MDPI

Article

On the Convergence and Law of Large Numbers for the Non-Euclidean $\mathcal{L}^p$-Means

George Livadiotis

Space Science and Engineering, Southwest Research Institute, San Antonio, TX 78238, USA;
glivadiotis@swri.edu; Tel.: +1-210-522-3415

Abstract: This paper describes and proves two important theorems that compose the Law of Large Numbers for the non-Euclidean $\mathcal{L}^p$-means, known to be true for the Euclidean $\mathcal{L}^2$-means: Let the $\mathcal{L}^p$-mean estimator, which constitutes the specific functional that estimates the $\mathcal{L}^p$-mean of N independent and identically distributed random variables; then, (i) the expectation value of the $\mathcal{L}^p$-mean estimator equals the mean of the distributions of the random variables; and (ii) the limit $N \to \infty$ of the $\mathcal{L}^p$-mean estimator also equals the mean of the distributions.

1. Introduction: Definition of $\mathcal{L}^p$-Means and Their Basic Properties

In [1–3], a 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 a special case for $p = 1$ ($\mathcal{L}^1$) and the ordinary Euclidean mean for $p = 2$ ($\mathcal{L}^2$) (see also: [4,5]). Let the set of y-values $\{y_k\}_{k=1}^W$ ($y_k \in D_y \subseteq \Re$), associated with the probabilities $\{p_k\}_{k=1}^W$; then, the non-Euclidean means μ_p, based on $\mathcal{L}^p$-norms, are defined by

$$\sum_{k=1}^{W} p_k |y_k - \mu_p|^{p-1} sign(y_k - \mu_p) = 0 , \tag{1}$$

where the median μ_1 and the arithmetic mean μ_2 follow as special cases when the Taxicab $\mathcal{L}^1$ and Euclidean $\mathcal{L}^2$-norms are respectively considered. Both the median μ_1 and arithmetic μ_2 means can be implicitly written in the form of Equation (1) as $\sum_{k=1}^W p_k sign(y_k - \mu_1) = 0$ and $\sum_{k=1}^W p_k |y_k - \mu_2| sign(y_k - \mu_2) = 0$ ($\Leftrightarrow \mu_2 = \sum_{k=1}^W p_k y_k$), respectively.

Note that the solution of Equation (1) is a specific case of the so-called *M*-estimators [6], while it is also related to the Fréchet Means [7]. The Euclidean norm $\mathcal{L}^2$ is also known as "Pythagorean" norm. In [3], we preferred 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. As an example of an application in physics, the $\mathcal{L}^p$ expectation value of an energy spectrum $\{\varepsilon_k\}_{k=1}^W$ is defined by representing the non-Euclidean adaptation of internal energy U_p [8].

Figure 1 illustrates an example of $\mathcal{L}^p$-means. We use the Poisson distribution $p_k = e^{-\lambda} \lambda^k / k!$ and the dataset $y_k = k$, for $k = 1, ..., W$; hence, the $\mathcal{L}^p$-means are implicitly given by $\sum_{k=1}^W \frac{\lambda^k}{k!} |k - \mu_p|^{p-1} sign(k - \mu_p) = 0$ (note that the constant term $e^{-\lambda}$ can be ignored). The function $\mu_p = \mu_p(\lambda)$ is examined for various values of the p-norm, either (a) super-Euclidean, $p > 2$, or (b) sub-Euclidean $p < 2$. The mean value for the Euclidean case, $p = 2$, is $\mu_2 = \lambda$, which is represented by

the diagonal line in both panels. We observe that for $p > 2$ we always have $\mu_p > \lambda$, while for $p < 2$ there is a critical value $\lambda^*(p)$, for which $\mu_p > \lambda$ for $\lambda > \lambda^*$ and $\mu_p < \lambda$ for $\lambda < \lambda^*$. The critical value $\lambda^*(p)$ increases with p, and as $p \to 2$, $\mu_p \to \lambda$. For $\lambda = 1$, $\mu_p = 1$ for any $p \leq 2$, while for $\lambda = 0$, $\mu_p = 0$ for any values of p.
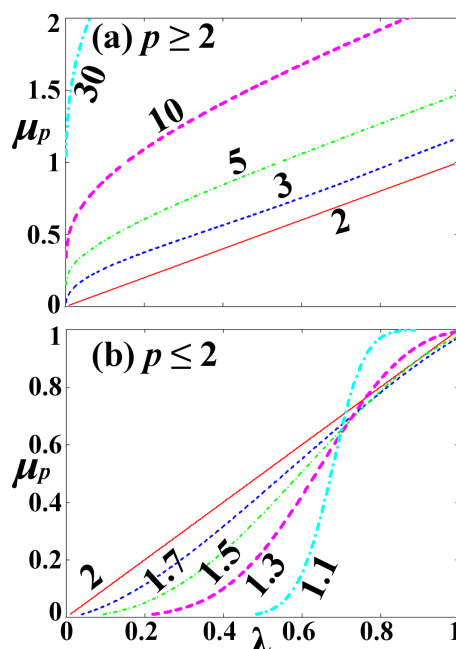

Figure 1. Example of $\mathcal{L}^p$-means of a dataset following the Poisson distribution. The relation $\mu_p = f(\lambda)$ is plotted for (**a**) $p \geq 2$, i.e., $p = 2$ (red solid line), $p = 3$ (blue dash), $p = 5$ (green dash–dot), $p = 10$ (purple thick dash), $p = 30$ (light-blue thick dash–dot); and (**b**) $p \leq 2$, i.e., $p = 2$ (red solid line), $p = 1.7$ (blue dash), $p = 1.5$ (green dash–dot), $p = 1.3$ (purple thick dash), $p = 1.1$ (light-blue thick dash–dot).

The Law of Large Numbers is a theorem that guarantees the stability of long-term averages of random events, but is valid only for Euclidean metrics based on L^2-norms. The purpose of this paper is to extend the theorem of the "Law of Large Numbers" to the non-Euclidean, $\mathcal{L}^p$-means. Namely, (i) the expectation value of the $\mathcal{L}^p$-mean estimator (that corresponds to Equation (1)) equals the mean of the distribution of each of the random variables; and (ii) the limit $N \to \infty$ of the $\mathcal{L}^p$-mean estimator also equals the mean of the distributions. These are numbered as *Theorems 2 and 3*, respectively. The paper is organized as follows: In Section 2, we prove the theorem of uniqueness of the $\mathcal{L}^p$-means (*Theorem 1*). This will be used in the proofs of *Theorems 2 and 3*, shown in Sections 3 and 4, respectively. Finally, Section 5 briefly summarizes the conclusions. Several examples are used to illustrate the validity of the Theorems 1–3, that is, the Poisson distribution (discrete description) and a superposition of normal distributions (continuous description).

2. Uniqueness of $\mathcal{L}^p$-Means

Here, we show the theorem of uniqueness of the $\mathcal{L}^p$-means for any $p > 1$. The theorem will be used in the *Theorem 2 and 3* of the next sections.

