axioms

MDPI

Article

Probability Distributions Approximation via Fractional Moments and Maximum Entropy: Theoretical and Computational Aspects

Pier Luigi Novi Inverardi *,† **and Aldo Tagliani** †

Department of Economics & Management, University of Trento, I-38122 Trento, Italy; aldo.tagliani@unitn.it
* Correspondence: pierluigi.noviinverardi@unitn.it
† These authors contributed equally to this work.

Abstract: In the literature, the use of fractional moments to express the available information in the framework of maximum entropy (MaxEnt) approximation of a distribution F having finite or unbounded positive support, has been essentially considered as a computational tool to improve the performance of the analogous procedure based on integer moments. No attention has been paid to two formal aspects concerning fractional moments, such as conditions for the existence of the maximum entropy approximation based on them or convergence in entropy of this approximation to F. This paper aims to fill this gap by providing proofs of these two fundamental results. In fact, convergence in entropy can be involved in the optimal selection of the order of fractional moments for accelerating the convergence of the MaxEnt approximation to F, to clarify the entailment relationships of this type of convergence with other types of convergence useful in statistical applications, and to preserve some important prior features of the underlying F distribution.

Keywords: entropy; convergence in entropy; integer moments; fractional moments; Tchebycheff-systems

MSC: 62E17; 62G07; 62B10; 94A17

check for updates

Citation: Novi Inverardi, P.L.; Tagliani, A. Probability Distributions Approximation via Fractional Moments and Maximum Entropy: Theoretical and Computational Aspects. *Axioms* **2024**, *13*, 28. https://doi.org/10.3390/axioms13010028

Academic Editor: Jong-Min Kim

Received: 16 November 2023
Revised: 20 December 2023
Accepted: 23 December 2023
Published: 30 December 2023

Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

In statistical estimation, one often wants to guess an unknown probability distribution F, given certain observations based on it. There are generally infinitely many distributions consistent with the available data, and the question of which of these to select is an important one in many fields. The notion of entropy has been proposed as a remarkable tool for performing this choice. More precisely, the principle of maximum entropy was established by [1,2] as a tool for inference under uncertainty and consists of finding the most suitable probability distribution under the available information. As Jaynes [1] expressed it, the resulting MaxEnt distribution "*. . . is the least biased estimate possible on the given information*". In summary, the MaxEnt method dictates what are the most "*reasonable and objective*" distribution *subject to given constraints* expressing the available information concerning the data generating mechanism: the analytical form of that constraints are chosen to look at the features of the distribution that we want to preserve in the MaxEnt approximation to guarantee its capability of modeling specific features of F.

It is a common choice to express the constraints in terms of expectations of some functions g_j of X, i.e.,

$$\boldsymbol{E}[g_j(X)] = \int_{\mathcal{U}} g_j(x) f(x) dx = c_j, \quad j = 1, 2, \ldots, n \tag{1}$$

and the resulting maximum entropy distribution (better, its density) emerges by maximizing the Shannon (differential) entropy

$$h_f = -\int_{\mathcal{U}} f(x) \ln f(x) dx \tag{2}$$

under a set of constraints (1) using calculus of variations and Lagrange's multipliers method. The general solution, assuming an arbitrary set of $n+1$ constraints g_j, is given by [1]

$$f_n(x) = \exp\left\{-\lambda_0 - \sum_{j=1}^{n} \lambda_j g_j(x)\right\} \tag{3}$$

where $\lambda_1, \ldots, \lambda_n$, are the Lagrange multipliers linked to the set of adopted constraints (1) while the multiplier λ_0 guarantees the legitimacy of the distribution and $\mathbb{E}(g_j(X))$, $j = 1, 2, \ldots, n$ are the *characterizing moments* of the distribution. It becomes clear that the resulting maximum entropy distribution having density (3) is uniquely driven by the choice of the imposed constraints. This implies that this choice is the most important and determinative part of the MaxEnt method. In the end, the form of the MaxEnt approximation (3) of f is problem-dependent that is, its analytical form depends on the choice of the constraints g_j describing the features of the distribution F that must be preserved in the approximation process.

As we said above, in constructing a density with the MaxEnt methodology, for practical purposes only partial information can be used. However, this does not preclude that we must have at our disposal some physical knowledge of the underlying problem that amounts to

1. the solution to the problem is unique.
2. The entire moment curve from which to pick up a finite number of arbitrary fractional moments is known.

The aim of the paper is focused on the opportunity offered by expressing the system of constraints (1) by fractional moments. This choice gives back a great flexibility to model a wide class of problems where the traditional integer moment constraints may be inefficient in recovering the available information. The theorem of entropy convergence stated for fractional moments setup and related other modes of convergence, offer a formal basis for the optimal choice of number and order of the fractional moments to be involved in the MaxEnt approximation procedures. In the paper we will stress the fact that theoretical and numerical aspects are inextricably linked to each other, justifying why both theoretical and computational aspects must be treated simultaneously in the paper. More precisely, in Section 2 some properties of fractional moments motivating their use in MaxEnt reconstruction of distribution will be discussed, in Section 3 some basics about Tchebycheff systems (T-systems, for brevity) will be recalled and using this tool, the existence and convergence in entropy of the MaxEnt distribution constrained by fractional moments will be proven in Section 4.1 and in Section 4.2, respectively. Finally, in Section 5 two crucial results concerning the optimal choice of the orders α's and the optimal number n of the fractional moments both based on the convergence in entropy of the MaxEnt distribution will be presented.

2. The Role of Fractional Moments in MaxEnt Setup

Constraints (1) expressed by integer moments play an important role in the inverse Hausdorff and Stieltjes classical moment problem that consists in determining an unknown probability mass or density function f corresponding to the distribution F from the knowledge of the sequence of its integer moments $m_j = \mathbb{E}(X^j) = \int_{\mathcal{U}} x^j f(x) dx$, $j \geq 1$, $m_0 = 1$, where $\mathcal{U} = [0, 1]$ in Hausdorff case or $\mathcal{U} = [0, \infty)$ in Stieltjes case.

For practical purposes, if only n prefixed moments are taken into account to express the available information about the distribution, then many different (even an infinity!)

probability distributions could be compatible with that information and non-uniqueness of the distribution recovered from them follows immediately. Hence, the question: *What probability distribution is the best and with respect to what criterion?* The answer follows naturally from Jaynes' principle [1]: from the set of all probability distributions compatible with the n prefixed moments, choose the one that maximizes Shannon's entropy that is, the so-called MaxEnt distribution.

The MaxEnt approximation of f constrained by first $n+1$ integer moments that is, $\boldsymbol{E}[g_j(X)] = \boldsymbol{E}(X^j) = m_j$, $j = 1, \ldots, n$ where n is arbitrary large and $m_0 = 1$ being any density a normalized function, comes immediately from (3) that is,

$$f_n(x) = \exp\left\{-\lambda_0 - \sum_{j=1}^{n} \lambda_j x^j\right\} \tag{4}$$

where λ_j, $j = 1, \ldots, n$, are the Lagrange multipliers linked to the set of adopted constraints while the multiplier λ_0 guarantees the legitimacy of the distribution. Widely known references are the books [3,4] and more recently [5]. These sources contain comprehensive details about a series of remarkable results paving the progress in moment problems for more than a century. Theoretical and computational aspects are inextricably linked to each other.

It is a well-known fact that in a determinate moment problem, the sequence of integer moments $\{m_j\}_{j=0}^{\infty}$ carries all the information concerning the distribution F; hence it may happen that the moments of high order also contain a considerable amount of it as in the case of asymmetric or heavy-tailed distributions. But it is also well known that the moment problem becomes ill-conditioned when the number (hence, the order) n of moments increases and to avoid numerical instability due to the ill-conditioning of Hankel matrices only the moments of small order n are involved. Neglecting higher-order moments implies losing the information carried by them with consequences on the quality of f_n as an approximation of f. Furthermore, if the first few moments are not informative with respect to the distribution, the situation is even worse with f_n being a definitely bad approximation of f.

For a practical example, Ref. [6] discuss the role played by the constraints choice in modeling probability distributions via the MaxEnt method for complex geophysical processes concluding that the usual choice based on integer moments in virtue of their physical meaning cannot be able, both for theoretical and empirical, i.e., data-driven, reasons to describe the relevant geophysical features that have to be preserved because characterizing the distribution (for more details, see the mentioned paper pp. 52–53).

At this point, a fundamental question is how to reformulate the MaxEnt solution of the moment problem in a suitable way to permit a reliable and efficient approximation through f_n of the target density f. Or, equivalently, how to choose the (optimal) analytical form of the set of the MaxEnt constraints?

To try to find an answer to this crucial question, combining (2) and (3), the entropy of f_n is given by

$$h_{f_n} = \sum_{j=0}^{n} \lambda_j m_j \geq h_f. \tag{5}$$

and point out that the quantity

$$h_{f_n} - h_f \geq 0 \tag{6}$$

is a measure of residual uncertainty about the distribution of the random variable X associated with the MaxEnt approximation f_n of f and corresponds to the maximum residual entropy-or equivalently the minimum information gain-associated to the knowledge of the first n moments of F not captured by (4).

When integer moments are used as constraints in a MaxEnt procedure, due to the mechanical choice of integer moments, the rate of reduction of the residual entropy (6) is very slow and becomes more and more negligible as the number n of moments increases. Conse-

quently, it is urgent to look for a class of alternative approximants with faster convergence of h_{f_n} to h_f: the class of MaxEnt distributions constrained by fractional moments

$$m_{\alpha_j} = \boldsymbol{E}(X^{\alpha_j}) = \int_{\mathcal{U}} x^{\alpha_j} f(x) dx, \;\; \alpha_j \in \mathbb{R}^+ \tag{7}$$

represents a natural alternative to (4) where a proper choice of number n and α exponents built on the convergence in entropy theorem of Section 4.2, allows us to control the residual uncertainty reduction and at the end, to accelerate the rate of the convergence of f_n to f, mitigating the effects of ill-conditioning due to the large value of n used. Indeed, since any fractional moment m_α can be obtained as a function of many (as computationally feasible) integer moments m_j ([7,8] for details), the information available in the sequence of integer moments can be squeezed into a few fractional moments: for example, in case of heavy-tailed distributions where integer moments of high order are required, the use of a few fractional moments permit to avoid (or mitigate) ill-conditioning without losing the tail information which is crucial to model, for example, the risk of extreme events or their predictability in a computationally tractable environment. Refs. [9,10] give an application of the fractional moment method in the structural reliability analysis which is typically based on a model that describes the response, such as maximum deformation or stress, as a function of several random variables and they base the derivation of that model on the MaxEnt principle where constraints are specified in terms of the fractional moments, in place of commonly used integer moments to avoid the well known ill-conditioning problem, studying the numerical accuracy and efficiency of the proposed method. From the life-cycle perspective, Ref. [11] observes that probabilistic lifetime modeling of an engineering system provides important information for risk assessment of the system by evaluating the mean-time to failure, survival probability, dynamic hazard rate, and among others. This can be conducted again using fractional moments into the MaxEnt technique to approximate the distribution of interest. Further, in virtue of the flexibility due to the continuous nature of their order α, fractional moments represent a valid alternative when physical principles of momentum are invalid and consequently, the use of integer moments to express the MaxEnt constraints reveals to be improper as it happens in many geophysical processes like daily rainfall distribution [6] or tree diameter distribution modeling [12].

