Article

Stieltjes and Hamburger Reduced Moment Problem When MaxEnt Solution Does Not Exist

Pier Luigi Novi Inverardi * [iD] **and Aldo Tagliani**

Department of Economics and Management, University of Trento, 38122 Trento, Italy; aldo.tagliani@unitn.it
* Correspondence: pierluigi.noviinverardi@unitn.it

Abstract: For a given set of moments whose predetermined values represent the available information, we consider the case where the Maximum Entropy (MaxEnt) solutions for Stieltjes and Hamburger reduced moment problems do not exist. Genuinely relying upon MaxEnt rationale we find the distribution with largest entropy and we prove that this distribution gives the best approximation of the true but unknown underlying distribution. Despite the nice properties just listed, the suggested approximation suffers from some numerical drawbacks and we will discuss this aspect in detail in the paper.

Keywords: probability distribution; Stieltjes and Hamburger reduced moment problem; entropy; maximum entropy; moment space

1. Problem Formulation and MaxEnt Rationale

In the context of testable information that is, when a statement about a probability distribution whose truth or falsity is well-defined, the principle of maximum entropy states that the probability distribution which best represents the current state of knowledge is the one with largest entropy. In this spirit, Maximum Entropy (MaxEnt) methods are traditionally used to select a probability distribution in situations when some (prior) knowledge about the true probability distribution is available and several (up to an infinite set of) different probability distributions are consistent with it. In such a situation MaxEnt methods represent correct methods for doing inference about the true but unknown underlying distribution generating the data that have been observed.

Suppose that X be an absolutely continuous random variable having probability density function (pdf) f defined on an unbounded support $\mathcal{S}_X$ and that $\{\mu_k^*\}_{k=1}^M$, with $\mu_0^* = 1$, be M finite integer moments whose values are pre-determined that is,

$$\mu_k^* = \int_{\mathcal{S}_X} x^k f(x)\,dx, \quad k = 0, \ldots, M, \tag{1}$$

for an arbitrary $M \in \mathbb{N}$. Quantities such as in (1) may be intended to represent the available (pre-determined) information relatively to X.

The Stieltjes (Hamburger) reduced moment problem [1] consists of recovering an unknown pdf f, having support $\mathcal{S}_X = \mathbb{R}^+$ ($\mathcal{S}_X = \mathbb{R}$), from the knowledge of prefixed moment set $\{\mu_k^*\}_{k=1}^M$.

Due to the non-uniqueness of the recovered density, the best choice among the (potentially, infinite) competitors may be done by invoking the Maximum Entropy (MaxEnt) principle [2] which consists in maximizing the Shannon-entropy

$$H_f = -\int_{\mathcal{S}_X} f(x) \ln f(x) dx$$

under the constraints (1). Since entropy may be regarded as an objective measure of the uncertainty in a distribution, "... the MaxEnt distribution is uniquely determined as the

one which is maximally non-committal with regard the missing information" ([2], p. 623) so that "...It agrees with is known but expresses maximum uncertainty with respect to all other matters, and thus leaves a maximum possible freedom for our final decisions to be influenced by the subsequent sample data" ([3], p. 231). In other words, the MaxEnt method dictates the most "reasonable and objective" distribution subject to given constraints.

More formally, in such situation we have to manage a constrained optimization problem involving Shannon entropy and a set of given constraints (here the first M integer moments and the normalization constraint given by (1) when $k = 0$).

This problem is typically solved using the method of Lagrange multipliers leading to a MaxEnt distribution whose density function f_M is given by [4], p. 59

$$f_{M|(\mu_1^*,\dots,\mu_M^*)}(x) = \exp\left(-\sum_{j=0}^{M} \lambda_j x^j\right),\tag{2}$$

fulfils the given constraints $\{\mu_k^*\}_{k=0}^{M}$ since

$$\mu_k^* = \int_{\mathcal{S}_X} x^k f_M(x)\, dx, \qquad k = 0, 1, \dots, M\tag{3}$$

and has entropy

$$H_{f_M}(\mu_1^*, \dots, \mu_M^*) = -\int_{\mathcal{S}_X} f_M(x) \ln f_M(x)\, dx = \lambda_0 + \sum_{k=1}^{M} \lambda_j \mu_k^*\tag{4}$$

where

$$H_{f_M}(\mu_1^*, \dots, \mu_M^*) = \max_{f \in \mathcal{C}_M} H_f(\mu_1^*, \dots, \mu_M^*)$$

and

$$\mathcal{C}_M =: \left\{ f \geq 0 \mid \int_{\mathcal{S}_X} x^k f(x)\, dx = \mu_k^*,\ k = 0, \dots, M \right\} = \left\{ f_{(\mu_1^*, \dots, \mu_M^*)} \right\}.\tag{5}$$

From now on, for sake of brevity, we will write each member f of $\mathcal{C}_M$ omitting the dependency on $(\mu_1^*, \dots, \mu_M^*)$, predetermined set of moments; hence, f and f_M will stand for $f_{(\mu_1^*, \dots, \mu_M^*)}$ and $f_{M|(\mu_1^*, \dots, \mu_M^*)} \in \mathcal{C}_M$, respectively. The same will be done for the corresponding entropies: we will write H_f and H_{f_M} in place of $H_f(\mu_1^*, \dots, \mu_M^*)$ and $H_{f_M}(\mu_1^*, \dots, \mu_M^*)$.