Theorem 1. *The curve $\mu_p(p)$ is univalued, namely, for each $p > 1$, there is a unique value of the $\mathcal{L}^p$-mean $\mu_p(p)$.*

Proof of Theorem 1. Using the implicit function theorem [9], we can easily show the uniqueness in a sufficiently small neighbourhood of $p = 2$. Indeed, there is at least one point, that is the Euclidean point $(p = 2, \mu_p = \mu_2)$, for which the function $\mu_p(p)$ exists and is univalued. Then, the values of $\mu_p(p)$,

$\forall p > 1$, can be approximated to any accuracy, starting from the Euclidean point. The implicit function $F(p, \mu_p) = 0$, defined by Equation (1), is continuous and $\partial F(p, \mu_p)/\partial \mu_p = (p-1) \sum_{k=1}^{W} p_k |y_k - \mu_p|^{p-2} \neq 0$; then $\mu_p(p)$ is univalued in some domain around $p = 2$. The first derivative $\mu_p'(p)$ is finite $\forall p > 1$ (e.g., for $p = 2$ we have $\mu_2' = [\partial \mu_p(p)/\partial p]_{p=2} = \sum_{k=1}^{W} p_k (y_k - \mu_2) \ln(|y_k - \mu_2|)$). Indeed, the inverse derivative is non-zero for any p, i.e.,

$$\frac{\partial p}{\partial \mu_p} = \frac{(p-1) \sum_{k=1}^{W} p_k |y_k - \mu_p|^{p-2}}{\sum_{k=1}^{W} p_k |y_k - \mu_p|^{p-1} sign(y_k - \mu_p) \ln(|y_k - \mu_p|)}, \forall p > 1. \tag{2}$$

$\square$

The inverse function, $p(\mu_p)$, should be continuous and differentiable according to Equation (2). If $\mu_p(p)$ were multi-valued, then, it should have local minima or maxima. However, the derivative $dp/d\mu_p$ is non-zero. Therefore, we conclude that $p(\mu_p)$ cannot be multi-valued, and there is a unique curve $\mu_p(p)$ that passes through ($p = 2, \mu_p = \mu_2$).

As an example, Figure 2 plots the $\mathcal{L}^p$-means of the Poisson distribution shown in Figure 1, but now as a function of the p-norm, and for various values of $0 < \lambda < 1$. For $\lambda < \ln 2$, the function $\mu_p(p)$ is monotonically increasing with p. On the contrary, for $\lambda > \ln 2$, the function $\mu_p(p)$ is not monotonic, having a minimum in the region of sub-Euclidean norms, $1 < p < 2$. The separatrix between these two behaviors of $\mu_p(p)$ is given for $\lambda = \ln 2$. We observe that the function $\mu_p(p)$ is differentiable, $\partial \mu_p/\partial p$ is always finite or $\partial p/\partial \mu_p$ is always non-zero, thus $\mu_p(p)$ is unique for any value of p.

Finally, we note that the uniqueness of μ_p for a given p does not ensure monotonicity, as different values of p may lead to the same $\mathcal{L}^p$-mean. Such an example is the $\mathcal{L}^p$-means of the Poisson distribution for $\lambda > \ln 2$, shown in Figure 2. As stated and illustrated in [3], when the examined probability distribution is symmetric, then 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 rather generated.
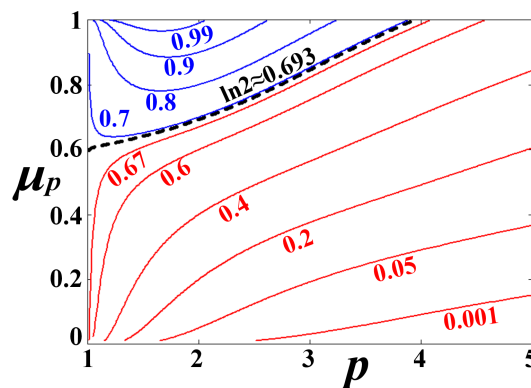

Figure 2. Uniqueness of the $\mathcal{L}^p$-means of the Poisson distribution. The means are plotted as a function of the p-norm, and for various values of $0 < \lambda < 1$, that is, $\lambda < \ln 2$ (red solid), $\lambda = \ln 2$ (black dash), and $\lambda > \ln 2$ (blue solid).

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

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

$$\sum_{i=1}^{N} \left| y_i - \widehat{\mu}_{p,N}(\{y_j\}_{j=1}^{N}; p) \right|^{p-1} sign\left[y_i - \widehat{\mu}_{p,N}(\{y_j\}_{j=1}^{N}; p) \right] = 0. \tag{3}$$

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

$$\int \cdots \int_{\{y_j \in D_y\}_{j=1}^N} \left| \widehat{\mu}_{p,N}(\{y_j\}_{j=1}^N; p) - \langle \widehat{\mu}_{p,N} \rangle_p \right|^{p-1} sign \left[\widehat{\mu}_{p,N}(\{y_j\}_{j=1}^N; p) - \langle \widehat{\mu}_{p,N} \rangle_p \right]$$
$$\times \, \mathcal{P}(\{y_j\}_{j=1}^N) \, dy_1 \ldots dy_N = 0 \,, \tag{4}$$

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

$$\int \cdots \int_{\{y_j \in D_y\}_{j=1}^N} \mathcal{P}(\{y_j\}_{j=1}^N) \, dy_1 \ldots dy_N = 1 \,. \tag{5}$$

Definition 1. *Let the sampling $\{y_i\}_{i=1}^N$, $y_i \in D_y \subseteq \Re, \forall i = 1, \ldots, N$, of the set of random variables $\{Y_i\}_{i=1}^N$. This set is called symmetrically distributed if the joint distribution density has the property $\mathcal{P}(\{y_j\}_{j=1}^N) = \mathcal{P}(y_1 \ldots y_k \ldots y_i \ldots y_N) = \mathcal{P}(y_1 \ldots y_i \ldots y_k \ldots y_N), \forall i, k(\neq i) = 1, \ldots, N$. This property is formally called Exchange-ability [10] and will be used in Lemmas 1 and 2.*

　　Next, we postulate and prove Lemmas 1 and 2, which are necessary for the following Theorem 2 about the expectation value of the $\mathcal{L}^p$-mean estimator.

Lemma 1. *The symmetrically distributed random variables $\{Y_i\}_{i=1}^N$ are characterized by the same $\mathcal{L}^p$ expectation value, namely, $\langle Y_i \rangle_p = \hat{E}_p(Y_i) = \mu_p \in \Re, \forall i = 1, \ldots, N$, which is implicitly given by*

$$\int \cdots \int_{\{y_j \in D_y\}_{j=1}^N} \left| y_i - \mu_p \right|^{p-1} sign \left(y_i - \mu_p \right) \mathcal{P}(\{y_j\}_{j=1}^N) \, dy_1 \ldots dy_N$$
$$= \int_{y_i \in D_y} \left| y_i - \mu_p \right|^{p-1} sign \left(y_i - \mu_p \right) \mathcal{P}_y(y_i) \, dy_i = 0 \,, \tag{6}$$

where $\mathcal{P}_{y_i}(u) \equiv \mathcal{P}_y(u), \forall i = 1, \ldots, N$, is the marginal distribution density, that is identical for all the random variables $\{Y_i\}_{i=1}^N$.