Looking for more practical reasons motivating the use of fractional moments in addition to computational feasibility issues, it is interesting to note that sometimes the available information could be better exploited if the search for its optimal summaries took place on the entire moment curve rather than on a predetermined sequence of equispaced points (i.e., integer moments). For example, a large number of common families of probability distributions largely used in reliability and risk theory such as Gamma, Pareto, Rayleigh, and Lognormal to name a few, belong to the exponential family having logarithmic characterizing moments as shown in Table 1 (recalling that $x^\alpha = e^{\alpha \, \ln(x)}$).

Table 1. Some families of distributions and their characterizing moments.

Families of Distribution	Density	Characterizing Moments
Gamma (γ, β)	$\frac{x^{\gamma-1} \exp\{-x/\beta\}}{\Gamma(\gamma)\,\beta^\gamma} \mathbb{1}_{\mathbb{R}^+}(x)$	$\boldsymbol{E}[X]$, $\boldsymbol{E}[\ln(X)]$
Pareto (γ, k)	$\gamma\, k^\gamma \, x^{-(\gamma-1)} \, \mathbb{1}_{[k,+\infty)}(x)$	$\boldsymbol{E}[\ln(X)]$
Lognormal (μ, σ)	$\frac{\exp\{-[\ln(x)-\mu]^2/2\sigma^2\}}{\sqrt{2\,\pi}\sigma x} \mathbb{1}_{\mathbb{R}^+}(x)$	$\boldsymbol{E}[\ln(X)]$, $\boldsymbol{E}[\ln^2(X)]$
Rayleigh (σ^2)	$\frac{x \exp\{-(x^2/2\sigma^2)\}}{\sigma^2} \mathbb{1}_{\mathbb{R}^+}(x)$	$\boldsymbol{E}[X^2]$, $\boldsymbol{E}[\ln(X)]$

It is possible to show that these families of distributions can be considered MaxEnt distributions with fractional moments as characterizing moments, in the sense that the analytic form of the latter is appropriate to capture the relevant (that is, characterizing) information and features of the corresponding distribution. For example, in the Lognormal

case, the characterizing moments are $\Big(\boldsymbol{E}[\ln(X)], \boldsymbol{E}[\ln^2(X)]\Big)$, respectively. Then, by the known relationships

$$\lim_{\alpha\to 0}\frac{x^\alpha-1}{\alpha}=\ln(x)\quad\text{and}\quad\lim_{\alpha\to 0}\left(\frac{x^\alpha-1}{\alpha}\right)^2=\lim_{\alpha\to 0}\frac{x^{2\,\alpha}-2\,x^\alpha+1}{\alpha^2}=\ln^2(x)$$

it follows that the Lognormal density can be reconsidered as a MaxEnt one having $\{\boldsymbol{E}(X^\alpha), \boldsymbol{E}(X^{2\alpha})\}$, $\alpha\to 0$, as characterizing fractional moments.

The same line of reasoning and related results hold true if we consider a random sample $(X_1, X_2, \ldots, X_N)$ and the associated sample fractional moments

$$\hat{m}_{\alpha_j}=\frac{1}{N}\sum_{i=1}^{N}X_i^{\alpha_j},\ \alpha_j\in\mathbb{R}^+ \tag{8}$$

to summarize the sample information needed for the MaxEnt estimation of the density f [7]. In this setup, the MaxEnt estimate f_n represents a genuine non-parametric estimate of f where the constraints of appropriate number and order expressed by (8) represent the features of the distribution of X that must be preserved.

For assessing the feasibility of the moment problem solution based on fractional moments, three aspects must be now considered: the first one concerns the existence of the MaxEnt approximation based on fractional moments as a tool to express the constraints set, the second one consists in finding a formal proof of the convergence in entropy to f of the sequence of MaxEnt "fractional" approximation (Equation (10) below). The third aspect concerns the choice of the number n and the set of the orders α_j for $j=1,2,\ldots,n$ of the fractional moments m_{α_j}. Convergence in entropy of the MaxEnt approximation (10) f_n to f guarantees the residual uncertainty $h_{f_n}-h_f$ about the distribution of the random variable X associated with the approximation f_n of f is minimal. And, for fixed n, it is natural to base the choice of the fractional order $\alpha_j, j=1,2,\ldots,n$, looking for the α's values that minimize the residual uncertainty (6) in the framework established by two important results due to [13].

More precisely, Lin's Theorems 1 and 2 based on asserting that an analytic function on the right half complex plane is completely determined by its values on a sequence of points having an accumulation point there, guarantee that the fractional moments corresponding to the posed restrictions on the exponents to catch the aspects of the process that must be preserved, still characterize the underlying distribution. Specifically, these theorems are:

Theorem 1 (Lin (1992), Thm. 1)**.** *A positive r.v. X is uniquely characterized by an infinite sequence of positive fractional moments $\{m_{\alpha_j}\}_{j=1}^{\infty}$ with distinct exponents $\alpha_j\in(0,\alpha^*)$, $m_{\alpha^*}<\infty$, for some $\alpha^*>0$.*

and

Theorem 2 (Lin (1992), Thm. 2)**.** *If X is a r.v. assuming values from a bounded interval $[0,1]$ and $\{\alpha_j\}_{j=1}^{\infty}$ an infinite sequence of positive and distinct numbers satisfying*

$$\lim_{j\to\infty}\alpha_j=0\quad\textit{and}\quad\sum_{j=1}^{\infty}\alpha_j=+\infty$$

then the sequence of moments $\{m_{\alpha_j}\}_{j=1}^{\infty}$ characterizes X.

The following sections provide formal proofs and results related to each of the three aspects mentioned. But before proceeding, let us briefly recall an important technical result, which plays a pivotal role in performing the proofs.

3. A Reminder about T-Systems

T-systems represent a technical tool that plays a crucial role in proving both the existence and convergence in entropy of the MaxEnt approximation f_n of f. For this reason, we will revisit briefly their main aspects, considering the two cases $X \in [0,\infty)$ and $X \in [0,1]$, separately. In the sequel, notations and results are borrowed from [14,15] where the T-systems are extensively investigated including general functions $\{u_j(t)\}_{j=0}^n$ on abstract set $\mathcal{E}$.

1. $X \in \mathcal{U} = [0,\infty)$.
The starting point is to consider that the set of continuous linearly independent real-valued functions $\{u_j(t)\}_{j=0}^n$, defined on the interval $\mathcal{U} = [0,\infty)$, constitutes a T-system of order n if any polynomial

$$P(t) = \sum_{j=0}^{n} a_j u_j(t), \quad \text{with} \quad \sum_{j=0}^{n} a_j^2 > 0$$

has no more than n zeros on $[0,\infty)$. Equivalently, it is readily seen that $\{u_j(t)\}_{j=0}^n$ is a T-system if and only if the determinants of order $(n+1)$

$$\det \parallel u_0(t), u_1(t), \ldots, u_n(t) \parallel_0^n =: \begin{vmatrix} u_0(t_0) & u_0(t_1) & \cdots & u_0(t_n) \\ u_1(t_0) & u_1(t_1) & \cdots & u_1(t_n) \\ \vdots & \vdots & & \vdots \\ u_n(t_0) & u_n(t_1) & \cdots & u_n(t_n) \end{vmatrix}$$

are strictly positive for any choice of (distinct) pairs of elements $0 \le t_0 < t_1, \cdots < t_n$ in $[0,\infty)$. According with the above definition the special set $\{u_j(t) = t^{\alpha_j}\}_{j=0}^n$ we are interested in, with distinct $0 = \alpha_0 < \alpha_1, \cdots < \alpha_n$, is a T-system having the properties

(a) $u_j(t) = t^{\alpha_j} > 0$ for each $0 \le j \le n$

(b) $\lim_{t\to\infty} \frac{t^{\alpha_j}}{t^{\alpha_n}} = 0$ for each $j = 0, \ldots, n-1$

(c) if the set $\{u_j(t) = t^{\alpha_j}\}_{j=0}^n$ is a T-system, then $\{u_j(t) = t^{\alpha_j}\}_{j=0}^{n+1}$ is it too.

The space $\mathcal{M}_{n+1}$ of moments, given by the convex hull generated by the points $\{t^{\alpha_j}\}_{j=0}^n$ has a nonempty interior. This set is convex but has a complex geometry. A good deal of the geometry of the classical moment spaces induced by the special T-system $\{1, t, t^2, ..., t^n\}$ can be generalized to the case of the investigated T-system $\{u_j(t) = t^{\alpha_j}\}_{j=0}^n$. If the sequence of prescribed moments $\{m_{\alpha_j}\}_{j=0}^n$ is an inner point of $\mathcal{M}_{n+1}$ then there are uncountably many probability measure $d\sigma(t)$ having such prescribed moments, one of them being $d\sigma(t) = f_n(t)dt$. Elsewhere, if the sequence of prescribed moments $\{m_{\alpha_j}\}_{j=0}^n$ belongs to $\partial\mathcal{M}_{n+1}$, the boundary of $\mathcal{M}_{n+1}$, a unique measure supported on a finite set of points exists (the so-called *lower principal representation*) and the determinant of the below-defined Gram matrix G_n becomes zero. For an arbitrary n, let $0 = \alpha_0 < \alpha_1 < \cdots < \alpha_n$. For notational convenience we set

$$m_{\alpha_i,\alpha_j} = \boldsymbol{E}(X^{\alpha_i} X^{\alpha_j}) = \int_{\mathcal{U}} t^{\alpha_i} t^{\alpha_j} f(t)\, dt = \int_{\mathcal{U}} t^{\alpha_i+\alpha_j} f(t)\, dt.$$

Let us now consider the probability measure $d\sigma(t) = f_n(t)\, dt$. Then $t^{\alpha_j} \in L^2_{d\sigma}(\mathcal{U})$, where, as usual

$$L^2_{d\sigma}\Big(\mathcal{U}\Big) = \{t^{\alpha_j} : \int_{\mathcal{U}} t^{2\alpha_j}\, d\sigma(t) = \int_{\mathcal{U}} t^{2\alpha_j} f_n(t)\, dt < +\infty\}$$

Thus the matrix $G_n = [m_{\alpha_i,\alpha_j}]_{i,j=0}^n$ is the positive definite Gram matrix.