A few words about our notation are now opportune. Since in the sequel an arbitrary moment μ_j may play different roles, we establish to use

1. μ_j^* for prescribed moments;
2. μ_j for variable (free to vary) moments;
3. $\mu_{j,f_{j-1}}$ for the j-th moment of f_{j-1}, that is $\mu_{j,f_{j-1}} = \int_{\mathcal{S}_X} x^j f_{j-1}(x)\, dx$ (in general $\mu_j \neq \mu_j^*$)
4. μ_j^- to indicate the smallest value of μ_j, once $(\mu_1^*, \dots, \mu_{j-1}^*)$ are prescribed.

Our attention is solely addressed towards sequences $\{\mu_k^*\}_{k=1}^{\infty}$ whose underlying density f has finite entropy H_f. More precisely, only distributions with $H_f = -\infty$ are not considered. Indeed, once $\{\mu_k^*\}_{k=0}^{\infty}$ is assigned, $H_f = +\infty$ is not feasible, as it is well known in MaxEnt setup that $H_f \leq H_{f_2} = \frac{1}{2}\ln[2\pi e(\mu_2^* - (\mu_1^*)^2)]$ is finite because Lyapunov's inequality $\mu_2^* - (\mu_1^*)^2$ (Hamburger case) and $H_f \leq H_{f_1} = 1 + \ln \mu_1^*$ is finite for every $\mu_1^* > 0$ (Stieltjes case).

Here $(\lambda_0, \dots, \lambda_M)$ is the vector of Lagrange multipliers, with $\lambda_M \geq 0$ to guarantee integrability of f_M. If it is possible to determine Lagrange multipliers from the constraints $\{\mu_k^*\}_{k=1}^{M}$ then the moment problem admits solution and f_M is MaxEnt solution (which is unique in S due to strict concavity of (4)).

The above non negativity condition on λ_M which is a consequence of unbounded support $\mathcal{S}_X$, is crucial and renders the moment problem solvable only under certain

restrictive assumptions on the prescribed moment vector $(\mu_1^*, \ldots, \mu_M^*)$. This is the ultimate reason upon which the present paper relies.

The existence conditions of the MaxEnt solution f_M have been deeply investigated in literature ([5–9] just to mention some widely cited papers); over the years an intense debate—combining the results of the above papers—has established the correct existence conditions underlying the Stieltjes and Hamburger moment problem (more details on this topic may be found in the Appendix A).

On the other hand, when the existence conditions for f_M are not satisfied, the non-existence of the MaxEnt solution in Stieltjes and Hamburger reduced moment problem poses a series of interesting and important questions about how to find an approximant of the unknown density f least committed to the information not given to us (still obeying to Jaynes' Principle). This problem is addressed the present paper.

More formally, take $\mathcal{C}_M$ to be the set of the density functions satisfying the $M + 1$ moment constraints (that is, they share the same $M + 1$ predetermined moments) and let $\mu(\mathcal{C}_M)$ be the moment space associated to $\mathcal{C}_M$; hence, the indeterminacy of the moment problem (1) follows.

A common way to regularize the problem, as recalled before, consists in applying the MaxEnt Principle obtaining $\mathcal{E}_M$, the set of MaxEnt densities functions which is a subset of $\mathcal{C}_M$; consequently, let $\mu(\mathcal{E}_M)$ be the moment space relative to the set of MaxEnt densities functions $\mathcal{E}_M$. Because, in general, $\mu(\mathcal{C}_M)$ strictly includes $\mu(\mathcal{E}_M)$ there are admissible moment vectors in $\text{Int}(\mu(\mathcal{C}_M))$, the interior of $\mu(\mathcal{C}_M)$, for which the moment problem (1) is solvable but the MaxEnt problem (3) has no solution and the usual regularization based on MaxEnt strategy is therefore precluded.

The implications of such issue are often understated in practical applications where the usual procedure limits itself to:

1. In the Stieltjes or symmetric Hamburger cases: to replace the support $\mathbb{R}^+$ ($\mathbb{R}$) with an arbitrarily large interval $[a, b]$. As a consequence of it, the problem is numerically solved within a proper interval $[a, b]$, changing the original Stieltjes (Hamburger) moment problem into Hausdorff one. In the MaxEnt setup, the latter admits a solution for each set $\{\mu_k^*\}_{k=1}^M \in \text{Int}(\mu(\mathcal{C}_M))$ ([7], Theorem 2).

2. If f_M does not exist (conclusion drawn uniquely from numerical evidence), f_{M-1} (Stieltjes case) or f_{M-2} (symmetric Hamburger case) always exist (see Appendix A). In such a case, although the first M moments are known, we have to settle for a density constrained by $\{\mu_k^*\}_{k=1}^{M-1}$ or $\{\mu_k^*\}_{k=1}^{M-2}$. However, this is not completely coherent with the MaxEnt principle that prescribes to use not only the available but all the available information; hence discarding available information seems to be conceptually in contrast with the MaxEnt spirit. However, from the point of practical applications, to consider or not to consider the prefixed moment μ_M^* seems to have negligible effects on the summarizing quantities of the underlying distribution (mostly expected values of suitable functions) in which we may be interested in. We will resume this issue, after having carefully motivated and proved the proposed solution, in the last section of the paper devoted to discussion and conclusions.