Proof of Lemma 1. The y_i-marginal probability density, $\mathcal{P}_{y_i}(y_i)$, is

$$\mathcal{P}_{y_i}(y_i) = \int \cdots \int_{\{y_j \in D_y\}_{j=1, j \neq i}^N} \mathcal{P}(\{y_j\}_{j=1}^N) \, dy_1 \ldots dy_{i-1} dy_{i+1} \ldots dy_N \,, \tag{7}$$

so that

$$\int_{y_i \in D_y} \mathcal{P}_{y_i}(y_i) \, dy_i = 1 \,. \tag{8}$$

Given the symmetrical joint distribution, we have

$$\mathcal{P}_{y_i}(y_i) = \int \cdots \int_{\{y_j \in D_y\}_{j=1, j \neq i}^N} \mathcal{P}(y_1 \ldots y_i \ldots y_k \ldots y_N) \, dy_1 \ldots dy_{i-1} dy_{i+1} \ldots dy_N$$
$$= \left\{ \int \cdots \int_{\{y_j \in D_y\}_{j=1, j \neq i}^N} \mathcal{P}(y_1 \ldots y_k \ldots y_i \ldots y_N) \, dy_1 \ldots dy_{k-1} dy_{k+1} \ldots dy_N \right\}_{y_k = y_i}$$
$$= \left\{ \int \cdots \int_{\{y_j \in D_y\}_{j=1, j \neq i}^N} \mathcal{P}(y_1 \ldots y_i \ldots y_k \ldots y_N) \, dy_1 \ldots dy_{k-1} dy_{k+1} \ldots dy_N \right\}_{y_k = y_i} \tag{9}$$
$$= \mathcal{P}_{y_k}(y_i) \, \forall i, k(\neq i) = 1, \ldots, N \,.$$

　　Hence, the expression of the marginal distribution density $\mathcal{P}_{y_i}(u)$ is identical $\forall i = 1, \ldots, N$, i.e., for all the random variables, $\mathcal{P}_{y_i}(u) \equiv \mathcal{P}_y(u)$.

Then, we readily derive that the random variables $\{Y_i\}_{i=1}^N$ are characterized by the same $\mathcal{L}^p$ expectation value, namely, $\langle Y_i \rangle_p = \hat{E}_p(Y_i) = \mu_p \in \Re, \forall\, i = 1, \ldots, N$, which is implicitly expressed by

$$
\begin{aligned}
&\int \cdots \int_{\{y_j \in D_y\}_{j=1}^N} \left| y_i - \mu_p \right|^{p-1} sign\left(y_i - \mu_p \right) \mathcal{P}(\{y_j\}_{j=1}^N)\, dy_1 \ldots dy_N \\
&= \int_{y_i \in D_y} \left| y_i - \mu_p \right|^{p-1} sign\left(y_i - \mu_p \right) \mathcal{P}_y(y_i)\, dy_i = 0.
\end{aligned}
\tag{10}
$$

Indeed, if we had $\langle Y_i \rangle_p = \mu_{pi} \in \Re,\, \forall\, i = 1, \ldots, N$, then

$$
\int_{y_i \in D_y} \left| y_i - \mu_{pi} \right|^{p-1} sign\left(y_i - \mu_{pi} \right) \mathcal{P}_y(y_i)\, dy_i = 0 \,,
\tag{11}
$$

and for $k \neq i$,

$$
\begin{aligned}
&\int_{y_k \in D_y} \left| y_k - \mu_{pk} \right|^{p-1} sign\left(y_k - \mu_{pk} \right) \mathcal{P}_y(y_k)\, dy_k = 0 \,, \\
\Rightarrow\ &\int_{y_i \in D_y} \left| y_i - \mu_{pk} \right|^{p-1} sign\left(y_i - \mu_{pk} \right) \mathcal{P}_y(y_i)\, dy_i = 0.
\end{aligned}
\tag{12}
$$

However, given the uniqueness of the $\mathcal{L}^p$-means, Equations (11) and (12) lead to $\mu_{pi} = \mu_{pk}\, \forall\, i, k(\neq i) = 1, \ldots, N$, or $\mu_{pi} = \mu_p\,, \forall\, i = 1, \ldots, N$. $\square$

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

Proof of Lemma 2. The $\mathcal{L}^p$ expectation value of $\langle G_i \rangle_p$ is implicitly given by

$$
\begin{aligned}
&\int \cdots \int_{\{y_j \in D_y\}_{j=1}^N} \left| y_i - \hat{\mu}_{p,N}(\{y_j\}_{j=1}^N; p) - \langle G_i \rangle_p \right|^{p-1} sign\left[y_i - \hat{\mu}_{p,N}(\{y_j\}_{j=1}^N; p) - \langle G_i \rangle_p \right] \\
&\quad \times \mathcal{P}(\{y_j\}_{j=1}^N)\, dy_1 \ldots dy_N = 0.
\end{aligned}
\tag{13}
$$

If $\langle G_i \rangle_p = 0$, then

$$
\begin{aligned}
&\int \cdots \int_{\{y_j \in D_y\}_{j=1}^N} \left| y_i - \hat{\mu}_{p,N}(\{y_j\}_{j=1}^N; p) \right|^{p-1} sign\left[y_i - \hat{\mu}_{p,N}(\{y_j\}_{j=1}^N; p) \right] \\
&\quad \times \mathcal{P}(\{y_j\}_{j=1}^N)\, dy_1 \ldots dy_N = 0,
\end{aligned}
\tag{14}
$$

while if $\langle G_i \rangle_p \neq 0$, then the above functional has to be non-zero, because of the uniqueness of $\mathcal{L}^p$ expectation values, namely,

$$
\begin{aligned}
&\int \cdots \int_{\{y_j \in D_y\}_{j=1}^N} \left| y_i - \hat{\mu}_{p,N}(\{y_j\}_{j=1}^N; p) \right|^{p-1} sign\left[y_i - \hat{\mu}_{p,N}(\{y_j\}_{j=1}^N; p) \right] \\
&\quad \times \mathcal{P}(\{y_j\}_{j=1}^N)\, dy_1 \ldots dy_N = C_i(p, N) \neq 0,
\end{aligned}
\tag{15}
$$

Now, rewriting Equation (15) for an index $k(\neq i)$, we have

$$
\begin{aligned}
&\int \cdots \int_{\{y_j \in D_y\}_{j=1}^N} \left| y_k - \widehat{\mu}_{p,N}(\{y_j\}_{j=1}^N; p) \right|^{p-1} sign \left[y_k - \widehat{\mu}_{p,N}(\{y_j\}_{j=1}^N; p) \right] \\
&\times \mathcal{P}(\{y_j\}_{j=1}^N) \, dy_1 \ldots dy_N = C_k(p,N) \\
&= \int \cdots \int_{\{y_j \in D_y\}_{j=1}^N} \left| y_k - \widehat{\mu}_{p,N}(y_1 \ldots y_i \ldots y_k \ldots y_N; p) \right|^{p-1} sign \left[y_i - \widehat{\mu}_{p,N}(y_1 \ldots y_i \ldots y_k \ldots y_N; p) \right] \\
&\times \mathcal{P}(y_1 \ldots y_i \ldots y_k \ldots y_N) \, dy_1 \ldots dy_i \ldots dy_k \ldots dy_N \\
&= \int \cdots \int_{\{y_j \in D_y\}_{j=1}^N} \left| y_i - \widehat{\mu}_{p,N}(y_1 \ldots y_k \ldots y_i \ldots y_N; p) \right|^{p-1} sign \left[y_i - \widehat{\mu}_{p,N}(y_1 \ldots y_i \ldots y_k \ldots y_N; p) \right] \\
&\times \mathcal{P}(y_1 \ldots y_k \ldots y_i \ldots y_N) \, dy_1 \ldots dy_k \ldots dy_i \ldots dy_N \\
&= \int \cdots \int_{\{y_j \in D_y\}_{j=1}^N} \left| y_i - \widehat{\mu}_{p,N}(y_1 \ldots y_i \ldots y_k \ldots y_N; p) \right|^{p-1} sign \left[y_i - \widehat{\mu}_{p,N}(y_1 \ldots y_k \ldots y_i \ldots y_N; p) \right] \\
&\times \mathcal{P}(y_1 \ldots y_i \ldots y_k \ldots y_N) \, dy_1 \ldots dy_i \ldots dy_k \ldots dy_N = C_i(p,N),
\end{aligned}
\tag{16}
$$