The following Markov-Krein theorem ([14] Thm 5.1, p. 157; [15] Thm 1.1, p. 177) is fundamental to prove the convergence in entropy of the MaxEnt distribution: here we adapt it to fractional moments.

Theorem 3 (Markov–Krein theorem). *Given values of the first fractional moments* $\{m_{\alpha_j}\}_{j=0}^{n}$ $\in Int(\mathcal{M}_{n+1})$ *so that the Gram matrix* G_n *be positive definite, the integral* $\int_{\mathcal{U}} u_{n+1}(t)\, d\sigma(t)$ *over all the distributions* $\sigma(t)$ *having the assigned moments* $\{m_{\alpha_j}\}_{j=0}^{n}$, *has a minimum value* $m_{\alpha_{n+1}}^{-}$ *where*

$$m_{\alpha_{n+1}}^{-} = \int_{\mathcal{U}} u_{n+1}(t)\, d\underline{\sigma}(t) \tag{9}$$

The corresponding measure $\underline{\sigma}$, *under the form of a sum weighted Dirac delta function for which is uniquely determined, is the so-called lower principal representation. Furthermore the point* $\{m_{\alpha_0}, \cdots, m_{\alpha_n}, m_{\alpha_{n+1}}^{-}\}$ *belongs to* $\partial\mathcal{M}_{n+2}$, *the boundary of* $\mathcal{M}_{n+2}$.

2. $X \in \mathcal{U} = [0,1]$.
In this case the procedure proposed for $X \in [0,\infty)$ runs more or less similarly since the involved functions $\{u_j(t) = t^{\alpha_j}\}_{j=0}^{n}$, $t \in [0,1]$ are T-systems too and an analogous Markov-Krein theorem ([14] Thm 1.1, p. 80; [15] Thm. 1.1, p. 109) is available. It should only be recalled that, in analogy with Theorem 3, given $\{m_{\alpha_j}\}_{j=0}^{n} \in Int(\mathcal{M}_{n+1})$, the moment $m_{\alpha_{n+1}}$ admits minimum and maximum value $m_{\alpha_{n+1}}^{-}$, $m_{\alpha_{n+1}}^{+}$, respectively, where

$$\begin{pmatrix} m_{\alpha_{n+1}}^{+} \\ m_{\alpha_{n+1}}^{-} \end{pmatrix} = \int_{\mathcal{U}} u_{n+1}(t) d \begin{pmatrix} \overline{\sigma} \\ \underline{\sigma} \end{pmatrix}$$

Here, the corresponding measures $\underline{\sigma}$ and $\overline{\sigma}$ under the form of a sum-weighted Dirac delta function are uniquely determined and they are the so-called *lower and upper principal representation*, respectively, and the points $\{m_{\alpha_0}, \cdots, m_{\alpha_n}, m_{\alpha_{n+1}}^{\pm}\} \in \partial(\mathcal{M}_{n+2})$.

4. MaxEnt Solution of the Fractional Moment Problem

Once in both cases $X \in [0,1]$ and $X \in [0,\infty)$ the moment curve $m_X(\alpha) = \int_{\mathcal{U}} t^{\alpha} f(t)\, dt$ has been obtained, the probability distribution constrained by fractional moments can be estimated (approximated) through the MaxEnt technique, which is essentially an extension of the commonly used integer moment-based MaxEnt procedure.

Note, that given a finite collection of (population or sample) fractional moments $\{m_{\alpha_j}\}_{j=0}^{n}$, with $\alpha_0 = 0$, the corresponding MaxEnt solution for f is

$$f_n(x) = \exp\left\{ -\sum_{j=0}^{n} \lambda_j x^{\alpha_j} \right\} \tag{10}$$

where the λ_j are such that the (fractional) constraints

$$\int_{\mathcal{U}} x^{\alpha_j} f_n(x)\, dx = m_{\alpha_j}, \qquad j = 0, \ldots, n \tag{11}$$

are satisfied, and f_n depends on the m_{α_j} (thus on the α_j) through the λ_j. Note, that in $[0,\infty)$ case λ_n must take positive values in $[0,\infty)$ to guarantee f_n integrability. The MaxEnt approximation f_n of f has entropy

$$h_{f_n} = \sum_{j=0}^{n} \lambda_j m_{\alpha_j} \tag{12}$$

Here $(\lambda_0, \ldots, \lambda_n)$ is the vector of Lagrange multipliers: if it is possible to determine Lagrange multipliers from the constraints $\{m_{\alpha_j}\}_{j=0}^{n}$, then the moment problem admits solution

and f_n is MaxEnt approximation of f which is unique in $\mathcal{U}$ due to strict concavity of (12). In this setup, two fundamental theoretical questions must be now addressed: the existence of the MaxEnt distribution F_n and its convergence in entropy to F. The last two are crucial to exploit in real-world applications; the MaxEnt technique aims to recover the distribution F by the available information on X here summarized by a proper set of constraints expressed in terms of fractional moments, just to take into account what is discussed in Section 2.

4.1. Existence of MaxEnt Distribution

For the MaxEnt distribution existence, a close and evident analogy between the two cases integer and fractional moments there exist, being both the set of functions $\{u_j(t) = t^{\alpha_j}\}_{j=0}^{n}$ and $\{u_j(t) = t^{j}\}_{j=0}^{n}$ T-systems. The proof simply replaces Hankel matrices with Gram matrices above defined.

1. Suppose that X has unbounded support, $\mathcal{U} = \mathbb{R}^+$, and the first $n+1$ moments $\{m_{\alpha_j}\}_{j=0}^{n} \in Int(\mathcal{M}_{n+1})$ have been assigned, $\lambda_n \geq 0$ has to be to guarantee integrability of f_n. In analogy with the case of integer moments, being both integer and fractional moments T-systems, the above nonnegativity condition on λ_n is crucial and renders the moment problem solvable only under certain restrictive assumptions on the prescribed moment vector $\{m_{\alpha_j}\}_{j=0}^{n}$. Consider (11) with n replaced by $n+1$, the first $n+1$ moments $\{m_{\alpha_j}\}_{j=0}^{n}$ held constant, whilst $m_{\alpha_{n+1}}$ varies continuously, so that the Lagrange multipliers $\lambda_j = \lambda_j(m_{\alpha_{n+1}})$, $j = 0, \ldots, n+1$ are depending on $m_{\alpha_{n+1}}$. Differentiating both sides with respect to $m_{\alpha_{n+1}}$ one has

$$G_{n+1} \cdot \left[\frac{d\lambda_0}{dm_{\alpha_{n+1}}}, \ldots, \frac{d\lambda_{n+1}}{dm_{\alpha_{n+1}}}\right]' = [0, \ldots, 0, -1]' \tag{13}$$

where $'$ denotes the transpose. From G_{n+1} symmetric and positive definite it follows

$$\begin{aligned} 0 < \left[\frac{d\lambda_0}{dm_{\alpha_{n+1}}}, \ldots, \frac{d\lambda_{n+1}}{dm_{\alpha_{n+1}}}\right] \cdot G_{n+1} \cdot \left[\frac{d\lambda_0}{dm_{\alpha_{n+1}}}, \ldots, \frac{d\lambda_{n+1}}{dm_{\alpha_{n+1}}}\right]' = \\ = \left[\frac{d\lambda_0}{dm_{\alpha_{n+1}}}, \ldots, \frac{d\lambda_{n+1}}{dm_{\alpha_{n+1}}}\right] \cdot \left[0, \ldots, 0, -1\right]' = -\frac{d\lambda_{n+1}}{dm_{\alpha_{n+1}}} \end{aligned} \tag{14}$$

Then $\frac{d\lambda_{n+1}}{dm_{\alpha_{n+1}}} < 0$ and λ_{n+1} monotonic decreasing function. MaxEnt machinery leads us to consider a further quantity

$$m^{+}_{\alpha_{n+1}} = \int_{\mathcal{U}} t^{\alpha_{n+1}} f_n(t)\, dt \tag{15}$$

with, in general, $m^{+}_{\alpha_{n+1}} \neq m_{\alpha_{n+1}}$.
From now on, for the sake of brevity, in the arguments of f_{n+1} and $h_{f_{n+1}}$ we will mention only those that take continuously varying values.

(i) Assume f_n exists. Once $\{m_{\alpha_j}\}_{j=0}^{n+1}$ are assigned and $m_{\alpha_{n+1}}$ varies continuously, combine together the following facts: $\lambda_{n+1}(m_{\alpha_{n+1}})$ is a monotonic decreasing function, $f_{n+1}(m^{+}_{\alpha_{n+1}}) = f_n$ and take into account (9) and (15). One concludes that, if f_n exists, the necessary and sufficient condition for the existence of f_{n+1} is $m^{-}_{\alpha_{n+1}} < m_{\alpha_{n+1}} \leq m^{+}_{\alpha_{n+1}}$, in analogy with the past investigated case concerning integer moments ([16], Appendix A).

(ii) Assume f_n does not exist. In such a case $\lambda_{n+1} > 0$. Indeed, if it were $\lambda_{n+1} = 0$ then we would have both $m_{\alpha_{n+1}} = m^{+}_{\alpha_{n+1}}$ and then $f_{n+1} = f_n$, contradicting the fact that f_n does not exist. Consequently, f_{n+1} exists for every set $\{m_{\alpha_j}\}_{j=0}^{n+1} \in Int(\mathcal{M}_{n+2})$. For practical purposes, f_n doesn't exist, both f_{n-1} and f_{n+1} exist. We can state that the problem of the non-existence of the MaxEnt density can be easily bypassed.