We call the solutions 1. and 2. "forced" pseudo-solutions; they might indeed lead to the unpleasant fact that a MaxEnt solution always exists, although the original Stieltjes (Hamburger) moment problem does not admit any solution. Hence the crucial question is: does there exist a way to regularize the (indeterminate) moment problem (1) coherently with all and only the available information exploiting the MaxEnt rationale setup without forcing to unnatural solutions, i.e., based on totally inappropriate application of the MaxEnt principle?

Before proceed recall $\mathcal{C}_M$ and define the following class of density functions:

$$\tilde{\mathcal{C}}_M =: \{f \geq 0 \mid \int_S x^k f(x) \, dx = \mu_k^*, \ k = 0, \ldots, M, \ \mu_{M+1} = +\infty\} \tag{6}$$

with $\tilde{\mathcal{C}}_M \subset \mathcal{C}_M$, whose entries satisfy the given constraints expressed in terms of $M+1$ assigned integer moments $\mu_k^* = \mathbb{E}(X^k)$, $k = 0, 1, \ldots, M$.

Now the question is: once $\{\mu_k^*\}_{k=1}^M \in \mu(\mathcal{C}_M)\backslash\mu(\mathcal{E}_M)$ are pre-determined (that is, the MaxEnt problem does not admit solution) what is the optimal choice of the pdf that we can select in place of f_M? Relying upon the MaxEnt rationale, the best substitute of the missing f_M should be given by suitable one $\tilde{f}_M \in \mathcal{C}_M$ having the overall largest entropy; that is, select $\tilde{f}_M \in \mathcal{C}_M$ actually satisfying the relationship

$$\sup_{f \in \mathcal{C}_M} H_f - H_{\tilde{f}_M} < \varepsilon \tag{7}$$

for an arbitrarily small ε.

We are aimed to find $\sup_{f \in \mathcal{C}_M} H_f$, $\tilde{f}_M$ and the corresponding entropy $H_{\tilde{f}_M}$, proving that it may be accomplished by MaxEnt machinery (see Equations (9)–(11) below).

The remainder of the paper is organized as follows. Sections 2 and 3 are devoted to evaluating the best pdf in Stieltjes and Hamburger cases respectively. We devote Section 4 to numerical aspects and in Section 5 we round up with some concluding remarks. In Appendix A. the existence conditions of MaxEnt distributions in Stieltjes and Hamburger case are shortly reviewed.

2. Stieltjes Case

Let us consider $\{\mu_k^*\}_{k=1}^M \in \mu(\mathcal{C}_M)\backslash\mu(\mathcal{E}_M)$; consequently f_M does not exist. In this section we provide a formal justification about motivation (rationale) and optimality of the proposed substitute $\tilde{f}_M$ of the MaxEnt density f_M. We deal with the issue of selecting the "best" pdf both satisfying the constraints (given by predetermined integer moments) and with the overall largest entropy.

Before start, some relevant facts need to be collected together. Since MaxEnt density f_M does not exist both f_{M-1} with its M-th moment $\mu_{M,f_{M-1}} \leq \mu_M^*$ and

$$f_{M+1} = f_{M+1}(\mu_{M+1}) = f_{M+1|\mu_1^*,\ldots,\mu_M^*,\mu_{M+1}} \in \mathcal{C}_M$$

exist; the latter exists for any value $\mu_{M+1} > \mu_{M+1}^-$ (see Appendix A for more details).

Since the procedure here adopted remains valid for each value $\mu_M^* > \mu_{M,f_{M-1}}$, as $\mu_M^* \to +\infty$ from Lyapunov's inequality, we have $\mu_{M+1} > (\mu_M^*)^{(1+\frac{1}{M})}$ and consequently $\mu_{M+1} \to +\infty$ too. As well, since MaxEnt density does not exist, some additional information not given to us must be added; of course, μ_{M+1} is the most suitable candidate to represent it.

Once this is established, the relevant question is: what value for μ_{M+1}? Recalling that $H_{f_{M+1}}(\mu_1^*,\ldots,\mu_M^*,\mu_{M+1}) = H_{f_{M+1}}(\mu_{M+1})$ is monotonic increasing ([4], Equation (2.73), p. 59), with upper bound $H_{f_{M-1}}$ so that $\lim_{\mu_{M+1} \to \infty} H_{f_{M+1}}(\mu_{M+1})$ exists, μ_{M+1} should assume the overall largest value, so that the decreasing of entropy is as small as possible.

Since $\{\mu_k^*\}_{k=1}^M \in \mu(\mathcal{C}_M)\backslash\mu(\mathcal{E}_M)$, f_M and then H_{f_M} are meaningless, $\mathcal{C}_M$ includes infinitely many f and $\sup_{f \in \mathcal{C}_M} H_f$ must be calculated. The moment set $(\mu_1^*,\ldots,\mu_{M-1}^*,\mu_M^* > \mu_{M,f_{M-1}},\mu_{M+1})$ is considered too, where $\mu_{M,f_{M-1}}$ is the M-th order moment of f_{M-1} and μ_{M+1} varies continuously within $\mu(\mathcal{E}_{M+1})$ with

$$(\mu_1^*,\ldots,\mu_M^*,\mu_{M+1}^-) \in \partial\mu(\mathcal{C}_{M+1}). \tag{8}$$

If $f_{M+1}(\mu_{M+1})$ is density corresponding to the set of moments (8), the following theorem holds.