because of the symmetrical distribution of random variables $\{y_j\}_{j=1}^N$, i.e., $\mathcal{P}(y_1 \ldots y_k \ldots y_i \ldots y_N) = \mathcal{P}(y_1 \ldots y_i \ldots y_k \ldots y_N)$, $\forall i, k(\neq i) = 1, \ldots, N$, (the same symmetry holds also for the estimator $\widehat{\mu}_{p,N}$, while the integration on each y_i spans the same integral D_y. Hence, $C_i(p,N) = C_k(p,N) \equiv C(p,N)$. Then, by summing both sides of Equation (15) with $\sum_{i=1}^N$, we conclude in

$$
\int \cdots \int_{\{y_j \in D_y\}_{j=1}^N} \sum_{i=1}^N \left| y_i - \widehat{\mu}_{p,N}(\{y_j\}_{j=1}^N; p) \right|^{p-1} sign \left[y_i - \widehat{\mu}_{p,N}(\{y_j\}_{j=1}^N; p) \right]
$$

$$
\times \mathcal{P}(\{y_j\}_{j=1}^N) \, dy_1 \ldots dy_N = 0 = \sum_{i=1}^N C_i(p,N) = N\, C(p,N),
$$

or $C(p,N) = 0$. Thus, Equation (14) holds, and given the uniqueness of $\mathcal{L}^p$ expectation values, we conclude in $\langle G_i \rangle_p = 0$, $\forall i = 1, \ldots, N$. $\quad\square$

Theorem 2. *Consider the sampling $\{y_i\}_{i=1}^N$, $y_i \in D_y \subseteq \Re$, $\forall i = 1, \ldots, N$, of the symmetrically distributed random variables $\{Y_i\}_{i=1}^N$. According to* Lemma 1, *the random variables are characterized by the same $\mathcal{L}^p$ expectation value (assuming that this exists), namely, $\langle Y_i \rangle_p = \hat{E}_p(Y_i) = \mu_p \in \Re$, $\forall i = 1, \ldots, N$, which is implicitly expressed by Equation (6). Then, the $\mathcal{L}^p$ expectation value of the $\mathcal{L}^p$-mean estimator $\widehat{\mu}_{p,N}(\{y_j\}_{j=1}^N; p)$ is equal to μ_p, i.e., $\langle \widehat{\mu}_{p,N} \rangle_p = \hat{E}_p[\widehat{\mu}_{p,N}(\{y_j\}_{j=1}^N; p)] = \mu_p$ or*

$$
\begin{aligned}
&\int \cdots \int_{\{y_j \in D_y\}_{j=1}^N} \left| \widehat{\mu}_{p,N}(\{y_j\}_{j=1}^N; p) - \mu_p \right|^{p-1} sign \left[\widehat{\mu}_{p,N}(\{y_j\}_{j=1}^N; p) - \mu_p \right] \\
&\times \mathcal{P}(\{y_j\}_{j=1}^N) \, dy_1 \ldots dy_N = 0.
\end{aligned}
\tag{17}
$$

Proof of Theorem 2. (For useful inequalities, see [11]) Apparently, the following integral inequalities hold:

$$
\begin{aligned}
0 = &\left| \int \cdots \int_{\{y_j \in D_y\}_{j=1}^N} \left| y_i - \widehat{\mu}_{p,N}(\{y_j\}_{j=1}^N; p) \right|^{p-1} sign \left[y_i - \widehat{\mu}_{p,N}(\{y_j\}_{j=1}^N; p) \right] \right. \\
&\left. \times \mathcal{P}(\{y_j\}_{j=1}^N) \, dy_1 \ldots dy_N \right| \\
\leq &\int \cdots \int_{\{y_j \in D_y\}_{j=1}^N} \left| y_i - \widehat{\mu}_{p,N}(\{y_j\}_{j=1}^N; p) \right|^{p-1} \mathcal{P}(\{y_j\}_{j=1}^N) \, dy_1 \ldots dy_N,
\end{aligned}
\tag{18}
$$

and, $\forall\, i = 1, \ldots, N$,

$$
\begin{aligned}
0 &= \left| \int_{y_i \in D_y} |y_i - \mu_p|^{p-1} sign(y_i - \mu_p) \mathcal{P}_y(y_i)\, dy_i \right| \\
&\leq \int_{y_i \in D_y} |y_i - \mu_p|^{p-1} \mathcal{P}_y(y_i)\, dy_i.
\end{aligned}
\tag{19}
$$

Furthermore, we consider the $\mathcal{L}^p$ expectation value of the functional $g(\{y_j\}_{j=1}^N; p) \equiv \widehat{\mu}_{p,N}(\{y_j\}_{j=1}^N; p) - \mu_p$, namely, $\langle g \rangle_p = \hat{E}_p(g(\{y_j\}_{j=1}^N; p))$, which is implicitly given by

$$
\begin{aligned}
&\int \cdots \int_{\{y_j \in D_y\}_{j=1}^N} \left| \widehat{\mu}_{p,N}(\{y_j\}_{j=1}^N; p) - \mu_p - \langle g \rangle_p \right|^{p-1} \\
&\times sign \left[\widehat{\mu}_{p,N}(\{y_j\}_{j=1}^N; p) - \mu_p - \langle g \rangle_p \right] \mathcal{P}(\{y_j\}_{j=1}^N)\, dy_1 \ldots dy_N = 0.
\end{aligned}
\tag{20}
$$