Collecting together the items (i) and (ii) we conclude that the existence of f_n is iteratively and numerically determined, starting from f_1 which exists.
Proving the conditions of existence of the MaxEnt distribution we remarked the close analogy between the cases of fractional moments and integer moments. It is reasonable to expect similar analogies to arise also in the case in which an entropy value is to be attributed to the density in the case in which it does not exist so that the sequence of entropies $\{h_{f_n}\}_{n=1}^{\infty}$ is defined for every n. The issue was addressed in ([16], Thm. 1) and taking into account of the laboriousness of the proof, we limit ourselves to illustrating the tools involved and the results obtained.
Some relevant facts need to be collected together. Since MaxEnt density f_n does not exist, both f_{n-1} and f_{n+1} exist with entropies $h_{f_{n-1}}$ and $h_{f_{n+1}}$, respectively. Introduce now the following class of densities all having the same first moments $\{m_{\alpha_j}\}_{j=0}^{n}$

$$\mathcal{C}_n =: \left\{ f \geq 0 \,|\, \int_{\mathcal{U}} x^j f(x)\, dx = m_{\alpha_j},\, j = 0, \ldots, n \right\} \tag{16}$$

In particular, we direct our attention to the density $f_{n+1} = f_{n+1}(m_{\alpha_{n+1}}) \in \mathcal{C}_n$, which thanks to Theorem 4 exists for any value $m_{\alpha_{n+1}} > m^{-}_{\alpha_{n+1}}$. As in integer moments case, f_n may not exist so that h_{f_n} is meaningless ([16], Thm. 1) proved the relationship $\lim_{(m_{n+1})\to\infty} h_{f_{n+1}}(m_{n+1}) = h_{f_{n-1}}$, from which $\sup_{f\in\mathcal{C}_n} h_f = h_{f_{n-1}}$, although the current use of MaxEnt fails (here the last recalled $\mathcal{C}_n$ is the analog of (16) with m_{α_j} replaced by m_j). Since the entropy is non-increasing as n increases, the latter equality enables us to set $h_{f_n} = h_{f_{n-1}}$, filling the gap left by the nonexistence of the density f_n. We reformulate such a result in terms of fractional moments as $\lim_{(m_{\alpha_{n+1}})\to\infty} h_{f_{n+1}}(m_{\alpha_{n+1}}) = h_{f_{n-1}}$, from which $\sup_{f\in\mathcal{C}_n} h_f = h_{f_{n-1}}$. That leads us to conclude, whenever f_n does not exist the missing entropy h_{f_n} is replaced with $h_{f_{n-1}}$, so that the sequence of entropies $\{h_{f_n}\}_{j=1}^{\infty}$ is defined for every n.
We can thus reformulate the conditions of existence according to integer moments by means

Theorem 4. *Once the moment set $\{m_{\alpha_j}\}_{j=0}^{n-1} \in Int(\mathcal{M}_n)$ is prescribed, suppose f_{n-1} exists with its n-th moment $m^{+}_{\alpha_n} = \int_{\mathcal{U}} t^{\alpha_n} f_{n-1}(t)\, dt$.*

(i) If $m_{\alpha_n} \leq m^{+}_{\alpha_n}$, then f_n exists; conversely if $m_{\alpha_n} > m^{+}_{\alpha_n}$ f_n does not exist. Thus the existence of f_n is iteratively (and numerically only) determined from f_{n-1} starting from f_1 which exists.

(ii) If f_n does not exist, both f_{n-1} and f_{n+1} exist for every $m_{\alpha_{n-1}} > m^{-}_{\alpha_{n-1}}$ and $m_{\alpha_{n+1}} > m^{-}_{\alpha_{n+1}}$, respectively. In addition, $h_{f_n} = h_{f_{n-1}}$ can be set.

2. Suppose now that $X \in [0,1]$: the procedure employed in the unbounded support case 1. runs similarly since the involved functions $\{u_j(t) = t^{\alpha_j}\}_{j=0}^{n}$, $t \in [0,1]$ are T-systems too and an analogous Markov-Krein theorem ([15], Thm. 1.1, p. 109) is available. It should only be recalled that, in analogy with the above Theorem 4, given $\{m_{\alpha_j}\}_{j=0}^{n-1} \in Int(\mathcal{M}_n)$, the moment m_{α_n} admits minimum and maximum value $m^{-}_{\alpha_n}$ and $m^{+}_{\alpha_n}$, respectively. The corresponding measures $\sigma = \underline{\sigma}$ and $\sigma = \overline{\sigma}$ under the form of a sum weighted Dirac delta function are uniquely determined and they are the so-called *lower and upper principal representation*, respectively, and the points $\{m_{\alpha_0}, \cdots, m_{\alpha_{n-1}}, m^{\pm}_{\alpha_n}\} \in \partial(\mathcal{M}_{n+1})$. Thanks to MaxEnt formalism Equation (13) continues to hold. Once the first moments $\{m_{\alpha_j}\}_{j=0}^{n} \in Int(\mathcal{M}_{n+1})$ have been assigned, the bounded support does not imply any restriction on the Lagrange multipliers, in particular, λ_n can take on any real value. From (13), as m_{α_n} varies within the bounded range of its admissible values $(m^{-}_{\alpha_n}, m^{+}_{\alpha_n})$, $det(G_{n-1}) > 0$ is bounded. As $m_{\alpha_n} \to m^{\pm}_{\alpha_n}$, f_n coincides with the measures $\underline{\sigma}$ and $\overline{\sigma}$. As a consequence $det(G_n) \to 0$, from which $\frac{d\lambda_{n+1}}{dm_{\alpha_{n+1}}} = -\frac{det(G_{n-1})}{det(G_n)} \to -\infty$ and then $\lambda_{n+1} \to -\infty$ follows.

Analog conclusions hold for the remaining Lagrange multipliers, pre and post-multiplying in $\frac{d\lambda_j}{dm_{\alpha_n}}$, with $j < n$, the matrix at the numerator by a suitable permutation matrix.
In conclusion, given $\{m_{\alpha_j}\}_{j=0}^{n-1} \in Int(\mathcal{M}_n)$ and assuming f_{n-1} exists, f_n exists if $\{m_{\alpha_j}\}_{j=0}^{n} \in Int(\mathcal{M}_{n+1})$. Equivalently, the existence of f_n is iteratively determined, starting from f_0 (the uniform distribution) which exists. On the other hand, thanks again to the MaxEnt formalism the previous proof of existence continues to hold. The solvability of the problem under certain restrictive assumptions on the prescribed moment vector ceases to exist and consequently the following theorem holds:

Theorem 5. *If $X \in [0,1]$ a necessary and sufficient condition for the existence of the MaxEnt distribution f_n is that the vector of moments is internal to the space of moments, that is $\{m_{\alpha_j}\}_{j=0}^{n} \in Int(\mathcal{M}_{n+1})$.*

4.2. Entropy Convergence of MaxEnt Distribution

Convergence in entropy of f_n to f in the case where the entropy (12) is finite or $-\infty$ and its implications, play a fundamental role in many applied problems where the focus is often put on the behavior of the tails of the distribution F that are crucial to study extreme events behavior and to evaluate the probability of their occurrence. In this direction, Ref. [17] stresses the fact that "*...at the tails, the MaxEnt distribution oscillates because of the nonmonotonic nature of the polynomial embedded in the f_n. Thus, only the lower-order moments are typically considered, but in such cases, f_n hardly models tails fatter than the Gaussian. Therefore, the tails of many distributions cannot be well fitted by the MaxEnt distribution with $n \leq 4$*" thus questioning the utility of the MaxEnt approach and consequent solution in this case.

We will prove the almost everywhere nature of the convergence in entropy to f of the MaxEnt approximation f_n based on an optimal set of fractional moments and this will permit us to disprove the above claim "*...MaxEnt distribution oscillates because of the nonmonotonic nature of the polynomial embedded in the f_n*" and state that f_n represents a reliable reconstruction of f and of the main features of the corresponding distribution F, including the tail behavior. Further, exploiting the convergence in entropy of f_n to f, it is possible to formulate a criterion for choosing the optimal number n and the values $\{\alpha_j\}_{j=1}^{n}$ of the fractional exponents and then, the best set of fractional moments $\{m_{\alpha_j}\}_{j=1}^{n}$ (see Equation (29) below).

Finally, even if the tails of the distribution oscillate as stated by [17], if the focus is on evaluating (or estimating) appropriate numerical summaries of the distribution usually expressed in terms of expected values or quantiles, convergence in entropy ensures that the approximation error (Equations (25) and (26) below) can be controlled by a proper choice of the number and the orders of fractional moments and, consequently, the goodness and reliability of such summaries regardless of the oscillating nature of the tails of the MaxEnt distribution.

We are now in a position to prove the main result of this paper, which we enunciate below.

Theorem 6 (Main result)**.** *If X is a positive random variable, having the moments sequence $\{m_{\alpha_j}\}_{j=0}^{\infty}$ characterizing a unique distribution, MaxEnt approximations converge in entropy to the underlying distribution, that is*

$$\lim_{n\to\infty} h_{f_n} = h_f \tag{17}$$

with h_f either finite or $-\infty$.

Proof. We begin giving the proof of Theorem 6 for $X \in [0,\infty)$. Then we just adjust the proof for $X \in [0,1]$.

1. Suppose $X \in [0,\infty)$.

As $m_{\alpha_{n+1}} > m^{-}_{\alpha_{n+1}}$ varies, both $f_{n+1} = f_{n+1}(m_{\alpha_{n+1}})$ (equivalently $\lambda_j = \lambda_j(m_{\alpha_{n+1}})$, $j = 0, \ldots, n+1$) and then $h_{f_{n+1}} = h_{f_{n+1}}(m_{\alpha_{n+1}})$ hold.
Consider $h_{f_{n+1}}(m_{\alpha_{n+1}})$ and collect together (12) and the first equation of (13), we have

$$\frac{dh_{f_{n+1}}(m_{\alpha_{n+1}})}{dm_{\alpha_{n+1}}} = \sum_{j=0}^{n} m_{\alpha_j} \frac{d\lambda_j(m_{\alpha_{n+1}})}{dm_{\alpha_{n+1}}} + \lambda_{n+1}(m_{\alpha_{n+1}}) = \lambda_{n+1}(m_{\alpha_{n+1}})$$

from which, taking into account (14), $\frac{d^2 h_{f_{n+1}}(m_{\alpha_{n+1}})}{dm^2_{\alpha_{n+1}}} = \frac{d\lambda_{n+1}(m_{\alpha_{n+1}})}{dm_{\alpha_{n+1}}} < 0$. Thus $h_{f_{n+1}}(m_{\alpha_{n+1}})$ is a differentiable concave function.
Enter Markov–Krein's Theorem. From Theorem 4 and its consequences, as $m_{\alpha_{n+1}} \to m^{-}_{\alpha_{n+1}}$, $f_{n+1}(m_{\alpha_{n+1}})$ can be assimilated to Dirac's deltas set, equivalently to discrete distribution, the so-called *lower principal representation* $\underline{\sigma}$.
We recall for consistency between the differential entropy of a continuous random variable and the entropy of its discretization, the differential entropy of any discrete measure (being compared to the delta Dirac function) is assumed to be $-\infty$ ([18], pp. 247–249). As a consequence, $h_{f_{n+1}}(m^{-}_{\alpha_{n+1}}) = -\infty$ can be set. On the other hand, as $m_{\alpha_{n+1}}$ takes its own prescribed value, $h_{f_{n+1}} \geq h_f$ holds. Then, with $h_{f_{n+1}}(m_{\alpha_{n+1}})$ being a continuous function, there exists a value, say $\tilde{m}_{\alpha_{n+1}} \in (m^{-}_{\alpha_{n+1}}; m_{\alpha_{n+1}}]$, such that $h_{f_{n+1}}(\tilde{m}_{\alpha_{n+1}}) = h_f$. Summarizing, we have seen that:

(i) If $\{\alpha_j\}_0^{n+1}$ are assigned and f_{n+1} is the corresponding MaxEnt density with entropy $h_{f_{n+1}}$, the sequence $\{h_{f_{n+1}}\}$ is monotonically decreasing and then convergent, with $\lim_{n\to\infty} h_{f_{n+1}} \geq h_f$;
(ii) for each n, $h_{f_{n+1}}(m_{\alpha_{n+1}})$ is concave function in $(m^{-}_{\alpha_{n+1}}; m_{\alpha_{n+1}}]$; as $m_{\alpha_{n+1}} \to m^{-}_{\alpha_{n+1}}$, $h_{f_{n+1}}(m^{-}_{\alpha_{n+1}}) = -\infty$;
(iii) there exists $\tilde{m}_{\alpha_{n+1}} \in (m^{-}_{\alpha_{n+1}}; m_{\alpha_{n+1}}]$ such that $h_{f_{n+1}}(\tilde{m}_{\alpha_{n+1}}) = h_f$.
Enter Lin's Theorem. Consider the Theorem 1 and without loss of generality, it will be assumed the sequence $\{m_{\alpha_j}\}_0^{\infty}$ is asymptotically monotonic increasing. From Theorem 1, the sequence $\{m_{\alpha_j}\}_0^{\infty}$ is convergent and, under the above assumption, is asymptotically monotonic increasing. As $n \to \infty$, from both relationships $m_{\alpha_n} < m^{-}_{\alpha_{n+1}} < \tilde{m}_{\alpha_{n+1}} < m_{\alpha_{n+1}}$ and $(m_{\alpha_{n+1}} - m_{\alpha_n}) \to 0$, it follows

(iv) both $m^{-}_{\alpha_{n+1}} \to m_{\alpha_{n+1}}$ and $\tilde{m}_{\alpha_{n+1}} \to m_{\alpha_{n+1}}$.

Combining together just the above items (i)–(iv) drawn from Theorem 1 and Theorem 4, respectively, it follows

$$\lim_{n\to\infty} h_{f_{n+1}} = \lim_{n\to\infty} h_{f_{n+1}}(\tilde{m}_{\alpha_{n+1}}) = h_f. \tag{18}$$

The employed methodology for the proof clearly suggests that the convergence in entropy holds true in both cases h_f finite and $h_f = -\infty$. Indeed, assuming $h_f = -\infty$, as $n \to \infty$, $m_{\tilde{\alpha}_{n+1}}$ tends to $m_{\alpha_{n+1}}$, so that Equation (18) leads to $\lim_{n\to\infty} h_{f_{n+1}} = -\infty$ too. Previously we proved that whenever f_n does not exist the missing entropy h_{f_n} is replaced with $h_{f_{n-1}}$, so that the sequence of entropies $\{h_{f_n}\}_1^{\infty}$ is defined for every n. That fact gives full significance to (18).

2. Suppose $X \in [0, 1]$.
The procedure previously employed in $X \in [0, \infty)$ case, is likewise extended to $X \in [0, 1]$ since the involved functions $\{u_j(t) = t^{\alpha_j}\}_0^{n+1}$, $t \in [0, 1]$ are T-systems too and both an analogous Markov–Krein theorem ([15]—Thm. 1.1, p. 109) and Lin's Theorem 6 are available (in the latter case, although the sequence $\{m_{\alpha_j}\}_{j=0}^{\infty}$ has to be monotonically decreasing, the proof is similar). Thanks to MaxEnt formalism, both Equation (13) and the used methodology to prove Theorem (6) hold true.

In conclusion, if h_f is finite or $-\infty$, for $X \in [0,1]$ or $X \in [0,\infty)$, MaxEnt formalism enables us to prove the entropy convergence (Theorem (6)) by means of a unified procedure. □

We recall that the entropy convergence had been proved in [19]—Theorem 3.1, for the case with $X \in [0,1]$ and h_f finite, by transforming the problem of Laplace transform inversion into a fractional moment one on $[0,1]$. Here, the author mentions Lin's theorem, although the statements of such theorem are not actually used in the proof at all. Theorem 6 allows us to selecting $(\alpha_1, \ldots, \alpha_n)$ in both cases $X \in [0,1]$ and $X \in [0,\infty)$. This choice is driven by the minimization of the residual $h_{f_n} - h_f$. For this purpose, limited to the case in which f has finite entropy h_f, a valid guide is given by the different modes of convergence stemming from just above proved entropy convergence and shortly recalled.

Thanks to MaxEnt formalism the below-described procedure holds true in both cases $X \in [0,1]$ and $X \in [0,\infty)$, so that with $\mathcal{U}$ we mean, without distinction, the support of $X \in [0,1]$ or $X \in [0,\infty)$.

4.3. Further Convergence Modes for Finite h_f

In the case in which h_f is finite and then $\inf_n h_{f_n}$ is finite too, the following additional results may be drawn. These results configure in a chain of implications starting from the (almost everywhere) convergence in entropy and ending with the convergence in distribution: the aim is to justify the MaxEnt reconstruction of particular features of the distribution in which we are interested in governing their reliability by controlling their approximation error in terms of residual entropy $h_{f_n} - h_f$.

Let $m > n$ and f_m and f_n be the maxentropic solution of the truncated fractional moment problem, with m and n moments, respectively. Combining together the following two facts:

(a) the monotonically non-increasing sequence $\{h_{f_n}\}$ converges to h_f and then it is a Cauchy sequence

(b) the Kullback–Leibler distance between f_m and f_n that share the same first n fractional moments given by

$$D(f_m, f_n) = \int_{\mathcal{U}} f_m \ln \frac{f_m}{f_n} dx \tag{19}$$

implies

$$D(f_m, f_n) = h_{f_n} - h_{f_m}. \tag{20}$$

Hence, taking into account Pinsker's inequality ([20], p. 390), it follows

$$\frac{1}{2} \parallel f_m - f_n \parallel_1^2 \leq D(f_m, f_n) = h_{f_n} - h_{f_m} \tag{21}$$

By replacing f_m with f, letting $n \to \infty$, recalling Theorem 6 and the completeness of the $\mathcal{L}^1$ space, it holds

$$\frac{1}{2} \parallel f_n - f \parallel_1^2 \leq D(f, f_n) = h_{f_n} - h_f \to 0 \tag{22}$$

and hence $\{f_n\}_{n=1}^{\infty}$ has limit f. Then $\{f_n\}_{n=1}^{\infty}$ has a subsequence pointwise convergent a.e. to f and the whole sequence $\{f_n\}_{n=1}^{\infty}$ is also convergent a.e. to the same limit, that is

$$\lim_{n\to\infty} f_n = f \quad a.e. \tag{23}$$

that explains the goodness of the approximation (or estimation, if in a sample setup) f through the MaxEnt f_n based on fractional moments.

Since $\{f_n\}_{n=1}^{\infty}$ converges in $\mathcal{L}^1$-norm to f, then it converges to f also (in probability and) in distribution so that,

$$\lim_{n\to\infty} F_n(x) = F(x)$$

for all x at which $F(x)$ is continuous, where F_n and F denote the cumulative distribution functions corresponding to f_n and f, respectively. Then the approximation f_n is particularly suitable for an accurate calculation of the expected values, since as $n \to \infty$ convergence in distribution is equivalent to

$$\lim_{n\to\infty} \int_{\mathcal{U}} g(x) f_n(x) dx = \int_{\mathcal{U}} g(x) f(x) dx \tag{24}$$

for each bounded function g. Then from (21) and (24) it follows

$$| \boldsymbol{E}_{f_n}(g) - \boldsymbol{E}_f(g) | \leq \| g \|_\infty \sqrt{2(h_{f_n} - h_f)} \tag{25}$$

The argument runs similarly whether quantiles have to be calculated. They may be configured as expected values of proper bounded functions: indeed, for fixed x, $F(x) = \boldsymbol{E}[g(t)]$ with $g(t) = 1$ if $t \in [0, x]$ and $g(t) = 0$ if $t \in \mathcal{U} \setminus [0, x]$. Then we have

$$| F_n(x) - F(x) | \leq \int_0^x | f_n(t) - f(t) | dt \leq \int_{\mathcal{U}} | f_n(t) - f(t) | dt \leq \sqrt{2(h_{f_n} - h_f)}. \tag{26}$$

The above convergence results suggest that a rapid convergence in entropy allows an accurate approximation of the desired density or its features that have to be preserved. So choosing an optimal set of α_j's indices in terms of numbers and values becomes the priority.

Remark 1. *For $X \in [0,1]$ or $X \in [0,\infty)$ with h_f finite, combining Theorem 6 with (21), from $\{h_{f_n}\}_{n=1}^{\infty}$ Cauchy sequence, the continuous functions sequence $\{f_n\}_{n=1}^{\infty}$ is Cauchy sequence too and then uniformly convergent to f. Hence, f is a continuous function. Consequently, an accurate reconstruction of the distribution requires that the underlying density be continuous as well. From an engineering point of view, this request may seem obvious. This explains the reason why in some numerical tests appearing in the literature and concerning the reconstruction of discontinuous densities with entropic techniques using integer or fractional moments, the reconstruction obtained had proven to be somewhat inaccurate.*

5. Optimal Choice and Optimal Number of α's

As recalled in Section 1, in real-world problems the proposal of a stochastic model or a probabilistic law F for a phenomenon X must necessarily take into account the aspects of X that must be preserved and, in some sense, the proposal process is guided by them. As a consequence, although considering the same phenomenon, it could be necessary to compute several distributions, each referring to a specific aspect that has to be preserved and then identify the proposal according to it. Because of their flexible choice, fractional moments can be considered a valuable tool in this regard and for operational reasons, two main questions need now to be addressed: the choice of the number n and of the orders α's of the fractional moments involved in the MaxEnt approximation of F. Both questions have strong relationship with the notion of convergence in entropy of F_n to F which plays a strategic role in finding appropriate answers to these questions.

5.1. The Choice of $(\alpha_1, \ldots, \alpha_n)$

Once the theoretical problem of existence, and convergence in entropy from an arbitrary fractional moment sequence according to Lin's theorems are solved, the approximation of the distribution becomes essentially a computational issue. For this reason, it is a matter of choosing a suitable set of exponents $(\alpha_1, \ldots, \alpha_n)$.