Theorem 1. *The following two relationships hold*

$$\sup_{f \in \mathcal{C}_M} H_f = H_{f_{M-1}} \tag{9}$$

and

$$\lim_{\mu_{M+1} \to \infty} H_{f_{M+1}}(\mu_{M+1}) = H_{f_{M-1}}. \tag{10}$$

Now, $\tilde{f}_M$ is identified with $f_{M+1}(\tilde{\mu}_{M+1})$ where $\tilde{\mu}_{M+1}$ is such that

$$H_{f_{M-1}} - H_{f_{M+1}}(\tilde{\mu}_{M+1}) < \varepsilon \tag{11}$$

and ε indicates a fixed tolerance.

Proof. If f_M does not exist, f_{M-1} exists with entropy $H_{f_{M-1}}$ and M-th moment $\mu_{M,f_{M-1}} < \mu_M^*$ respectively. The function $H_{f_M}(\mu_M)$, with $\mu_M > \mu_M^-$, is monotonic increasing. As

1. $\mu_M^* \leq \mu_{M,f_{M-1}}$ one has $\sup_{f \in \mathcal{C}_M} H_f = \max_{f \in \mathcal{C}_M} H_f = H_{f_M} \leq H_{f_{M-1}}$. The latter represents the maximum attainable entropy once $(\mu_1^*, \dots, \mu_M^*)$ are prescribed;

2. $\mu_M^* > \mu_{M,f_{M-1}}$, from monotonicity of $H_{f_M}(\mu_M)$ it follows $\sup_{f \in \mathcal{C}_M} H_f = H_{f_{M-1}}$ independent on μ_M^*. Equivalently, $H_{f_{M-1}}$ is strict upper bound for the entropies of all densities which have same lower moments $(\mu_1^*, \dots, \mu_{M-1}^*)$ as f_{M-1} but whose highest moment μ_M^* exceeds $\mu_{M,f_{M-1}}$.

Hence Equation (9) is proved.

Let us now consider the suitable class

$$\mathcal{E}_{M+1}(\mu_{M+1}) = \left\{ f_{M+1}(\mu_{M+1}) = f_{M+1|\mu_1^*, \dots, \mu_M^*, \mu_{M+1}}(x) \right\} \tag{12}$$

where $\mu_{M+1} \in (\mu_{M+1}^-, \infty)$ is assumed as parameter and $\mu_{M,f_{M-1}}$ is the M-th order moment of f_{M-1}. Equivalently, the entries of $\mathcal{E}_{M+1}(\mu_{M+1})$ are MaxEnt pdfs constrained by $(\mu_1^*, \dots, \mu_M^*, \mu_{M+1})$, belong to $\mathcal{C}_M$ and, primarily, they all have analytically tractable entropy.

In (5), (6) and (12) three classes of functions $\mathcal{C}_M$, $\tilde{\mathcal{C}}_M$ and $\mathcal{E}_{M+1}(\mu_{M+1})$ had been defined. Relying upon the identity $\mathcal{C}_M = \mathcal{E}_{M+1}(\mu_{M+1}) \cup (\mathcal{C}_M \backslash \mathcal{E}_{M+1}(\mu_{M+1}) \backslash \tilde{\mathcal{C}}_M) \cup \tilde{\mathcal{C}}_M$ we investigate the entropy H_f of functions f belonging to (a) $\mathcal{E}_{M+1}(\mu_{M+1})$, (b) $\mathcal{C}_M \backslash \mathcal{E}_{M+1}(\mu_{M+1}) \backslash \tilde{\mathcal{C}}_M$ and (c) $\tilde{\mathcal{C}}_M$ respectively.

1. Consider that $H_{f_{M+1}}(\mu_{M+1})$, bounded by $H_{f_{M-1}}$ from above, is a differentiable monotonic increasing function of μ_{M+1} and then it tends to a finite limit so that

$$\sup_{f_{M+1|\mu_{M+1} \in \mathcal{E}_{M+1}(\mu_{M+1})}} H_{f_{M+1}}(\mu_{M+1}) = \lim_{\mu_{M+1} \to \infty} H_{f_{M+1}}(\mu_{M+1}).$$

2. Each $f \in \mathcal{C}_M \backslash \mathcal{E}_{M+1}(\mu_{M+1}) \backslash \tilde{\mathcal{C}}_M$ has entropy H_f and its $(M+1)$-th finite moment, say $\mu_{M+1,f}$. Since f and $f_{M+1|\mu_{M+1,f}}$ share same moments $(\mu_1^*, \dots, \mu_M^*, \mu_{M+1,f})$ then $H_f \leq H_{f_{M+1}}(\mu_{M+1,f})$ holds, from which

$$\sup_{f \in (\mathcal{C}_M \backslash \mathcal{E}_{M+1}(\mu_{M+1}) \backslash \tilde{\mathcal{C}}_M)} H_f \leq \sup_{f_{M+1|\mu_{M+1} \in \mathcal{E}_{M+1}(\mu_{M+1})}} H_{f_{M+1}}(\mu_{M+1}) = \lim_{\mu_{M+1} \to \infty} H_{f_{M+1}}(\mu_{M+1}).$$