If $\langle g \rangle_p = 0$, then,

$$
\begin{aligned}
&\int \cdots \int_{\{y_j \in D_y\}_{j=1}^N} \left| \widehat{\mu}_{p,N}(\{y_j\}_{j=1}^N; p) - \mu_p \right|^{p-1} sign \left[\widehat{\mu}_{p,N}(\{y_j\}_{j=1}^N; p) - \mu_p \right] \\
&\times \mathcal{P}(\{y_j\}_{j=1}^N)\, dy_1 \ldots dy_N = 0,
\end{aligned}
\tag{21}
$$

while if $\langle g \rangle_p \neq 0$, then the above functional has to be non-zero, because of the uniqueness of $\mathcal{L}^p$ expectation values, namely,

$$
\begin{aligned}
&\int \cdots \int_{\{y_j \in D_y\}_{j=1}^N} \left| \widehat{\mu}_{p,N}(\{y_j\}_{j=1}^N; p) - \mu_p \right|^{p-1} sign \left[\widehat{\mu}_{p,N}(\{y_j\}_{j=1}^N; p) - \mu_p \right] \\
&\times \mathcal{P}(\{y_j\}_{j=1}^N)\, dy_1 \ldots dy_N = D(p, N) \neq 0,
\end{aligned}
\tag{22}
$$

or

$$
\begin{aligned}
|D(p, N)| &= \left| \int \cdots \int_{\{y_j \in D_y\}_{j=1}^N} \left| \widehat{\mu}_{p,N}(\{y_j\}_{j=1}^N; p) - \mu_p \right|^{p-1} sign \left[\widehat{\mu}_{p,N}(\{y_j\}_{j=1}^N; p) - \mu_p \right] \right. \\
&\quad \left. \times \mathcal{P}(\{y_j\}_{j=1}^N)\, dy_1 \ldots dy_N \right| \\
&\leq \int \cdots \int_{\{y_j \in D_y\}_{j=1}^N} \left| \widehat{\mu}_{p,N}(\{y_j\}_{j=1}^N; p) - \mu_p \right|^{p-1} \mathcal{P}(\{y_j\}_{j=1}^N)\, dy_1 \ldots dy_N.
\end{aligned}
\tag{23}
$$

$\square$

First case, $p \leq 2$: Hence, $p - 1 \leq 1$, and from the power inequality $(|f| + |g|)^s \leq |f|^s + |g|^s$ holding $\forall\, s \leq 1$, we have the following:

The triangle inequality gives $|\widehat{\mu}_{p,N}(\{y_j\}_{j=1}^N; p) - \mu_p| \leq |\widehat{\mu}_{p,N}(\{y_j\}_{j=1}^N; p) - y_i| + |y_i - \mu_p|$. Then, applying the above power inequality for $s = p - 1$, $f = \widehat{\mu}_{p,N}(\{y_j\}_{j=1}^N; p) - y_i$, and $g = y_i - \mu_p$, we have $|\widehat{\mu}_{p,N}(\{y_j\}_{j=1}^N; p) - \mu_p|^{p-1} \leq (|\widehat{\mu}_{p,N}(\{y_j\}_{j=1}^N; p) - y_i| + |y_i - \mu_p|)^{p-1} \leq |\widehat{\mu}_{p,N}(\{y_j\}_{j=1}^N; p) - y_i|^{p-1} + |y_i - \mu_p|^{p-1}$. Thereafter, Equation (23) becomes

$$
\begin{aligned}
&|D(p, N)| \leq \\
&\int \cdots \int_{\{y_j \in D_y\}_{j=1}^N} \left| \widehat{\mu}_{p,N}(\{y_j\}_{j=1}^N; p) - y_i \right|^{p-1} \mathcal{P}(\{y_j\}_{j=1}^N)\, dy_1 \ldots dy_N \\
&+ \int \cdots \int_{\{y_j \in D_y\}_{j=1}^N} |y_i - \mu_p|^{p-1} \mathcal{P}(\{y_j\}_{j=1}^N)\, dy_1 \ldots dy_N, \\
&\Rightarrow |D(p, N)| \leq \\
&\int \cdots \int_{\{y_j \in D_y\}_{j=1}^N} \left| \widehat{\mu}_{p,N}(\{y_j\}_{j=1}^N; p) - y_i \right|^{p-1} \mathcal{P}(\{y_j\}_{j=1}^N)\, dy_1 \ldots dy_N \\
&+ \int_{y_i \in D_y} |y_i - \mu_p|^{p-1} \mathcal{P}_y(y_i)\, dy_i,
\end{aligned}
\tag{24}
$$

or

$$\frac{1}{2}|D(p,N)| \leq Max\left\{ \int\cdots\int_{\{y_j\in D_y\}_{j=1}^N} \left|\widehat{\mu}_{p,N}(\{y_j\}_{j=1}^N;p) - y_i\right|^{p-1} \right.$$

$$\left. \mathcal{P}(\{y_j\}_{j=1}^N)\, dy_1\dots dy_N \, , \int_{y_i\in D_y} |y_i - \mu_p|^{p-1}\, \mathcal{P}_y(y_i)\, dy_i \right\}. \tag{25}$$

Second case, $p \geq 2$: Hence, $p - 1 \geq 1$, and applying the Minkowski inequality on Equation (23), we have

$$|D(p,N)|^{\frac{1}{p-1}} \leq$$

$$\left[\int\cdots\int_{\{y_j\in D_y\}_{j=1}^N} \left|\widehat{\mu}_{p,N}(\{y_j\}_{j=1}^N;p) - \mu_p\right|^{p-1} \mathcal{P}(\{y_j\}_{j=1}^N)\, dy_1\dots dy_N \right]^{\frac{1}{p-1}}$$

$$\leq \left[\int\cdots\int_{\{y_j\in D_y\}_{j=1}^N} \left|\widehat{\mu}_{p,N}(\{y_j\}_{j=1}^N;p) - y_i\right|^{p-1} \mathcal{P}(\{y_j\}_{j=1}^N)\, dy_1\dots dy_N \right]^{\frac{1}{p-1}} \tag{26}$$

$$+ \left[\int_{y_i\in D_y} |y_i - \mu_p|^{p-1}\, \mathcal{P}_y(y_i) \right]^{\frac{1}{p-1}} dy_i \, .$$

or

$$\frac{1}{2^{p-1}}|D(p,N)| \leq$$

$$Max\left\{ \int\cdots\int_{\{y_j\in D_y\}_{j=1}^N} \left|\widehat{\mu}_{p,N}(\{y_j\}_{j=1}^N;p) - y_i\right|^{p-1} \mathcal{P}(\{y_j\}_{j=1}^N)\, dy_1\dots dy_N, \right.$$

$$\left. \int_{y_i\in D_y} |y_i - \mu_p|^{p-1}\, \mathcal{P}_y(y_i)\, dy_i \right\}. \tag{27}$$

Combining Equations (25) and (27), we conclude in an inequality that holds $\forall\, p \geq 1$,

$$0 \neq Min\left\{ \left[\frac{1}{2}|D(p,N)|\right]_{p\leq 2}, \left[\frac{1}{2^{p-1}}|D(p,N)|\right]_{p\geq 2} \right\} \equiv \tilde{D}(p,N) \leq$$

$$Max\left\{ \int\cdots\int_{\{y_j\in D_y\}_{j=1}^N} \left|\widehat{\mu}_{p,N}(\{y_j\}_{j=1}^N;p) - y_i\right|^{p-1} \mathcal{P}(\{y_j\}_{j=1}^N)\, dy_1\dots dy_N, \right. \tag{28}$$

$$\left. \int_{y_i\in D_y} |y_i - \mu_p|^{p-1}\, \mathcal{P}_y(y_i)\, dy_i \right\}.$$

On the other hand, Equations (18) and (19) imply that

$$0 \leq Max\left\{ \int\cdots\int_{\{y_j\in D_y\}_{j=1}^N} \left|\widehat{\mu}_{p,N}(\{y_j\}_{j=1}^N;p) - y_i\right|^{p-1} \mathcal{P}(\{y_j\}_{j=1}^N)\, dy_1\dots dy_N, \right.$$

$$\left. \int_{y_i\in D_y} |y_i - \mu_p|^{p-1}\, \mathcal{P}_y(y_i)\, dy_i \right\}. \tag{29}$$

