mathcal{L}}_p(\varepsilon_k) = \hat{\mathcal{L}}_p(\varepsilon_j)$, so that by maximizing Boltzmann-Gibbs entropy we end up with $\rho_k \sim e^{-\beta \hat{\mathcal{L}}_p(\varepsilon_k)}$. (2) The connection with thermodynamics can be achieved if and only if $C = \frac{1}{(p-1)\phi_p}$ (Equation (15)).

5. Further Analytical and Numerical Examples

5.1. Analytical Example: The Spectrum of the $\mathcal{L}^p$ Means and Its Degeneration

For a continuous description of data $y(t) \in D_y \subseteq \mathbb{R}$, where $t \in D_t \subseteq \mathbb{R}$ is a continuous index, the $\mathcal{L}^p$ -expectation value $\langle y \rangle_p$ is given by

$$\int_{t \in D_t} |y(t) - \langle y \rangle_p|^{p-1} \text{sign}[y(t) - \langle y \rangle_p] dt = 0 \quad (53)$$

Given the probability distribution of y -values, $\rho(y)$, the $\mathcal{L}^p$ -expectation value is given by Equation (32). Consider now the equidistribution of data in the interval $[0, 1]$. From Equation (32) we have

$$\begin{aligned} \int_0^1 |y - \langle y \rangle_p|^{p-1} \text{sign}(y - \langle y \rangle_p) dy &= 0 \Leftrightarrow \int_0^{\langle y \rangle_p} (\langle y \rangle_p - y)^{p-1} dy = \int_{\langle y \rangle_p}^1 (y - \langle y \rangle_p)^{p-1} dy \\ &\Leftrightarrow \langle y \rangle_p^p = (1 - \langle y \rangle_p)^p \Leftrightarrow \langle y \rangle_p = \frac{1}{2} \quad \forall p \geq 1. \end{aligned}$$

The fact that $\langle y \rangle_p$ is independent of p , is a general result of symmetric probability distributions. Indeed, consider a distribution $\rho(y)$, $y \in [-c, c]$, symmetric at $y = 0$, *i.e.*, $\rho(-y) = \rho(y) \forall y \in [-c, c]$. Then,

$$\int_{-c}^c \rho(y) |y|^{p-1} \text{sign}(y) dy = 0,$$

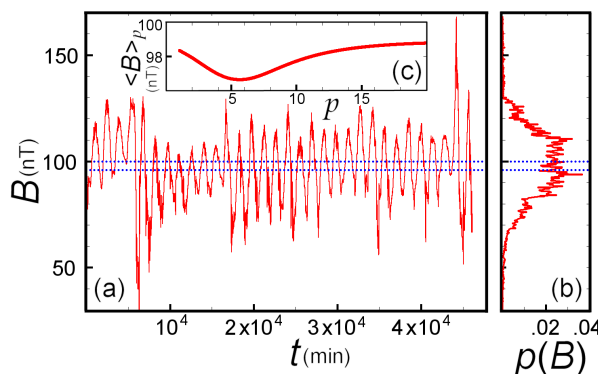
(the *sign* is odd function, while $\rho(y)|y|^{p-1}$ is even). Hence, given the uniqueness of $\langle y \rangle_p$ for a given p , we conclude in $\langle y \rangle_p = 0$, $\forall p \geq 1$.

Therefore, in the case where the distribution $\rho(y)$ is symmetric, the whole set of $\langle y \rangle_p$ -values degenerate to one single value, which can be found thus, by the usual Euclidean norm, namely $\langle y \rangle_p = \langle y \rangle_2$. (The opposite statement is also true.) However, when $\rho(y)$ is asymmetric, a spectrum-like range of different $\langle y \rangle_p$ -values is generated [7]. For example, the distribution $\rho(y) \simeq 1 + \delta(1 - 3y^2)$ in the interval $y \in [0, 1]$ is symmetric for $\delta = 0$, but becomes asymmetric for $0 < \delta \ll 1$. Then, we find $\langle y \rangle_p \simeq \frac{1}{2} - \frac{3}{4} \frac{1}{p+1} \delta + O(\delta^2)$.