3. In analogy with (12), let us introduce the following class

$$\mathcal{C}_{M+1}(\mu_{M+1}) = \left\{ f(\mu_{M+1}) = f_{|\mu_1^*, \dots, \mu_M^*, \mu_{M+1}}(x) \right\}$$

where $\mu_{M+1} > \mu_{M+1}^-$ assumes arbitrary values. For a fixed μ_{M+1}, each $f \in \mathcal{C}_{M+1}(\mu_{M+1})$ satisfies the following inequality

$$H_f \leq \sup_{f \in \mathcal{C}_{M+1}(\mu_{M+1})} H_f = H_{f_{M+1}}(\mu_{M+1}).$$

Taking $\mu_{M+1} \to \infty$, $\mathcal{C}_{M+1}(\mu_{M+1})$ coincides with $\tilde{\mathcal{C}}_M$ and then

$$\sup_{f \in \tilde{\mathcal{C}}_M} H_f = \lim_{\mu_{M+1} \to \infty} H_{f_{M+1}}(\mu_{M+1}).$$

Collecting together both the achieved results in above items (a), (b), (c) and taking into account (9) one has

$$\sup_{f \in \mathcal{C}_M} H_f = \lim_{\mu_{M+1} \to \infty} H_{f_{M+1}}(\mu_{M+1}) = H_{f_{M-1}}.$$

Hence Equation (10) is also proved.

Equation (10) is restated as follows: if ε indicates a fixed tolerance, there exists a value $\tilde{\mu}_{M+1}$ of μ_{M+1} such that $H_{f_{M-1}} - H_{f_{M+1}}(\tilde{\mu}_{M+1}) < \varepsilon$ holds. Next $\tilde{f}_M$ is identified with $f_{M+1}(\tilde{\mu}_{M+1})$ so that its entropy $H_{\tilde{f}_M}$ coincides with $H_{f_{M+1}}(\tilde{\mu}_{M+1})$. From which the wanted result $H_{f_{M-1}} - H_{\tilde{f}_M} < \varepsilon$ (or, equivalently (7)) follows. As a consequence $\tilde{f}_M$ is the proposed substitute of f_M and Equation (11) is proved. $\square$

In conclusion:

1. As f_M does not exist, although the current use of MaxEnt fails, a solution is found back to (9)–(11) from which the desired result (7).
2. The existence of MaxEnt f_M implies its uniqueness, unlike $\tilde{f}_M$ which depends on the assumed tolerance. In numerical Examples below just above remark will be actually used.

3. Hamburger Case

The non-symmetric Hamburger case when M even is here disregarded because the existence of the MaxEnt solution f_M is guaranteed. Now, we will concentrate our attention on the symmetric case with $M \geq 4$ and on the non-symmetric case with $M \geq 3$ odd. In both cases, thanks to MaxEnt formalism, the procedure used in Stieltjes case can be extended to Hamburger one (see [9]); this fact represents one of the main advantages of MaxEnt machinery.

3.1. Symmetric Case with $M \geq 4$ Even

We recall f_M is symmetric function for every M even so that Lagrange multipliers $\lambda_{2j-1} = 0$.

Theorem 2. *Suppose the moment set $(\mu_1^*, ..., \mu_M^*)$ is prescribed and f_M does not exist. Symmetric Hamburger case is analogous to Stieltjes one and then Theorem 1 holds true, with μ_{M+1} replaced by μ_{M+2} and μ_{M-1} by μ_{M-2}.*

Proof. Just remember that if f_M does not exist for a prescribed moment set $(\mu_1^*, ..., \mu_M^*)$ then f_{M-2} exists with its next moments $\mu_{M-1,f_{M-2}}$, $\mu_{M,f_{M-2}}$ and entropy $H_{f_{M-2}}$. If $\mu_M^* = \mu_{M,f_{M-2}}$ holds, then $H_{f_{M-2}}$ is the maximum attainable entropy. In analogy with (12) let

$$\mathcal{E}_{M+2} = \left\{ f_{M+2}(\mu_{M+2}) =: f_{M+2|(\mu_1^*,...,\mu_M^*,\mu_{M+1}^*=0,\mu_{M+2}},\ \mu_{M+2} \in (\mu_{M+2}^-, \infty) \right\}$$

where the parameter μ_{M+2} is introduced and thanks to MaxEnt machinery the proof continues analogously to the Stieltjes case. $\square$

3.2. Non-Symmetric Case with $M \geq 3$ Odd

If M is odd f_M does not exist for every set of moments belonging to $\mu(\mathcal{C}_M)$ because $\int_{\mathbb{R}} f_M(x)\, dx = +\infty$. We now look at the problem from a different point of view. Suppose $(\mu_1^*, ..., \mu_M^*) \in \mu(\mathcal{C}_M)$ is prescribed. In general f_{M-1} exists (equivalently, $(\mu_1^*, ..., \mu_{M-1}^*) \in \mu(\mathcal{E}_{M-1})$ (see Appendix), with its M-th moment $\mu_{M,f_{M-1}}$. Two alternatives are possible:

1. $\mu_M^* = \mu_{M,f_{M-1}}$ then f_M with $\lambda_M = 0$ exists and coincides with f_{M-1}. Then usual MaxEnt method may be used;

2. $\mu_M^* \neq \mu_{M,f_{M-1}}$ (highly probable case), then f_M does not exist.