Note, that Lin's theorems provide a theoretical guarantee to the process of reconstructing. However, due to the underlying computational issues, there are infinitely many possibilities in the choice of a few fractional moments. Such a choice must rest on the characteristics of the quantity to be calculated and that are intended to be retained in the approximation process. In an equivalent way, unlike the integer moments, in the approximation of the distribution with fractional moments it is possible to incorporate further

information that comes from the underlying physical problem. The idea that we follow here was originally proposed by [7] and further explored by [21]. The idea goes as follows.

1. We shall denote by f_n found in (10) to make explicit its dependence on n and implicitly on the $(\alpha_1, \ldots, \alpha_n)$. These will be chosen as to minimize the Kullback–Leibler divergence (19) between the "true" but unknown density f and the maxentropic solution f_n. From (12), $h_{f_n} = \sum_{j=0}^{n} \lambda_j m_{\alpha_j}$ and this quantity equals $-\int_{\mathcal{U}} f_n(x) \ln f_n(x)\, dx$ because f and f_n satisfy the same moments constraints. Therefore, minimizing (19) amounts to

$$\arg\min\left\{ \int_{\mathcal{U}} f(x) \ln \frac{f(x)}{f_n(x)} dx \mid \alpha_1, \ldots, \alpha_n \right\} = \arg\min\{h_{f_n} \mid \alpha_1, \ldots, \alpha_n\}. \tag{27}$$

In other words, f_n is obtained through two consecutive minimization procedures with respect to $(\alpha_1, \ldots, \alpha_n, \lambda_1, \ldots, \lambda_n) = (\boldsymbol{\alpha}, \boldsymbol{\lambda})$, namely

$$\min_{\boldsymbol{\alpha}} \min_{\boldsymbol{\lambda}} \left[h_{f_n}(\boldsymbol{\lambda}, \boldsymbol{\alpha}) \right] = \min_{\boldsymbol{\alpha}} \min_{\boldsymbol{\lambda}} \left[\ln\Big(\int_{\mathcal{U}} \exp(-\sum_{j=1}^{n} \lambda_j x^{\alpha_j})\, dx \Big) + \sum_{j=1}^{n} \lambda_j m_{\alpha_j} \right] \tag{28}$$

for $n = 1, 2, \ldots$. This method consists of an implementation of the nested minimization. That is for each fixed $\boldsymbol{\alpha}$, first minimize $\boldsymbol{\lambda} \to h_{f_n}(\boldsymbol{\lambda}, \boldsymbol{\alpha})$ and then carry on the outer minimization with respect to $\boldsymbol{\alpha}$. It is worth mentioning the choice criterion (27) stems from entropy-convergence Theorem 6. The inner minimization is easy because we are dealing with a convex function. But even though the function $\alpha \to m_\alpha = \boldsymbol{E}[X^\alpha]$ is log-convex, the linear combination $\sum_{j=0}^{n} \lambda_j m_{\alpha_j}$ need not be so. However, the existence conditions of a unique solution for (28) remain a theoretical open issue.
Being (28) multivariable and highly nonlinear unconstrained not convex optimization, the uniqueness of the MaxEnt solution may not be guaranteed, so the results greatly rely on the initial condition, i.e., different initial conditions may give different MaxEnt solutions. And even if the algorithm converges, there is no assurance that it will have converged to a global, rather than a local, optimum since conventional algorithms cannot distinguish between the two.
For problems where finding an approximate global optimum is more important than finding a precise local optimum in a fixed amount of time, the Simulated Annealing Method may be preferable to exact algorithms. This explores the function's entire surface and tries to optimize the function while moving both uphill and downhill. Thus, it is largely independent of the starting values, often a critical input in conventional algorithms. Further, it can escape from local optima and go on to find the global optimum.
In conclusion, the crucial issue consists of solving the nested minimization which ranges over two distinct sets of variables $\{\alpha_j\}_1^n$ and $\{\lambda_j\}_1^n$. While each α_j takes its values into the interval $(0, \alpha_{max}]$, where α_{max} relies upon physical or numerical reasons, each λ_j may assume any real value.
2. Alternatively, taking into account for each fixed set $(\alpha_1, \ldots, \alpha_n)$, the inner $\min_{\lambda_1,\ldots,\lambda_n}$ admits a unique solution being $h_{f_n} = \sum_{j=0}^{n} \lambda_j m_{\alpha_j}$ convex function, the outer one could be calculated by Monte Carlo technique, replacing $\min_{\alpha_1,\ldots,\alpha_n}$ with $\inf_{\alpha_1,\ldots,\alpha_n}$, that is

$$\inf_{\boldsymbol{\alpha}} \min_{\boldsymbol{\lambda}} \left[h_{f_n}(\boldsymbol{\lambda}, \boldsymbol{\alpha}) \right] = \inf_{\boldsymbol{\alpha}} \min_{\boldsymbol{\lambda}} \left[\ln\Big(\int_{\mathcal{U}} \exp(-\sum_{j=1}^{n} \lambda_j x^{\alpha_j})\, dx \Big) + \sum_{j=1}^{n} \lambda_j m_{\alpha_j} \right] \tag{29}$$

with $n = 1, 2, \ldots$. Indeed, Equation (29) is just a computational trick and replacing $\min_{\alpha_1,\ldots,\alpha_n}$ with $\inf_{\alpha_1,\ldots,\alpha_n}$ arises from the request for an estimator that guarantees faster convergence in entropy. This replacement does not conflict with the spirit of MaxEnt since, regardless of the estimation criterion (29) of α's, the resulting f_n continues to be a MaxEnt distribution. Then, according to Theorem 6, Equations (25) and (26), expected values or quantiles can be accurately calculated, up to a predetermined tolerance by means of (29). Note, also that the estimation method (29) lends itself easily to taking

into account the existence conditions of f_n stated in Theorem 4: if $(f_n \mid \alpha_1, \ldots, \alpha_{n-1}, \alpha_n)$ does not exist, $(f_{n-1} \mid \alpha_1, \ldots, \alpha_{n-1})$, with the same $(\alpha_1, \ldots, \alpha_{n-1})$ as f_n, does exist. Consequently, $h_{f_{n-1}}$ is recalculated and the value $h_{f_{n-1}}$ is assumed.
After having illustrated the $(\alpha_1, \ldots, \alpha_n)$ selection criteria in the distribution calculation procedure we can reconnect again to the previously introduced problem of choice of constraints in the construction of the MaxEnt distribution. As a constraint, we can also include the choice of range $(0, \alpha_{max}]$ in which to place $(\alpha_1, \ldots, \alpha_n)$ in the minimization procedure (29). As an example, if for physical reasons we know the underlying f has hazard rate function $h(x) = \frac{f(x)}{1-\int_0^x f(u)\,du}$ with prescribed properties (for instance, asymptotically decreasing to zero), the approximation f_n would save such property. As a consequence we choose once more (29), but, as it is easy to verify taking into account (10) with exponents $(\alpha_1, \ldots, \alpha_n) \in [0, 1)$ (therefore, not optimal for the purposes of rapid convergence in entropy). Conversely, with the hazard rate asymptotically increasing to $+\infty$, $(\alpha_1, \ldots, \alpha_n)$ with $\alpha_n > 1$ accomplish that request. Consequently, in both cases, it is important that entropy convergence is ensured.
In conclusion, the criterion (29) for the calculation of f_n is elastic and lends itself to correctly describing multiple scenarios.

5.2. A Single-Loop Strategy for Approximating f_n with $X \in [0, 1]$

In past Section, the difficulties related to nested minimization (28) have been circumvented with the procedure Monte Carlo (29) which allows the computation of $\boldsymbol{\lambda}$ uniquely through the minimization of a convex function. In the case $X \in [0, 1]$, exploiting the different modes of convergence previously proved, it is possible to further simplify the computation of $\boldsymbol{\lambda}$ in (29) by replacing the inner $\min_{\boldsymbol{\lambda}}$ with the solution of a suitable linear system of equations. Indeed, from (3) integrating by parts (see [21], for details), we have

$$\exp(-\sum_{k=0}^{n} \lambda_k) + \sum_{k=1}^{n} \alpha_k \lambda_k \boldsymbol{E}_{f_n}(X^{\alpha_j+\alpha_k}) = (1+\alpha_j)\boldsymbol{E}_{f_n}(X^{\alpha_j}), \quad j = 0, ..., n \tag{30}$$

Subtracting from each equation of index $j = 1, \ldots, n$ the one having index $j = 0$, the following system of equations in the unknowns $\lambda_1, \ldots, \lambda_n$ is obtained

$$\sum_{k=1}^{n} \alpha_k \lambda_k \left[\boldsymbol{E}_{f_n}(X^{\alpha_k+\alpha_{j-1}}) - \boldsymbol{E}_{f_n}(X^{\alpha_k+\alpha_j})\right] = (1+\alpha_{j-1})\boldsymbol{E}_{f_n}(X^{\alpha_{j-1}}) - (1+\alpha_j)\boldsymbol{E}_{f_n}(X^{\alpha_j}) \tag{31}$$

for $j = 1, \ldots, n$ where $\boldsymbol{E}_{f_n}(X^{\alpha_j}) = \boldsymbol{E}_f(X^{\alpha_j})$, $j = 0, \ldots, n$ are known, whilst $\boldsymbol{E}_{f_n}(X^{\alpha_k+\alpha_{j-1}})$ and $\boldsymbol{E}_{f_n}(X^{\alpha_k+\alpha_j})$ are generally unknown. Now observe that taking (25) into account, the moment curves $\boldsymbol{E}_{f_n}(X^{\alpha})$ and $\boldsymbol{E}_f(X^{\alpha})$ corresponding to f_n and f, respectively differ as follows

$$\mid \boldsymbol{E}_{f_n}(X^{\alpha}) - \boldsymbol{E}_f(X^{\alpha}) \mid \leq \sqrt{2(h_{f_n} - h_f)} \tag{32}$$

Note, as well that with $\boldsymbol{\alpha}$ solution of (29), the two moment curves $\boldsymbol{E}_f(X^{\alpha})$ and $\boldsymbol{E}_{f_n}(X^{\alpha})$ interpolate in the Birkhoff–Hermite sense at the nodes (see [22]); that is, they are both interpolating and tangent at the nodes $\boldsymbol{\alpha}$. This implies that

$$\begin{aligned} \boldsymbol{E}_f(X^{\alpha_j}) &= \boldsymbol{E}_{f_n}(X^{\alpha_j}), \;\; j = 0, 1, 2, \ldots, n \\ \boldsymbol{E}_f(X^{\alpha_j} \ln(X)) &= \boldsymbol{E}_{f_n}[X^{\alpha_j} \ln(X)], \;\; j = 1, 2, \ldots, n. \end{aligned} \tag{33}$$

By adopting a guessed choice of $\boldsymbol{\alpha}$ (and the Monte Carlo method may achieve this goal), from Theorem 6, $h_{f_n} = h_f$ follows, so that relying upon (32) and (33), both $\boldsymbol{E}_{f_n}(X^{\alpha_k+\alpha_{j-1}}) = \boldsymbol{E}_f(X^{\alpha_k+\alpha_{j-1}})$ and $\boldsymbol{E}_{f_n}(X^{\alpha_k+\alpha_j}) = \boldsymbol{E}_f(X^{\alpha_k+\alpha_j})$ can be set. Thus, with a guessed choice of $\boldsymbol{\lambda}$, it is legitimate to assimilate (31) to a linear system with unknown $\boldsymbol{\lambda}$ which admits a unique solution, being an identity relating $\boldsymbol{\lambda}$ with a set of values optimally picked up from the moment curve $\boldsymbol{E}_{f_n}(X^{\alpha}) \equiv \boldsymbol{E}_f(X^{\alpha})$.