5.2. Numerical Example: Earth's Magnetic Field

We consider the time series of the Earth's magnetic field magnitude (in nT). In particular, we focus on a stationary segment recorded by the GOES-12 satellite between the month 1/1/2008 and 1/2/2008, that is a sampling of one measurement per minute, constituting a segment of $N = 46,080$ data points, depicted in Figure 4a. This segment is characterized by a roughly symmetric distribution $p(B)$ (in nT^{-1}), depicted in Figure 4b, resulting to a narrow spectrum of $\langle B \rangle_p$ -values, depicted in Figure 4c. Notice that the extreme values of $\langle B \rangle_p$ do not coincide with the respective for $p = 1$ and $p \rightarrow \infty$. Indeed, a minimum of the $\mathcal{L}^p$ -expectation values, that is $\langle B \rangle_{p,\min} \approx 96.614$ nT, can be found for the non-Euclidean norm $p \approx 5.63$, while a maximum value of about $\langle B \rangle_{p,\max} \approx 98.819$ nT, is located at $p \approx 25.83$. As $p \rightarrow \infty$, $\langle B \rangle_p$ tends to $(B_{\min} + B_{\max})/2 \approx 98.535$ nT. Hence, it is evident that $\mathcal{L}^p$ means μ_p are not indispensably monotonic functions of p .

Figure 4. The magnitude of the Earth's total magnetic field. (a) The time series recorded between 1/1/2008 and 1/2/2008. (b) The relevant distribution $p(B)$ is roughly symmetric. As a result, the numerically calculated $\mathcal{L}^p$ -expectation values, $\langle B \rangle_p$, configure a narrow spectrum within the interval between the two horizontal dotted lines, where the dependence of $\langle B \rangle_p$ -values on the p -norm is shown within the magnified inset (c).

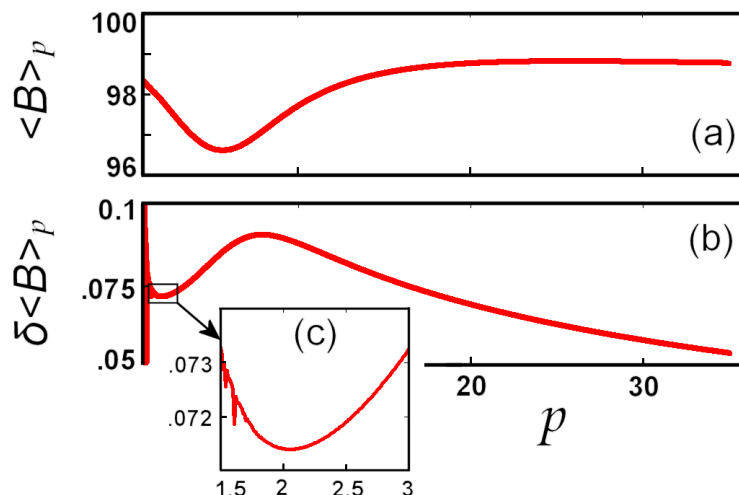


The expectation value $\langle B \rangle_p$ is given by the estimator $\hat{\mu}_{p,N}(\{B_i\}_{i=1}^N; p)$. On the other hand, the error $\delta \langle B \rangle_p$ is given by the square root of the variance $\widehat{S}_{p,N}^2$ of the estimator $\hat{\mu}_{p,N}(\{B_i\}_{i=1}^N; p)$, that is

$$\delta \langle B \rangle_p = \widehat{S}_{p,N} = \frac{1}{\sqrt{N}} \widehat{\sigma}_{p,N} = \sqrt{\frac{1}{N} \frac{1}{p-1} \frac{\sum_{i=1}^N |B_i - \langle B \rangle_p|^p}{\sum_{i=1}^N |B_i - \langle B \rangle_p|^{p-2}}} \quad (54)$$

In Figure 5a,b, the $\mathcal{L}^p$ -expectation value of the Earth's magnetic field magnitude (shown in Figure 4a), $\langle B \rangle_p$, together with its error $\delta \langle B \rangle_p$, are respectively depicted as functions of the p -norm. A local minimum of the error $\delta \langle B \rangle_p$ can be detected for $p \approx 2.05$, for which $\langle B \rangle_p \approx 97.88$ nT and $\delta \langle B \rangle_{p,\min} \approx 0.071$ nT, shown in the magnified inset of Figure 5c.