We construct the auxiliary random variables $\{X_i\}_{i=1}^N$, defined by $X_i = f_x(Y_i) \equiv Y_i \cdot \tilde{D}(p,N)^{-\frac{1}{p-1}}$, having values $\{x_i = f_x(y_i)\}_{i=1}^N$ in the domain $x_i \in D_x \equiv \{f_x(y_{Min}) \leq x \leq f_x(y_{Max})\} \subseteq \Re, \forall\, i = 1,\dots,N$ (where $y_{Min} \equiv D_{y,Min} \in D_y$ is the infimum of D_y, while $y_{Max} \equiv D_{y,Max} \in D_y$ is the supremum of D_y). The $\mathcal{L}^p$-mean estimator of the set $\{x_i\}_{i=1}^N$ is given by the functional $\mu_{p,N}^X = \mu_{p,N}^X(\{x_j\}_{j=1}^N;p) = \widehat{\mu}_{p,N}(\{y_j = x_j\tilde{D}(p,N)^{\frac{1}{p-1}}\}_{j=1}^N;p)\tilde{D}(p,N)^{-\frac{1}{p-1}}$, while the random variables $\{X_i\}_{i=1}^N$ have the common $\mathcal{L}^p$ expectation value $\langle X_i\rangle_p \equiv \hat{E}_p(X_i) = \mu_p^X = \mu_p\tilde{D}(p,N)^{-\frac{1}{p-1}} \in \Re, \forall\, i = 1,\dots,N$. The respective

joint probability density is given by $\mathcal{P}^X(\{x_j\}_{j=1}^N) = \mathcal{P}[\{y_j = x_j \tilde{D}(p,N)^{\frac{1}{p-1}}\}_{j=1}^N] \cdot \tilde{D}(p,N)^{\frac{N}{p-1}}$, so that, $\mathcal{P}^X(\{x_j\}_{j=1}^N)dx_1...dx_N = \mathcal{P}(\{y_j\}_{j=1}^N)dy_1...dy_N$. Then, Equations (28) and (29) become

$$1 \le Max\left\{ \int \cdot \cdot \int_{\{x_j \in D_x\}_{j=1}^N} \left| \mu_{p,N}^X(\{x_j\}_{j=1}^N; p) - x_i \right|^{p-1} \mathcal{P}^X(\{x_j\}_{j=1}^N) \, dx_1...dx_N, \right.$$
$$\left. \int_{x_i \in D_x} \left| x_i - \mu_p^X \right|^{p-1} \mathcal{P}_x^X(x_i) \, dx_i \right\}, \tag{30}$$

and

$$0 \le Max\left\{ \int \cdot \cdot \int_{\{x_j \in D_x\}_{j=1}^N} \left| \mu_{p,N}^X(\{x_j\}_{j=1}^N; p) - x_i \right|^{p-1} \mathcal{P}^X(\{x_j\}_{j=1}^N) \, dx_1...dx_N, \right.$$
$$\left. \int_{x_i \in D_x} \left| x_i - \mu_p^X \right|^{p-1} \mathcal{P}_x^X(x_i) \, dx_i \right\}, \tag{31}$$

respectively (where $\mathcal{P}_{x_i}^X(u) = \mathcal{P}_x^X(u)$, $\forall i = 1,...,N$ is the identical marginal distribution density for all the random variables $\{X_i\}_{i=1}^N$).

Moreover, we define the nonnegative quantities ℓ_p^X, determined by the integral operator $\hat{I}_p^X$, given by

$$\hat{I}_p^X \equiv Max\left\{ \int \cdot \cdot \int_{\{x_j \in D_x\}_{j=1}^N} \left| \mu_{p,N}^X(\{x_j\}_{j=1}^N; p) - x_i \right|^{p-1} \hat{1} \, dx_1...dx_N, \right.$$
$$\left. \int_{x_i \in D_x} \left| x_i - \mu_p^X \right|^{p-1} \hat{1} \, dx_i \right\}, \tag{32}$$

which acts on the probability densities $\mathcal{P}^X(\{x_j\}_{j=1}^N)$ so that

$$\ell_p^X = \hat{I}_p^X[\mathcal{P}^X(\{x_j\}_{j=1}^N)] . \tag{33}$$

Now consider the subset $\mathcal{M} \subseteq \Re$ of all the possible values of ℓ_p^X. Equation (30) yields $1 \le \ell_p^X$, so that the infimum $\mathcal{I}(\mathcal{M}) = 1$ (Note that, in this case, the infimum is element of $\mathcal{M}$, obtained for $p = 1$). On the other hand, Equation (31) yields $0 \le \ell_p^X$, which reads that the nonnegative quantities ℓ_p^X can be arbitrary small, even zero, so that the infimum is now given by $\mathcal{I}(\mathcal{M}) = 0$.

However, the infimum is unique. The false by contradiction comes from the statement $D(p,N) \ne 0$. Hence, $D(p,N) = 0$, and Equation (20) yields $\langle g \rangle_p = 0$, i.e., $\langle \hat{\mu}_{p,N}(\{y_j\}_{j=1}^N; p) - \mu_p \rangle_p = 0$, or $\langle \hat{\mu}_{p,N}(\{y_j\}_{j=1}^N; p)) \rangle_p = \mu_p = \langle y_i \rangle_p, \forall i = 1,...,N$.

4. Limit of the $\mathcal{L}^p$-Mean Estimator

The following theorem derives the limit of $\mathcal{L}^p$-mean estimator $\hat{\mu}_{p,N}(\{y_j\}_{j=1}^N; p)$.

Theorem 3. *Let the sampling $\{y_i\}_{i=1}^N$, $y_i \in D_y \subseteq \Re, \forall i = 1,...,N$, 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, $\langle \hat{\mu}_{p,N} \rangle_p = \hat{E}_p[\hat{\mu}_{p,N}(\{y_j\}_{j=1}^N; p)] \equiv \mu_p$, as $N \to \infty$, namely, $\lim_{N \to \infty} \hat{\mu}_{p,N} = \langle \hat{\mu}_{p,N} \rangle_p \equiv \mu_p$.*

Notes:

1. Obviously, the independent and identically distributed random variables are also symmetrically distributed. Thus, according to *Theorem 2*, we have $\langle \hat{\mu}_{p,N} \rangle_p = \hat{E}_p[\hat{\mu}_{p,N}(\{y_j\}_{j=1}^N; p)] = \langle Y_i \rangle_p = \hat{E}_p(Y_i) = \mu_p$.

2.　The $\mathcal{L}^p$ expectation value $\langle Y_i \rangle_p = \mu_p$ should be calculated given the marginal distribution density $\mathcal{P}_y(y_i)$, $\forall\, i = 1, \ldots, N$. However, the expression of this distribution is usually unknown, and thus, we estimate μ_p by means of $\lim_{N \to \infty} \widehat{\mu}_{p,N}$.