Both items 1. and 2. recall the Stieltjes case; more precisely, item (ii) may be solved taking into account $f_{M+1}(\mu_{M+1}) \in \mathcal{E}_{M+1}$ (which exists for each $\mu_{M+1} > \mu_{M+1}^-$). Then non-symmetric case, with M odd and for every set of moments belonging to $\mu(\mathcal{C}_M)$, is solved analogously to the Stieltjes case and Theorem 1 holds true. A consequence of achieved results in this section is the following. Let us consider a non symmetric Hamburger moment. In Theorem 1 we proved

$$H_{f_{M-1}} = \sup_{f \in \mathcal{C}_M} H_f = \lim_{\mu_{M+1} \to \infty} H_{f_{M+1}}(\mu_{M+1}).$$

Since the entropy is monotonic non increasing as M increases, the latter equalities enable us to set $H_{f_M} = H_{f_{M-1}}$. As a consequence, the two subsequences $\{H_{f_{2M}}\}$ and $\{H_{f_{2M-1}}\}$ have coinciding entries. In past paper Milev and Tagliani proved that $\{H_{f_{2M}}\}$ converges to H_f ([10], Theorem 1) . Joining together the two achievements, both $\{H_{f_{2M}}\}$ and $\{H_{f_{2M-1}}\}$ converge to the same limit H_f, then so does $\{H_{f_M}\}$, filling the gaps left by even moments.

4. Numerical Aspects

The procedure just above described and rooted on MaxEnt machinery suffers from some numerical drawbacks which will be here discussed. It deserves to recall similar drawbacks had been previously found ([11,12]) although for the special value $M = 4$ in Hamburger case, exploring special regions of the moment space. Essentially, numerical troubles arise because the expected solution f_{M+1} is contaminated with a small wiggle that (a) moving to infinity, (b) is scaled in such a way that its contribution to the $(M + 1)$-th order moment μ_{M+1} is always $O(1)$ and (c) may become invisible to numerical methods of quadrature.

Now we provide some theoretical ground to justify above heuristics, which holds true in both Hamburger and Stieltjes case thanks to MaxEnt formalism. First of all, under the constraints $(\mu_1^*, \dots, \mu_M^*, \mu_{M+1})$, we prove the wiggle exists. At this purpose both the relationships $\lambda_M < 0$ for each μ_{M+1} and $\lambda_{M+1} \to 0$ as $\mu_{M+1} \to \infty$ have to be proved.

1. $\lambda_M < 0$ for each μ_{M+1}. From Appendix we recalled if f_{M-1} exists with its M-th moment $\mu_{M,f_{M-1}}$, MaxEnt f_M does not exist if $\mu_M^* > \mu_{M,f_{M-1}}$. Let us consider f_M where μ_M varies continuously. Then λ_M is monotonically decreasing ([9], Equation (2.1)) with $\lambda_M = 0$ as $\mu_M = \mu_{M,f_{M-1}}$. As a consequence no set $(\lambda_1, \dots, \lambda_M)$ satisfies the constraints $(\mu_1^*, \dots, \mu_M^*)$ since the monotonicity of λ_M would require $\lambda_M < 0$. Let us consider f_{M+1} where μ_{M+1} varies continuously. Here, for each μ_{M+1}, $\lambda_{M+1} > 0$ guarantees integrability, so that λ_M may assume every real value. Collecting together the results about λ_M, the set $(\lambda_1, \dots, \lambda_{M-1}, \lambda_M < 0, \lambda_{M+1})$ satisfies the constraints $(\mu_1^*, \dots, \mu_M^*, \mu_{M+1})$ for each μ_{M+1}. Equivalently, we can assert $(\lambda_1, \dots, \lambda_{M-1}, \lambda_M < 0)$ are appointed to meet $(\mu_1^*, \dots, \mu_M^*)$, whilst λ_{M+1} to meet μ_{M+1}.

2. $\lambda_{M+1} \to 0$ as $\mu_{M+1} \to \infty$. Differentiating (4) with respect to λ_{M+1} and recalling the relationship ([9], Equation (2.1))

$$\sum_0^M \mu_j^* \frac{d\lambda_j}{d\mu_{M+1}} + \mu_{M+1} \frac{d\lambda_{M+1}}{d\mu_{M+1}} = 0$$

one has

$$\frac{dH_{f_{M+1}}(\mu_{M+1})}{d\mu_{M+1}} = \sum_0^M \mu_j^* \frac{d\lambda_j}{d\mu_{M+1}} + \mu_{M+1} \frac{d\lambda_{M+1}}{d\mu_{M+1}} + \lambda_{M+1} = \lambda_{M+1}.$$

From Theorem 1 we proved, as $\mu_{M+1} \to \infty$, $H_{f_{M+1}} \to H_{f_{M-1}}$, so that $\frac{dH_{f_{M+1}(\mu_{M+1})}}{d\mu_{M+1}} \to 0$ and then $\lambda_{M+1} \to 0$ too.