Figure 5. The $\mathcal{L}^p$ -expectation value of the magnitude of the Earth's total magnetic field (shown in Figure 4a), $\langle B \rangle_p$, together with its error $\delta \langle B \rangle_p$, are depicted as functions of the p -norm (panels (a) and (b), respectively). A local minimum of the error is found close to the Euclidean norm, *i.e.*, for $p \approx 2.05$, as it is shown within the magnified inset (c).



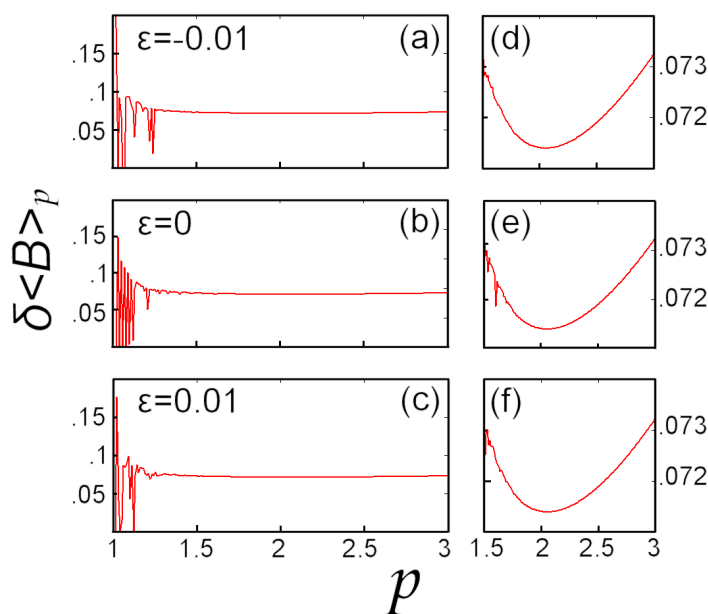
For $p \gtrsim 2.05$ the error increases as p increases, until it reaches a local maximum at $p \approx 7.88$, for which $\langle B \rangle_p \approx 97.07$ nT and $\delta \langle B \rangle_{p,max} \approx 0.091$ nT. Then, for $p \gtrsim 7.88$ the error decreases monotonically as p increases. We can readily derive that $\delta \langle B \rangle_p \approx \frac{1}{2}(B_{max} - B_{min}) \frac{1}{\sqrt{N}} \frac{1}{\sqrt{p-1}} \rightarrow 0$, as $p \rightarrow \infty$.

For $p \lesssim 2.05$ the error increases as p decreases, but numerous fluctuations appear that become more dense as $p \rightarrow 1$. This “instability cloud” is due to the reading errors of the data values, with their effect being magnified as $p - 1$ tends to zero. This effect can be demonstrated in Figure 6, where the error $\delta \langle B \rangle_p$ is depicted when an additive noise is inserted into the $\{B_i\}_{i=1}^N$ values. In particular, we consider the perturbed values $\{\tilde{B}_i\}_{i=1}^N$, where $\tilde{B}_i \equiv B_i + \epsilon r_i, \forall i = 1, \dots, N$, with $\{r_i\}_{i=1}^N$, being equidistributed in $[0, 1]$ and $|\epsilon|$ is the amplitude of the perturbations. We set $|\epsilon| = 0.01$ nT, which is equal to the resolution (reading error) of the values $\{B_i\}_{i=1}^N$. In Figure 6a we set $\epsilon = -0.01$ nT, while in Figure 6c we set $\epsilon = +0.01$ nT. In Figure 6b we depict the unperturbed error for convenience. We observe that the “instability cloud”, occurred for $p \rightarrow 1$, is different for the three cases $\epsilon = -0.01, 0$, and $+0.01$ (nT). However, the minimum at $p \approx 2.05$ remains unaffected.