We shall suppose that the solution to (31) coincides with that obtained by solving (29). This brings this section close to experimental mathematics. The necessary analysis to compute the error in the just above approximation is hard and the verification comes in a posteriori as the numerical results based on it make good sense. Indeed, the numerical evidence suggests that, with $n \simeq 6$ optimally chosen $\boldsymbol{\alpha}$ according with (29), $h_{f_n} \simeq h_f$ is usually observed. From which, combining together (20) with (23) (after replacing f_m with f) the relationship $f = f_n$ a.e. holds. Consequently, from (25) the last two densities have their respective moment curves coincident as well. Then in (31) $\boldsymbol{E}_{f_n}(X^{\alpha_k+\alpha_j}) = \boldsymbol{E}_f(X^{\alpha_k+\alpha_j})$ can be set which in turn enables us to state the solution of (31) coincides with the one obtained from (29). Which makes the ansatz plausible. The ill-conditioning of (31) remains to be investigated. In this regard, unlike integer moments, only a limited number $n \leq 6$ of fractional moments are sufficient for an accurate estimate of f_n, so that ill-conditioning issues are avoided. Once n is fixed, a final consideration concerns the choice of $\alpha_{max} = \max_{1\leq j\leq n}\{\alpha_j\}$. Since the simplified procedure just described essentially concerns the accurate calculation of expected values, the answer follows from (24) and (25): compatibly with numerical issues, α_{max} should be taken as large as possible so as not to raise further constraints on a rapid convergence in entropy. As a consequence, the approximate suggested procedure for replacing (29) is as follows:

1. Once $\boldsymbol{\alpha}$ is fixed, and $\boldsymbol{E}_{f_n}(X^{\alpha_k+\alpha_{j-1}}) = \boldsymbol{E}_f(X^{\alpha_k+\alpha_{j-1}})$, $\boldsymbol{E}_{f_n}(X^{\alpha_k+\alpha_j}) = \boldsymbol{E}_f(X^{\alpha_k+\alpha_j})$ are set, $\boldsymbol{\lambda}$ are drawn solving the linear system (31);
2. as f_n integrates into one, λ_0 is given by

$$\lambda_0 = \ln \int_{\mathcal{U}} \exp\Big(-\sum_{j=1}^{n} \lambda_j x^{\alpha_j}\Big)\, dx \tag{34}$$

3. combining (31) with (34), $h_{f_n} = \lambda_0 + \sum_{j=1}^{n} \lambda_j m_{\alpha_j}$ is calculated and finally

$$f_n^{(app)}: \quad h_{f_n}^{(app)} = \inf_{\alpha_1,\ldots,\alpha_n} h_{f_n} = \lambda_0 + \sum_{j=1}^{n} \lambda_j m_{\alpha_j} \tag{35}$$

In conclusion, the quick simplified approximate procedure permits to avoid the direct solution of (29) by solving the low order linear system (31) with unknown $\boldsymbol{\lambda}$, doing the numerical integration (34) and performing the one-loop procedure (35) that runs on $\boldsymbol{\alpha}$ by means of Monte Carlo technique, with a reduced number of unknowns. This procedure is computationally feasible and convenient and gives back an accurate approximation of f_n (hence of f), as we will see in the Example 1.

Remark 2. *In principle, thanks to MaxEnt formalism, the above outlined procedure in $X \in [0,1]$ might be pairwise extended to $X \in [0,\infty)$. Indeed, in that case the recursive relationship relating Lagrange multipliers $\{\lambda_j\}_{j=1}^n$ with higher order moments is quite similar to (31). With a similar procedure to the case of rv $X \in [0,1]$, integrating by parts (11), the following linear system follows*

$$\sum_{k=1}^{n} \alpha_k \lambda_k \boldsymbol{E}_{f_n}(X^{\alpha_k+\alpha_j}) = (1+\alpha_j)\, \boldsymbol{E}_{f_n}(X^{\alpha_j}), \quad k,j = 1,\ldots,n \tag{36}$$

with $\boldsymbol{E}_{f_n}(X^{\alpha_j}) = \boldsymbol{E}_f(X^{\alpha_j})$, $j = 1,\ldots,n$ known, whilst $\boldsymbol{E}_{f_n}(X^{\alpha_k+\alpha_j})$ are generally unknown and λ_0 given by (34). In analogy with $X \in [0,1]$, MaxEnt density f_n converges in entropy (and then in distribution) to f according to Theorem 6. Then, starting from moderate values of n, $\boldsymbol{E}_{f_n}(X^{\alpha_k+\alpha_j}) = \boldsymbol{E}_f(X^{\alpha_k+\alpha_j})$, for each j,k, can be set, where $\boldsymbol{E}_f(X^{\alpha_k+\alpha_j})$ are known quantities. As a consequence, (36) may considered a linear system admitting a unique solution $(\lambda_1,\ldots,\lambda_n)$, being the involved matrix a nonsingular Gram matrix with distinct $\{\alpha_j\}_{j=1}^n$. However, the method lacks a theoretical ground, since two main issues remain open that is,

1. *how to guarantee $\lambda_n \geq 0$ in (36) to ensure integrability in (34)*

2. *although the convergence in distribution is guaranteed, Equation (24) is not applicable being the function* $g(x) = x^{\alpha}$ *unbounded.*

Nevertheless, the above theoretical drawbacks don't preclude the possibility that, with a special given moment set $\{m_{\alpha_j}\}_{j=1}^{n}$, the method can guarantee accurate results.

Remark 3. *Suppose to consider a generic random sample* $(X_1, X_2, \ldots, X_N)$ *from a distribution having support* $\mathcal{U}$ *not necessarily* $[0,1]$*: for example,* $\mathcal{U} = \mathbb{R}^+$ *or* $\mathcal{U} = \mathbb{R}$*. The simplified procedure for the MaxEnt estimation of the density* f *proposed in Section 4.2 for the case* $X \in [0,1]$ *can be easily applied to the case* $X \in \mathbb{R}^+$ *or the case* $X \in \mathbb{R}$*. Below we will briefly sketch some details.*

1. *Case* $X \in \mathbb{R}$*: it is enough to transform the original sample data for instance through* $Y = g(X) = \frac{1}{2} + \frac{1}{\pi}\arctan(X)$*, to obtain a transformed sample* $Y_1, Y_2, \ldots, Y_N$ *in* $[0,1]$ *and apply the simplified procedure of Section 4.2.*
2. *Case* $X \in \mathbb{R}^+$*: in a similar way, the transformation* $Y = g(X) = e^{-X}$ *can be applied to the original data* $(X_1, X_2, \ldots, X_N)$ *obtaining a transformed sample* $Y_1, Y_2, \ldots, Y_N$ *once again in* $[0,1]$ *interval and the simplified procedure of Section 4.2 is immediately applicable. Note, that the empirical fractional moments of* Y *coincide with the empirical Laplace Transform of* X*, that is* $\frac{1}{N}\sum_1^N Y_j^{\alpha} = \frac{1}{N}\sum_1^N e^{-\alpha X_j}$*, which turns out to be the empirical version of the relationship* $\int_0^1 y^{\alpha} dF_Y(y) = \int_0^{\infty} e^{-\alpha x} dF_X(x)$*. This last relation leads us to conclude that also the numerical inversion of the Laplace transform can be reduced to a fractional moment problem in* $[0,1]$*.*
3. *More recently [23] investigated the feature of an estimator relying upon the fractional moments for random variables supported on* $\mathbb{R}$ *by allowing the fractional powers to take complex numbers. Unlike other authors, they are dealing with the case that the negative values of a random variable are not negligible at all.*

Once obtained the MaxEnt estimate $f_n^{(app)}$ of f in $[0,1]$, it is possible to come back to the original spaces $\mathbb{R}$ or $\mathbb{R}^+$ by the associated inverse transformations $f_n^{(app)}(x) = \mid g'(x) \mid f_n^{(app)}(g(x))$.

The case $X \in \mathbb{R}$ outlined in just above items 1. and 3. partially permits to disprove the criticism arising from [17] where it is asserted that the fractional moments technique is applicable in the case of positive r.v. X only.

5.3. The Choice of n

With just before considerations, as h_f is finite in both (28) and (29), as well in the single loop strategy, arises the optimal choice criterion of n. Recall that by employing integer moments, the adding of a further moment could result in a negligible or even zero entropy decrease, followed by not negligible entropy decreasing with the subsequent moments. Consequently, a stopping criteria based solely on the difference in entropy when adding an additional integer moment could be misleading in choosing the optimal number of moments to use. On the contrary, with the choice of fractional moments according to (29), it is easy to deduce that the sequence $\{h_{f_n}\}_{n \in \mathbb{N}}$ is strictly monotonic decreasing. Indeed, consider (29), fix n and calculate $(\alpha_1, \ldots, \alpha_n)$ from which h_{f_n}. Next, put $n+1$ in (29). As a first step, take the special set $(\alpha_1, \ldots, \alpha_{n+1})$ where the first entries $(\alpha_1, \ldots, \alpha_n)$ coincide with the just above found and $\alpha_{n+1} > \alpha_n$ is kept arbitrarily (that is constrained minimization running on α_{n+1} only, whilst $(\alpha_1, \ldots, \alpha_n)$ is held fixed). Calculate $h_{f_{n+1}}$ and call it $h^*_{f_{n+1}}$, with $h^*_{f_{n+1}} < h_{f_n}$. As a second step take $n+1$ in (29), where the minimum runs on $(\alpha_1, \ldots, \alpha_{n+1})$, from which $h_{f_{n+1}}$ (that is unconstrained minimization). It follows $h_{f_{n+1}} < h^*_{f_{n+1}} < h_{f_n}$. The sequence $\{h_{f_n}\}_{n \in \mathbb{N}}$ is strictly monotonic decreasing and converges to h_f. In the special case where the sequence $\{h_{f_n}\}_{n \in \mathbb{N}}$ is bounded below, has a finite limit, so it is a Cauchy sequence. It leads us to conclude from (29) the rate of entropy decrease becomes smaller and smaller as n increases and the difference between successive entropies becomes reasonably

small (which is up to the modeler to decide), one stops and accepts the density determined by the larger number of moments as the 'true' density.