We are ready to prove the statement concerning the fact that $f_{M+1}(\mu_{M+1})$ exhibits a small wiggle at $x \gg 1$ (analogously, in symmetric Hamburger case the wiggle is exhibited at $|x| \gg 1$). At $x \gg 1$

$$f_{M+1}(x) \sim \exp\left(-\lambda_M x^M - \lambda_{M+1} x^{M+1}\right)$$

so that f_{M+1} admits maximum value at

$$x_{\text{wig}} = -\frac{M\lambda_M}{(M+1)\lambda_{M+1}} > 0.$$

As μ_{M+1} increases we proved the relationships $\lambda_M < 0$ and $\lambda_{M+1} \to 0$, so that $x_{\text{wig}} > 0$ moves to infinity (from numerical evidence, as μ_{M+1} increases, $|\lambda_M| \to 0$ too much slower than λ_{M+1}). Since f_{M+1} has finite moments $(\mu_1^*, ..., \mu_M^*)$ for each μ_{M+1}, it follows the wiggle in a compact packet is scaled in such a way that its contribution to the $(M+1)$-th order moment μ_{M+1} is always $O(1)$ (whilst for all higher moments the contribution due to this maximum obviously grows without bound, as a consequence of Lyapunov's inequality).

An additional complication comes from the fact that height and position of wiggle is extremely sensitive to the parameters λ_M and λ_{M+1}, so that it becomes progressively smaller and smaller until to be "invisible" if an unsuitable numerical method of quadrature is adopted. As a consequence the procedure becomes increasingly ill-conditioned to such a degree that numerical error precludes finding a suitable solution. As remedy, for instance, the quadrature on the unbounded domain has to be mapped onto finite interval, as well an adaptive quadrature is required. Since the wiggle moves along x-axis as μ_{M+1} increases, a fixed nodes quadrature formula could be unsuitable as the wiggle could become invisible for some values of μ_{M+1}.

Above remedies are just a numerical trick, not a reduction of Stieltjes or Hamburger problem into Hausdorff one. Indeed, all the subsequent numerical examples consider and use random variables X having unbounded support $\mathbb{R}^+$ or $\mathbb{R}$.

As well the dual formulation, which evaluates $(\lambda_1, ..., \lambda_{M+1})$ minimizing the potential function

$$\{\lambda_j\}_{j=1}^{M+1}: \quad \min_{\lambda_1, ..., \lambda_{M+1}} \left[\ln\left(\int_S \exp(-\sum_{j=1}^{M+1} \lambda_j x^j)dx\right) + \sum_{j=1}^{M} \lambda_j \mu_j^* + \lambda_{M+1}\mu_{M+1}\right]$$

avoids the computation of higher moments, as required by Newton-type methods by solving (3).

The drawbacks just illustrated lead us to equip the stopping criterion (7) based on entropy with a further one based on the moments, which allows us the relationship $\mu_{j,f_{M+1}} = \mu_j^*$, $j = 1, ..., M$ holds true. That is,

$$\max_{1 \leq j \leq M} \left| \frac{\mu_j^* - \mu_{j,f_{M+1}}}{\mu_j^*} \right| < \varepsilon_1 \tag{13}$$

(or involving the absolute error) for a proper ε_1.

The following question arises: it is $\tilde{f}_M$, here identified with $f_{M+1}(\tilde{\mu}_{M+1})$ and $\tilde{\mu}_{M+1}$ is chosen so that stopping criteria (7) and (13) are verified, an acceptable approximation of underlying unknown density? Although the wiggle has non-physical meaning, nevertheless from the approximate density one like to calculate accurate and interesting quantities. We will resume the issue in the final part of the paper.

For practical purposes in both Stieltjes and Hamburger case $\tilde{f}_M$ is calculated according to (9)–(11) uniquely by means of MaxEnt machinery following these two distinct steps

1. First, the sequence $\{\mu_k^*\}_{k=0}^M$ is prescribed and f_M does not exists; then we know f_{M-1} exists with entropy $H_{f_{M-1}}$;

2. The next step relies upon on the monotonicity of $H_{f_{M+1}}(\mu_{M+1})$. If ε indicates a fixed tolerance, $f_{M+1}(\mu_{M+1})$ is calculated taking increasing values of μ_{M+1} until for some $\tilde{\mu}_{M+1}$ inequality $H_{f_{M-1}} - H_{f_{M+1}}(\tilde{\mu}_{M+1}) < \varepsilon$ is satisfied, assuming implicitly (13) is satisfied too. Next $\tilde{f}_M \equiv f_{M+1}(\tilde{\mu}_{M+1})$ is set.

Before to illustrate some numerical examples that confirm the goodness of the proposed method, it is worth spend some words discussing the outlined procedure. The calculation of $\tilde{f}_M$ is obtained through an approximate procedure and hence has a limited range of applicability. The main problem is the presence of wiggles; at the end to contain their detrimental effect it is necessary that convergence of $H_{f_{M+1}}(\mu_{M+1})$ to $H_{f_{M-1}}$ be fast. For example, the value μ_M^* that precludes the existence of f_M in the Stieltjes case, must be such that the difference $\mu_M^* - \mu_{M,f_{M-1}}$ be small. Larger values make the convergence of $H_{f_{M+1}}(\mu_{M+1})$ to $H_{f_{M-1}}$ slow, allowing the generation of a small wiggle at great distance. The latter may become invisible to numerical quadrature methods.

Below are some numerical examples which both take into account the above remarks about the difference $\mu_M^* - \mu_{M,f_{M-1}}$ and illustrate the theoretical and numerical aspects mentioned above.