The existence of a local minimum of the error, such as the minimum at $p \approx 2.05$, is of great importance. It suggests that for this specific norm, the expectation value comes with the minimum error. Therefore, after the total deviations minimization that leads to the normal Equation (1) from which the optimal parameter $\alpha^*(p)$ is derived, the optimization is completed by determining the specific norm p^* for which the variance $\hat{\sigma}_p^2(p)$ has a local minimum (if that exists). The importance of the local minimum of $\sigma_p^2(p)$ is that for any deviation of the norm at $p = p^*$, either $p < p^*$, or $p > p^*$, the error increases, introducing thus, a type of norm-stability. Hence, the $\mathcal{L}^p$ -norm that corresponds to the minimized error, $p \approx 2.05$, is distinguished. It is interesting that the norm $p \approx 2.05$ is very close to the Euclidean one ($p = 2$). However, it has to be stressed out that there is not any universally preferred norm, since this is dependent on the specific data values.

Both the diagrams of the $\mathcal{L}^p$ mean $\mu_p(p)$ and its relevant error $\delta\mu_p(p)$, depicted in terms of the norm p , constitute a “metricogram”. In a metricogram, we are able to observe the whole spectrum of the $\mathcal{L}^p$ mean and its error, and to recognize the preferable norms.

Figure 6. The error $\delta\langle B \rangle_p$ is depicted with additive equidistributed noise inserted into the $\{B_i\}_{i=1}^N$ values ($N = 46080$). The amplitude of the noise is equal to the resolution of the values $\{B_i\}_{i=1}^N$. Namely, we set $\epsilon = -0.01$ nT (a), and $\epsilon = +0.01$ nT (c). In panel (b) we depict the unperturbed error for convenience. The magnified panels (d), (e) and (f) of the respective panels (a), (b) and (c) demonstrate the minimum error at $p \approx 2.05$ that remains unaffected, for amplitudes of additive noise less or equal to the reading error, *i.e.*, $|\epsilon| \leq 0.01$ nT, in contrast to the fluctuations, appearing for $p - 1 \rightarrow 0$, which are affected by the additive noise.



6. Conclusions

This analysis introduced a possible generalization of the basic statistical concepts of the expectation value and variance for non-Euclidean metrics induced by $\mathcal{L}^p$ norms. The Euclidean $\mathcal{L}^2$ mean is derived by minimizing the sum of the total square deviations TSD , which is the Euclidean variance. Similarly, the non-Euclidean $\mathcal{L}^p$ means were developed by minimizing the sum of the $\mathcal{L}^p$ deviations TD_p^p , which is proportional to the $\mathcal{L}^p$ variance. The main advantage of the new statistical approach is that the p -norm is a free parameter, thus both the $\mathcal{L}^p$ -normed expectation values and their variance are flexible to analyze new phenomena that cannot be described under the notions of classical statistics based on Euclidean norms.

As it was shown, the $\mathcal{L}^p$ means embody a generic formal scheme of means characterization, given the sampling values $\{y_i\}_{i=1}^N$. This involves the existence of a univalued, N -multivariable function $M(\{y_i\}_{i=1}^N)$, fulfilling the following three preconditions: (i) Continuity; (ii) Internness; (iii) Symmetry. (This axiomatic scheme of means characterization generalizes the one proposed by Aczél [23], which led to the Euclidean φ -means.)

The $\mathcal{L}^p$ expectation values can be expressed in terms of the operator $\hat{\mathcal{L}}_p$, which helps to automatically retrieve the non-Euclidean representation of a given formulation, for which the Euclidean representation is known. This idea was utilized to derive the $\mathcal{L}^p$ -variance σ_p^2 , the derivative $\frac{\partial}{\partial \beta} \langle y \rangle_p$ (with respect to a parameter β , $\langle y \rangle_p = \langle y \rangle_p(\beta)$), and the novel representation of several fundamental notions of Statistical Mechanics (based on the $\mathcal{L}^p$ -norm), e.g., Canonical probability distribution and partition function.