We conclude the paper with a simple example just to see the fractional moment MaxEnt technique in action. It involves all the crucial theoretical results (fractional moments, entropy reduction, convergence in entropy, optimal choice of number and exponents) introduced in the previous sections of the paper.

Example 1. *Here our goal is to compare the performances of integer and fractional moments MaxEnt approximations of a given density function f in $[0,1]$. For the sake of comparison, we will also consider the approximation $f^{(app)}$ of f obtained by using fractional moments in the simplified MaxEnt procedure given in Section 4.2. Double arithmetic precision is used.*

Consider

$$f(x) = \frac{\pi}{2}\sin(\pi x)\,\mathbb{I}_{[0,1]}(x)$$

with $h_f = -0.14472327456$. The integer moments (im) have the following recursive relationship

$$m_j = \boldsymbol{E}(X^j) = \frac{1}{2} - \frac{j(j-1)}{\pi^2}\,m_{j-2},\ \ j \geq 2,\ m_0 = 1,\ m_1 = \frac{1}{2}$$

whilst fractional moments (fm) $m_{\alpha_j} = \boldsymbol{E}(X^{\alpha_j})$ are explicitly obtained by numerical integration. The associated MaxEnt densities ${f_n}^{(im)}$ and ${f_n}^{(fm)}$ are given by (4) and (10), respectively.

Optimal fractional moments determine a fast entropy decreasing $h_{f_n}^{(fm)} - h_f$ and 4 or 5 of them capture f. But, a definitely higher number of integer moments is required to have a comparable reconstruction of f, incurring drastic numerical instability due to ill-conditioning for $n > 12$: the first column of Table 2 gives evidence of it and actually, the sequence $\{h_{f_n}^{(im)}\}_{n\in\boldsymbol{N}}$ ceases to be a decreasing monotone sequence. The third column of Table 2 contains the residual entropy concerning the MaxEnt approximation $f_n^{(app)}$ of f given by the simplified MaxEnt procedure described in Section 4.2. A quick comparison allows us to conclude the closeness of the solution based on fractional moments and that based again on fractional moments but using the simplified procedure of Section 4.2. As a consequence of the chain of convergence implications descending from the convergence in entropy of $f_n^{(fm)}$ to f, the features of interest of the density f can be well approximated (or estimated, if in a sample setup) by the corresponding features of $f_n^{(fm)}$ or $f_n^{(app)}$ and governing the approximation error in terms of n.

Table 2. Residual entropy with integer (left), fractional (middle) moments and approximated method (right) for an increasing number n of moments.

n	$h_{f_n}^{(im)} - h_f$	n	$h_{f_n}^{(fm)} - h_f$	n	$h_{f_n}^{(app)} - h_f$
2	0.2577×10^{-1}	1	0.8774×10^{-1}	1	0.8893×10^{-1}
4	0.5051×10^{-2}	2	0.8077×10^{-2}	2	0.2802×10^{-2}
6	0.1626×10^{-2}	3	0.4488×10^{-3}	3	0.4851×10^{-3}
8	0.1443×10^{-2}	4	0.6043×10^{-5}	4	0.4693×10^{-4}
10	0.6698×10^{-3}	5	0.4000×10^{-6}	5	0.1196×10^{-4}
12	0.5923×10^{-3}			6	0.14152×10^{-5}

6. Conclusions

The approximation of probability distributions with finite or unbounded positive support using fractional moments to express available information has been reconsidered. Compared to the classical MaxEnt, a novel feature of the proposed method is that the fractional exponent of the MaxEnt distribution is determined through the entropy maximization process, instead of being assigned a priori by an analyst. Theorems of the existence of the maximum entropy distribution and its convergence in entropy are provided. The latter allows us to reconsider the selection criteria of the fractional exponents in order to

speed up the convergence in entropy and other related modes of convergence, as well as preserve some important prior features of the underlying distribution.

Since the release two decades ago of the first paper on the subject by the authors, numerous criticisms have been raised by several researchers who have used this methodology. The main criticism focused on the method of calculating the distribution consisting of nested minimization. There is no need to formalize ourselves on the existence or non-existence of the outer minimum in (28). With $\min_{\alpha_1,\ldots,\alpha_n}$ we would simply intend to carry down the entropy as fast as possible. Indeed, from the different ways of convergence that we have above listed, it can be deduced that the committed error in the calculation of expected values or quantiles is controlled precisely by residual entropy.

Relying on a solid proof of the entropy convergence theorem, different convergence modes are derived as well, we have tried to overcome that shortcoming. For this reason, a Monte Carlo method has been suggested which rests on solid foundations, being the only process of involved minimization performed on a function that is known to be convex.

Concerning the computational efforts of the suggested techniques, the algorithm associated with (29) requires a numerical integration subroutine while the simplified procedure of Section 4.2 needs a numerical routine for the solution of a linear system of equations, both tools available in any mathematical or statistical numerical package and seem to be definitely reasonable.

It is also remarkable to note that the use of a finite number of fractional moments to represent the available information properly tailored to the problem of interest is still possible when only a random sample from a given unknown distribution is available. The existence theorem and the convergence in entropy theorem give solid bases for the inferential procedures about the unknown f based on f_n.

After recalling that fractional moments can be included in the mathematical family of T-systems, they have helped to provide

1. the conditions of existence of the density f_n;
2. the convergence theorem in entropy from which other modes of convergence follow;
3. an optimal choice and optimal number of the fractional exponents $\boldsymbol{\alpha}$;
4. assuming $X \in [0,1]$, a single-loop algorithm for approximating f_n.

Author Contributions: The authors have contributed equally to both the conception and drafting as well as the revision of the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Data are contained within the article.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Jaynes, E.T. Information theory and statistical mechanics. *Phys. Rev.* **1957**, *106*, 620. [CrossRef]
2. Jaynes, E.T. Information theory and statistical mechanics II. *Phys. Rev.* **1957**, *108*, 171. [CrossRef]
3. Akhiezer, N.I. *The Classical Moment Problem and Some Related Questions in Analysis*; Oliver and Boyd: Edinburgh, UK, 1965.
4. Shohat, J.A.; Tamarkin, J.D. *The Problem of Moments*; Mathematical Surveys and Monographs-Volume I; American Mathematical Society: Providence, RI, USA, 1943.
5. Olteanu, O. Symmetry and asymmetry in moment, functional equations and optimization problems. *Symmetry* **2023**, *15*, 1471. [CrossRef]
6. Papalexiou, S.M.; Koutsoyiannis, D. Entropy based derivation of probability distributions: A case study to daily rainfall. *Adv. Water Resour.* **2012**, *45*, 51–57. [CrossRef]
7. Novi Inverardi, P.L.; Tagliani, A. Maximum Entropy Density Estimation from Fractional Moments. *Commun. Stat. Theory Methods* **2003**, *32*, 327–345. [CrossRef]
8. Novi Inverardi, P.L.; Petri, A.; Pontuale, G.; Tagliani, A. Stieltjes moment problem via fractional moments. *Appl. Math. Comput.* **2005**, *166*, 664–677. [CrossRef]
9. Xu, J.; Zhu, S. An efficient approach for high-dimensional structural reliability analysis. *Mech. Syst. Signal Process.* **2019**, *122*, 152–170. [CrossRef]

10. Zhang, X.; Pandey, M.D. Structural reliability analysis based on the concepts of entropy, fractional moment and dimensional reduction method. *Struct. Saf.* **2017**, *43*, 28–40. [CrossRef]
11. Zhang, X.; He, W.; Zhang, Y.; Pandey, M.D. An effective approach for probabilistic lifetime modelling based on the principle of maximum entropy with fractional moments. *Appl. Math. Model.* **2017**, *51*, 626–642. [CrossRef]
12. Ferreira de Lima, A.R.; Ferreira Batista, J.L.; Prado, P.I. Modelling Tree Diameter Distributions in Natural Forests: An Evaluation of 10 Statistical Models. *Forest Sci.* **2015**, *61*, 320–327. [CrossRef]
13. Lin, G.D. Characterizations of Distributions via moments. *Sankhya Indian J. Stat.* **1992**, *54*, 128–132.
14. Karlin, S.; Studden, W.J. *Tchebycheff Systems: With Applications in Analysis and Statistics*; Wiley Interscience: New York, NY, USA, 1966.
15. Krein, M.G.; Nudelman, A.A. *The Markov Moment Problem and Extremal Problems*; American Mathematical Society: Providence, RI, USA, 1977.
16. Novi Inverardi, P.L.; Tagliani, A. Stieltjes and Hamburger Reduced Moment Problem When MaxEnt Solution Does Not Exist. *Mathematics* **2021**, *9*, 309. [CrossRef]
17. Alibrandi, U.; Mosalam, K.M. Kernel density maximum entropy method with generalized moments for evaluating probability distributions, including tails, from a small sample of data. *Int. J. Numer. Methods Eng.* **2017**, *113*, 1904–1928. [CrossRef]
18. Cover, T.M.; Thomas, J.A. *Elements of Information Theory*; John Wiley & Sons: Hoboken, NJ, USA, 2006.
19. Gzyl, H. Super resolution in the maximum entropy approach to invert Laplace transforms. *Inverse Probl. Sci. Eng.* **2017**, *25*, 1536–1545. [CrossRef]
20. Kullback, S. *Information Theory and Statistics*; Dover: New York, NY, USA, 1967.
21. Tagliani, A. Hausdorff moment problem and fractional moments: S simplified procedure. *Appl. Math. Comput.* **2011**, *218*, 4423–4432. [CrossRef]
22. Gzyl, H.; Novi Inverardi, P.L.; Tagliani, A. Fractional moments and maximum entropy: Geometric meaning. *Commun. Stat. Theory Methods* **2014**, *43*, 3596–3601. [CrossRef]
23. Akaoka, Y.; Okamura, K.; Otobe, Y. Properties of complex-valued power means of random variables and their applications. *Acta Math. Acad. Sci. Hung.* **2023**, *171*, 124–175. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.