Proof of Theorem 3. We construct the set of auxiliary random variables $\{X_i\}_{i=1}^N$, defined by $X_i = f_x(Y_i) \equiv |Y_i - \mu_p|^{p-1} sign(Y_i - \mu_p)$, and having the relevant sampling values $\{x_i = f_x(y_i)\}_{i=1}^N$, with domain $x_i \in D_x \equiv \{-|f_x(y_{Min})| \leq x \leq |f_x(y_{Max})|\} \subseteq \Re$, $\forall\, i = 1, \ldots, N$. Apparently, $\{X_i\}_{i=1}^N$ are also independent and identically distributed random variables, and let $\mathcal{P}_{x_i}^X(u) = \mathcal{P}_x^X(u)$, $\forall\, i = 1, \ldots, N$ be the identical marginal distribution density for all the random variables $\{X_i\}_{i=1}^N$. □

Then, the Euclidean expectation value of each of the random variables is

$$\langle x_i \rangle_2 = \int_{x_i \in D_x} x_i\, \mathcal{P}_x(x_i)\, dx_i = \int_{y_i \in D_y} |y_i - \mu_p|^{p-1} sign(y_i - \mu_p)\, \mathcal{P}_y(y_i)\, dy_i = 0\,, \tag{34}$$

$\forall\, i = 1, \ldots, N$. Thereafter, from the "Law of Large Numbers" we have that $\sum_{i=1}^N (x_i - \langle x_i \rangle_2)$ converges to zero, as $N \to \infty$ [12–14]. Thus,

$$\sum_{i=1}^\infty |y_i - \mu_p|^{p-1} sign(y_i - \mu_p) = 0\,. \tag{35}$$

On the other hand, Equation (3), for $N \to \infty$ (assuming the convergence of the sum at this limit), is written as

$$\sum_{i=1}^\infty \left| y_i - \lim_{N \to \infty} \widehat{\mu}_{p,N} \right|^{p-1} sign\left(y_i - \lim_{N \to \infty} \widehat{\mu}_{p,N} \right) = 0\,, \tag{36}$$

while, given the uniqueness of μ_p, we conclude in $\lim_{N \to \infty} \widehat{\mu}_{p,N} = \mu_p$.

As an application, we examine the probability distribution $\mathcal{P}(x)$, constructed by the superposition of two different normal distributions $\mathcal{N}_1(y; \mu = 1, \sigma = 1)$ and $\mathcal{N}_2(y; \mu = 1 + \delta a, \sigma = 2)$ with $\delta a > 0$; namely, $\mathcal{P}(y; \delta a; \lambda) = [\mathcal{N}_1(y) + \lambda \mathcal{N}_2(y; \delta a)]/(1 + \lambda)$. For $\delta a = 0$, the constructed probability distribution is symmetric, and as explained in Section 3 and [3], the whole set of $\mathcal{L}^p$-means degenerates to one single value—in our case $\mu_p = 1$, $\forall\, p > 1$. First, we derive the $\mathcal{L}^p$-mean of the distribution, $\langle y \rangle_p = \mu_p$, where $\int_{-\infty}^{+\infty} |y - \mu_p|^{p-1} sign(y - \mu_p)\, \mathcal{P}(y)\, dy = 0$. Then, we compute the $\mathcal{L}^p$-mean estimator $\widehat{\mu}_{p,N}(\{y_j\}_{j=1}^N; p)$, where the y-values, $\{y_j\}_{j=1}^N$, follow the probability distribution constructed above, $\mathcal{P}(y; \delta a; \lambda)$. The N y-values are generated as follows: we derive the cumulative distribution of $\mathcal{P}$, that is $\mathcal{F}(y; \delta a; \lambda) = \frac{1}{2} + \left[erf\left(\frac{y-1}{\sqrt{2}} \right) + \lambda \cdot erf\left(\frac{y-1-\delta a}{2\sqrt{2}} \right) \right] / [2(1 + \lambda)]$, and then we set the parametrization $\mathcal{F}(y_i; \delta a; \lambda) = i/N$, from where we solve for y_i. The estimator $\widehat{\mu}_{p,N}(\{y_j\}_{j=1}^N; p)$ is implicitly given by $\sum_{i=1}^N |y_i - \widehat{\mu}_{p,N}|^{p-1} sign\left(y_i - \widehat{\mu}_{p,N} \right) = 0$; then, we demonstrate *Theorem 3* showing the equality $\lim_{N \to \infty} \widehat{\mu}_{p,N} = \mu_p$; in particular, we construct the sum $S(N) \equiv \sum_{i=1}^N |y_i - \mu_p|^{p-1} sign\left(y_i - \mu_p \right)$, and show that $lim_{N \to \infty} S(N) \to 0$, satisfying Equation (35).

Figure 3 illustrates the convergence $\lim_{N \to \infty} \widehat{\mu}_{p,N} = \mu_p$. Panel (a) plots the sum $S(N)$ as a function of N and for the norms $p = 1.5, 2$, and 2.5; we find a convergence rate of $S(N) \sim 1/N$. Panel (b) plots the summation $S(N)$ for large N $(= 10^3)$, which becomes zero only if the p-norm used by the estimator $\widehat{\mu}_{p,N}$, that is $p = p_1$, equals the p-norm used by the mean μ_p, that is $p = p_2$. Finally, panel (c) plots the value of the estimator $\widehat{\mu}_{p,N}$, as a function of the norm p, for two data numbers $N = 100$ and $N = 300$, showing the convergence to μ_p, which is co-plotted as a function of p.

Figure 4 plots the deviation between the $\mathcal{L}^p$-mean estimator $\widehat{\mu}_{p,N}$ and the mean μ_p, that is, $|\widehat{\mu}_{p,N} - \mu_p|$, and for a large number of data $(N = 10^3)$. The mean μ_p is taken for the norm $p = 3$, while the deviation is plotted as a function of the norm p of the estimator. We observe that the deviation is minimized, tending to zero, when the norm is $p \cong 3$. However, this result holds if the distribution is not symmetric. Once the parameter δa decreases approaching zero, the distribution $\mathcal{P}(y; \delta a \to 0; \lambda)$

becomes symmetric and the deviation $|\widehat{\mu}_{p,N} - \mu_p|$ obtains small values (while its minimization at a certain p loses its meaning). We observe that for $\delta a \approx 0.01$ or smaller, the deviation is small enough—of the order of $10^{-4} - 10^{-3}$ (it is non-zero because of the computation errors caused by the finite N), so that $\widehat{\mu}_{p,N} \approx \mu_p$.
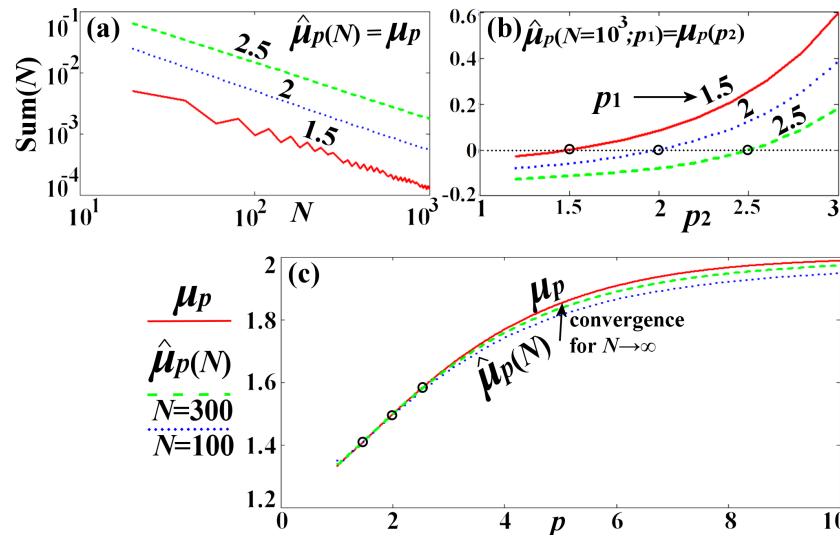