The $\mathcal{L}^p$ -mean estimator, $\hat{\mu}_{p,N} = \hat{\mu}_{p,N}(\{y_i\}_{i=1}^N; p)$, of a sampling values $\{y_j\}_{j=1}^N$ of the independent and identically distributed random variables $\{Y_i\}_{i=1}^N$ was defined. It is characterized by the following properties: (i) $\langle \hat{\mu}_{p,N}(\{y_i\}_{i=1}^N; p) \rangle_p = \mu_p$, and (ii) $\hat{\mu}_{p,N} \rightarrow \mu_p$ as $N \rightarrow \infty$, where μ_p is the $\mathcal{L}^p$ -expectation value of each of the $\{Y_i\}_{i=1}^N$ random variables. These properties are similar to the known Euclidean ones, namely, (i) $\langle \frac{1}{N} \sum_{i=1}^N y_i \rangle_2 = \mu_2$, and (ii) $\frac{1}{N} \sum_{i=1}^N y_i \rightarrow \mu_2$ as $N \rightarrow \infty$.

The expression of the $\mathcal{L}^p$ variance σ_p^2 was shown by four different ways: (a) The maximization of the likelihood function, constructed by N independent and identically distributed random variables $\{Y_i\}_{i=1}^N$, led to the $\mathcal{L}^p$ -mean estimator $\hat{\mu}_{p,N}$ and its $\mathcal{L}^p$ variance $\hat{\sigma}_{p,N}^2$, if $Y_i \sim \mathcal{GG}(\mu, \sigma; p)$, $\forall i = 1, \dots, N$, namely, the random variables are distributed according to the General Gaussian distribution with shape parameter equal to the p -norm. (b) The variance σ_p^2 is proportional to the total deviations (residuals) $A_0(p)$ (Equation (49)), and to the inverse of the curvature factor $A_2(p)$, given in Equation (50). The next two ways utilize the fact that the property (ii) of the operator $\hat{\mathcal{L}}_p$ holds if and only if the proportionality factor C , that connects, either the variance with the Total p -Deviations, i.e., $\sigma_p^2 = C \cdot T D_p^p$, or the energy states $\{\varepsilon_k\}_{k=1}^W$ with the “ $\mathcal{L}^p$ energy states” $\{\hat{\mathcal{L}}_p(\varepsilon_k)\}_{k=1}^W$, i.e., $\hat{\mathcal{L}}_p(\varepsilon_k - U_p) = C |\varepsilon_k - U_p|^{p-1} \text{sign}(\varepsilon_k - U_p)$, is given by the one single expression $C = \frac{1}{(p-1)\phi_p}$: (c) Derivation of the Canonical probability distribution of $\mathcal{L}^p$ -normed Statistical Mechanics. (d) Connection with thermodynamics, e.g., $U_p = -\frac{\partial \ln \mathcal{Z}_p}{\partial \beta}$.

We remark the difference between the extracted expression of the $\mathcal{L}^p$ deviation σ_p , and the one considered in the literature, $\tilde{\sigma}_p$, that is

$$\sigma_p = \sqrt{\frac{1}{p-1} \frac{\sum_{k=1}^W \rho_k |y_k - \langle y \rangle_p|^p}{\sum_{k=1}^W \rho_k |y_k - \langle y \rangle_p|^{p-2}}}, \quad \tilde{\sigma}_p = \sqrt[p]{\sum_{k=1}^W \rho_k |y_k - \langle y \rangle_2|^p} \quad (55)$$

which is supposed to represent the generalization of $\sigma_2 = \sqrt{\sum_{k=1}^W \rho_k |y_k - \langle y \rangle_2|^2}$. The correct expressions of the $\mathcal{L}^p$ means and variance can be crucial for providing insights into the fundamental numerical tools of data analysis, such as the moving averages techniques for smoothing, the multifractal detrended fluctuation analysis (MF-DFA) (e.g., see [32,33]), the singular spectrum analysis technique [34], etc. Having found a new class for the mean and variance, it is straightforward to develop an unbiased estimator for the population coefficient of variation (e.g., for Euclidean norm, see [35]).

Several analytical and numerical examples were examined. When the distribution of the data values is symmetric, the whole set of $\mathcal{L}^p$ means degenerates to one single value, while when it is asymmetric, a spectrum-like range of $\mathcal{L}^p$ means is generated. In addition, we dealt with the numerical data of the Earth's magnetic field magnitude.

The mean value is uniquely defined and dependent on the p -norm. The error of the mean is also dependent on the p -norm, leading to a variant weight of the mean (inverse square). In a metricogram, we observe the whole spectrum of the $\mathcal{L}^p$ mean $\mu_p(p)$ and its relevant error $\delta \mu_p(p)$, depicted in terms of the p -norm, where we can detect any preferable norms. For example, for the numerical example of the

Earth's magnetic field magnitude (time series recorded within January of 2008), a preferable norm was found for $p \approx 2.05$, for which $\frac{\partial}{\partial p} \delta \langle B \rangle_p(p) = 0$, $\frac{\partial^2}{\partial p^2} \delta \langle B \rangle_p(p) > 0$.

The classical concept of expectation values was generalized to the non-Euclidean $\mathcal{L}^p$ -normed representation, highlighting its implication in Statistical Mechanics. Indeed, the $\mathcal{L}^p$ expectation value of a given energy spectrum $\{\varepsilon_k\}_{k=1}^W$ represents the non-Euclidean adaptation of the internal energy U_p . This is an issue that has to be considered in Statistical Mechanics: Several pedagogical examples were examined: gas in thermal equilibrium, space plasmas out of thermal equilibrium, and multi-dimensional quantum harmonic oscillator at thermal equilibrium.

References

1. Abramowitz, M.; Stegun, I.A. *Handbook of Mathematical Functions*; Dover Publications: New York, NY, USA, 1965.
2. Kolmogorov, A.N. Sur la notion de la moyenne (in French). *Atti Accad. Naz. Lincei* **1930**, *12*, 388–391.
3. Nagumo, M. Uber eine Klasse der Mittelwerte (in German). *Jpn. J. Math.* **1930**, *7*, 71–79.
4. Hardy, G.H.; Littlewood, J.E.; Pólya, G. *Inequalities*; Cambridge University Press: Cambridge, UK, 1934.
5. Ben-Tal, A. On generalized means and generalized convex functions. *J. Optimiz.* **1977**, *21*, 1–13.
6. Páles, Z. On the characterization of quasi arithmetic means with weight function. *Aequationes Math.* **1987**, *32*, 171–194.
7. Livadiotis, G. Approach to block entropy modeling and optimization. *Physica A* **2008**, *387*, 2471–2494.
8. Bajraktarevic, M. Sur un equation fonctionnelle aux valeurs moyens (in French). *Glas. Mat.-Fiz. Astronom.* **1958**, *13*, 243–248.
9. Aczél, J.; Daroczy, Z. *On Measures of Information and Their Characterization*; Academy Press: New York, NY, USA, 1975.
10. Norries, N. General means and statistical theory. *Am. Stat.* **1976**, *30*, 8–12.
11. Kreps, D.M.; Porteus, E.L. Temporal resolution of uncertainty and dynamic choice theory. *Econometrica* **1978**, *46*, 185–200.
12. Flandrin, P. Separability, positivity, and minimum uncertainty in time-frequency energy distributions. *J. Math. Phys.* **1998**, *39*, 4016–4039.
13. Czachor, M.; Naudts, J. Thermostatistics based on Kolmogorov–Nagumo averages: Unifying framework for extensive and nonextensive generalizations. *Phys. Lett. A* **2002**, *298*, 369–374.
14. Dukkupati, A.; Narasimha-Murty, M.; Bhatnagar, S. Uniqueness of nonextensive entropy under Renyi's recipe. *Comp. Res. Rep.* **2005**, 511078.
15. Tzafestas, S.G. *Applied Control: Current Trends and Modern Methodologies*; Electrical and Computer Engineering Series; Marcel Dekker, Inc.: New York, 1993.
16. Koch, S.J.; Wang, M.D. Dynamic force spectroscopy of protein-DNA interactions by unzipping DNA. *Phys. Rev. Lett.* **2003**, *91*, 028103, doi:10.1103/PhysRevLett.91.028103.
17. Still, S.; Kondor, I. Regularizing portfolio optimization. *New J. Phys.* **2010**, *12*, 075034, doi:10.1088/1367-2630/12/7/075034.