Figure 3. Convergence of the estimator $\widehat{\mu}_{p,N}$ to the mean μ_p (ensemble average). (**a**) The summation $S(N)$ plotted against N for $p = 1.5$ (red solid), 2 (blue dot), and 2.5 (green dash); (**b**) The summation $S(N)$ plotted against $p = p_2$, for $p_1 = 1.5$ (red solid), 2 (blue dot), and 2.5 (green dash), where $S(N) = 0$ holds only for $p_2 = p_1$; (**c**) Estimator $\lim_{N \to \infty} \widehat{\mu}_{p,N}$, plotted against the norm p, for $N = 100$ and $N = 300$, showing the convergence to the co-plotted mean μ_p.
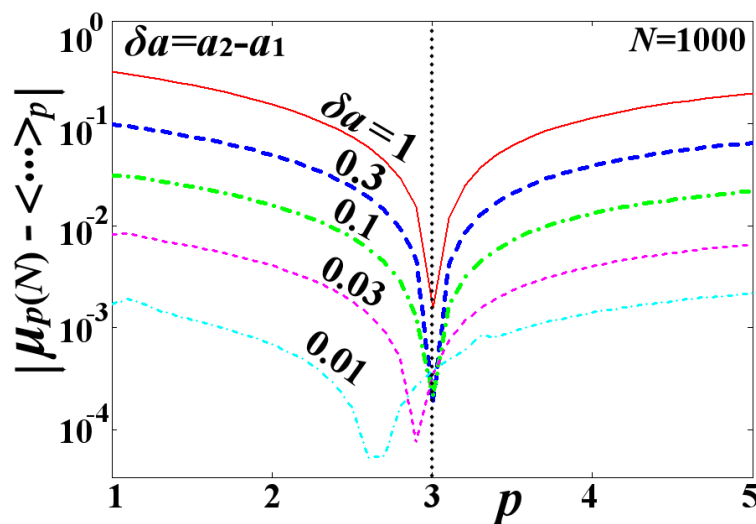

Figure 4. Deviation $|\widehat{\mu}_{p,N} - \mu_p|$ is plotted as a function of the p-norm of the estimator, and for various values of the parameter $\delta a = 1$ (red solid), 0.3 (blue thick dash), 0.1 (green thick dash–dot), 0.03 (purple dash), and 0.01 (light-blue dash–dot).

5. Conclusions

The Euclidean $\mathcal{L}^2$ means are derived by minimizing the sum of the total square deviations, i.e., the Euclidean variance. In a similar way, the non-Euclidean $\mathcal{L}^p$ means were developed by minimizing the sum of the $\mathcal{L}^p$ deviations, that is proportional to the $\mathcal{L}^p$ variance [3]. 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. The least square method based on the Euclidean norm, $p = 2$, and the least absolute deviations method based on the "Taxicab" norm, $p = 1$, are some cases of the general fitting methods based on the $\mathcal{L}^p$ -norms (e.g., [15]; for more applications on the fitting methods based on Lp norms, see: [2,4,16,17]; several other applications can be in signal processing optimization and block entropy analysis, e.g., [2]; in image processing, e.g., [18]; in general data analysis, e.g., [5]; in statistical mechanics, e.g., [3,8,19]. The Law of Large Numbers is a theorem that guarantees the stability of long-term averages of random events, but is valid only for metrics induced by the Euclidean $\mathcal{L}^2$ norm. The importance of this paper is in extending this theorem for $\mathcal{L}^p$-norms. Other interesting applications will be to establish a central limit theorem applied for the $\mathcal{L}^p$-means.

Conflicts of Interest: The author declare no conflict of interest.

References

1. Livadiotis, G. Approach to general methods for fitting and their sensitivity. *Physica A* **2007**, *375*, 518–536.
2. Livadiotis, G. Approach to the block entropy modeling and optimization. *Physica A* **2008**, *387*, 2471–2494.
3. Livadiotis, G. Expectation value and variance based on $\mathcal{L}^p$ norms. *Entropy* **2012**, *14*, 2375.
4. Livadiotis, G.; Moussas, X. The sunspot as an autonomous dynamical system: A model for the growth and decay phases of sunspots. *J. Stat. Distrib. Appl.* **2014**, *379*, 436–458.
5. Livadiotis, G. Chi-p distribution: Characterization of the goodness of the fitting using $\mathcal{L}^p$ norms. *Physica A* **2007**, *1*, 4.
6. Huber, P. *Robust Statistics*; John Wiley Sons: New York, NY, USA, 1981.
7. Fréchet, M. Les éléments aléatoires de nature quelconque dans un espace distancié. *Ann. L'Institut Henri Poincaré* **1948**, *10*, 215–310. (In French)
8. Livadiotis, G. Non-Euclidean-Normed Statistical Mechanics. *Physica A* **2016**, *445*, 240–255.
9. Scarpello, G.M.; Ritelli, D.E. A historical outline of the theorem of implicit functions. *Divulg. Mat.* **2002**, *10*, 171–180.
10. Ahmad, R. On the Structure and Application of Restricted Exchangeability. In *Exchangeability in Probability and Statistics*; Koch, G., Spizzichino, F., Eds.; Elsevier: Amsterdam, The Netherlands, 1982; pp. 157–164.
11. Williams, L.R.; Wells, J.H. $\mathcal{L}^p$ inequalities. *J. Math. Anal. Appl.* **1978**, *64*, 518.
12. Feller, W. *Law of Large Numbers for Identically Distributed Variables*; Wiley: New York, NY, USA, 1971.
13. Hu, T.C.; Chang, H.C. Complete convergence and the law of large numbers for arrays of random elements. *Nonlinear Anal.* **1997**, *30*, 4257–4266.
14. Hoffmann-Jørgensen, J.; Su, K.-L.; Taylor, R.L. The Law of Large Numbers and the Ito-Nisio Theorem for Vector Valued Random Fields. *J. Theor. Probab.* **1997**, *10*, 145–183.
15. Burden, R.L.; Faires, J.D. *Numerical Analysis*; PWS Publishing Company: Boston, MA, USA, 1993; pp. 437–438.
16. Sengupta, A. A rational function approximation of the singular eigenfunction of the monoenergetic neutron transport equation. *J. Phys. A* **1984**, *17*, 2743–2758.
17. Livadiotis, G.; McComas, D.J. Fitting method based on correlation maximization: Applications in Astrophysics. *J. Geophys. Res.* **2013**, *118*, 2863–2875.
18. Sharma, M.; Batra, A. Analysis of distance measures in content based image retrieval. *Glob. J. Comput. Sci. Technol. G Interdiscip.* **2014**, *14*, 11.
19. Livadiotis, G. *Kappa Distributions: Theory and Applications in Plasmas*; Elsevier: Amsterdam, The Netherlands; London, UK; New York, NY, USA, 2017.