18. Livadiotis, G. Approach to general methods for fitting and their sensitivity. *Physica A* **2007**, *375*, 518–536.
19. Bavaud, F. Aggregation invariance in general clustering approaches. *Adv. Data Anal. Classif.* **2009**, *3*, 205–225.
20. Gajek, L.; Kaluszk, M. Upper bounds for the L_1 -risk of the minimum L_1 -distance regression estimator. *Ann. Inst. Stat. Math.* **1992**, *44*, 737–744.
21. Huber, P. *Robust Statistics*; John Wiley & Sons: New York, NY, USA, 1981.
22. Hampel, F.R.; Ronchetti, E.M.; Rousseeuw, P.J.; Stahel, W.A. *Robust Statistics. The Approach Based on Influence Functions*; John Wiley & Sons: New York, NY, USA, 1986.
23. Aczél, J. On mean values. *Bull. Amer. Math. Soc.* **1948**, *54*, 392–400.
24. Aczél, J. *Lectures on Functional Equations and Their Applications*; Academy Press: New York, NY, USA, 1966.
25. Toader, G.; Toader, S. Means and generalized means. *J. Inequal. Pure Appl. Math.* **2007**, *8*, 45.
26. Livadiotis, G.; McComas, D.J. Beyond kappa distributions: Exploiting Tsallis statistical mechanics in space plasmas. *J. Geophys. Res.* **2009**, *114*, A11105, doi:10.1029/2009JA014352.
27. Tsallis, C. *Introduction to Nonextensive Statistical Mechanics*, Springer: New York, NY, USA, 2009.
28. Livadiotis, G.; McComas, D.J. Invariant kappa distribution in space plasmas out of equilibrium. *Astrophys. J.* **2011**, *741*, 88, doi:10.1088/0004-637X/741/2/88.
29. Livadiotis, G.; McComas, D.J. Measure of the departure of the q -metastable stationary states from equilibrium. *Phys. Scripta* **2010**, *82*, 035003, doi:10.1088/0031-8949/82/03/035003.
30. Livadiotis, G.; McComas, D.J. Exploring transitions of space plasmas out of equilibrium. *Astrophys. J.* **2010**, *714*, 971, doi:10.1088/0004-637X/714/1/971.
31. Melissinos, A.C. *Experiments in Modern Physics*; Academic Press Inc.: London, UK, 1966; pp. 438–464.
32. Kantelhardt, J.; Zschiegner, S.A.; Koscielny-Bunde, E.; Bunde, A.; Havlin, S.; Stanley, H.E. Multifractal detrended fluctuation analysis of nonstationary time series. *Physica A* **2002**, *316*, 87–114.
33. Varotsos, P.A.; Sarlis, N.V.; Skordas, E.S. Attempt to distinguish electric signals of a dichotomous nature. *Phys. Rev. E* **2003**, *68*, 031106, doi:10.1103/PhysRevE.68.031106.
34. Hassani, H.; Thomakos, D. A review on singular spectrum analysis for economic and financial time series. *Stat. Interface* **2010**, *3*, 377–397.
35. Mahmoudvand, R.; Hassani, H. Two new confidence intervals for the coefficient of variation in a normal distribution. *J. Appl. Stat.* **2009**, *36*, 429–442.