Example 1. *The Stieltjes case with $M = 2$ and prescribed (μ_1^*, μ_2^*) is considered. Now f_M exists if and only if the inequality $(\mu_1^*)^2 < \mu_2^* \leq 2(\mu_1^*)^2$ holds ([5], Theorem 2). The moment set $\{(\mu_1, \mu_2) \mid \mu_2 = 2\mu_1^2 \mid \mu_1 > 0\} \in Int(\mu(\mathcal{C}_M))$ represents an additional boundary in $\mu(\mathcal{C}_M)$. If the moments satisfy the reverse inequality $\mu_2^* > 2(\mu_1^*)^2$ there is no pdf which maximizes the entropy.*

We consider the latter case taking $\mu_1^* = 1$ and $\mu_2^* = 2.1$; then $f_{M+1}(\tilde{\mu}_{M+1})$ is calculated by means of (11), with $H_{f_{M-1}} = 1 + \ln \mu_1^* = 1$. Values of entropy $H_{f_{M+1}}(\mu_{M+1})$ with increasing values of $\mu_{M+1} > \mu_{M+1}^- = 4.41$ not reported here lead to conclusion that the entropy stabilizes rapidly as μ_{M+1} increases. This may be an evidence of high accuracy in the reconstruction.

Taking $\varepsilon = 10^{-4}$, Equation (11) is satisfied starting from $\tilde{\mu}_{M+1} = 20$. Then $\tilde{f}_M$, which is identified with $f_{M+1}(\tilde{\mu}_{M+1})$, jointly with f_{M-1} are displayed in Figure 1 (top). The difference between $\tilde{f}_M$ and f_{M-1} is insignificant since $\mu_M^* - \mu_{M,f_{M-1}} = 0.1$ was chosen to avoid the detrimental effect of wiggle. In Figure 1 (bottom) the same $f_{M+1}(\tilde{\mu}_{M+1})$, on a logarithmic scale and on extended x-axis scale, is reported to evidenciate the presence of small wiggle. The moments $\mu_{1,\tilde{f}_M}, \mu_{2,\tilde{f}_M}$ satisfy $\mid \mu_1^* - \mu_{1,\tilde{f}_M} \mid \sim 10^{-8}$, $\mid \mu_2^* - \mu_{2,\tilde{f}_M} \mid \sim 10^{-6}$, respectively. It can be concluded $\tilde{f}_M \equiv f_{M+1}(\tilde{\mu}_{M+1})$ satisfies all the expected theoretical properties and can be considered the "best" substitute of the missing f_M.
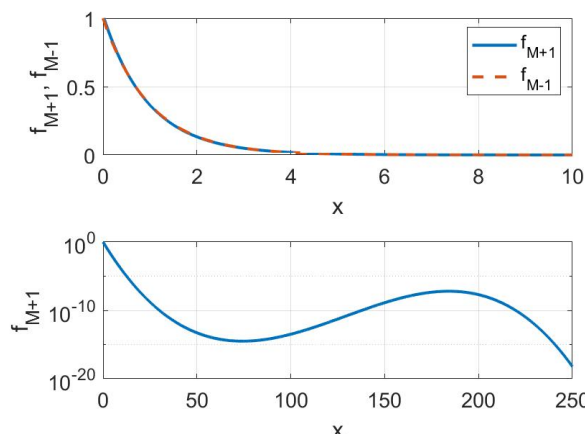
Figure 1. Stieltjes case, $M = 2$. $f_{M+1}(\tilde{\mu}_{M+1})$ and f_{M-1} (top). $f_{M+1}(\tilde{\mu}_{M+1})$ in logarithmic scale (bottom).

Example 2. *Hamburger case with $M = 3$; here $\mu_1^* = 0$, $\mu_2^* = 1$, $\mu_3^* = 0.5$, with the skewness = 0.5, are assumed. MaxEnt density f_M does not exist if the first 3 moments are specified and the skewness is required to be non-zero. Then f_M does not exists whilst f_{M-1} (Normal distribution) exists, with entropy $H_{f_{M-1}} = \frac{1}{2} \ln[2\pi e(\mu_2^* - (\mu_1^*)^2)] \simeq 1.4189385$. Taking $\varepsilon = 10^{-3}$, Equation (11) is satisfied starting from $\tilde{\mu}_{M+1} = 14.65$. Then $\tilde{f}_M$, which is identified with $f_{M+1}(\tilde{\mu}_{M+1})$, jointly with f_{M-1} are displayed in Figure 2 (top). In Figure 2 (bottom) the same $f_{M+1}(\tilde{\mu}_{M+1})$ is displayed but on a logarithmic scale and on extended x-axis scale, to highlight the presence of small wiggle.*

The moments $\mu_{1,\tilde{f}_M}$, $\mu_{2,\tilde{f}_M}$, $\mu_{3,\tilde{f}_M}$ satisfy $\mid \mu_1^* - \mu_{1,\tilde{f}_M} \mid \sim 10^{-7}$, $\mid \mu_2^* - \mu_{2,\tilde{f}_M} \mid \sim 10^{-7}$, $\mid \mu_3^* - \mu_{3,\tilde{f}_M} \mid \sim 10^{-5}$, respectively. It can be concluded $\tilde{f}_M$ satisfies all the expected theoretical properties and can be considered the "best" substitute of the missing f_M.
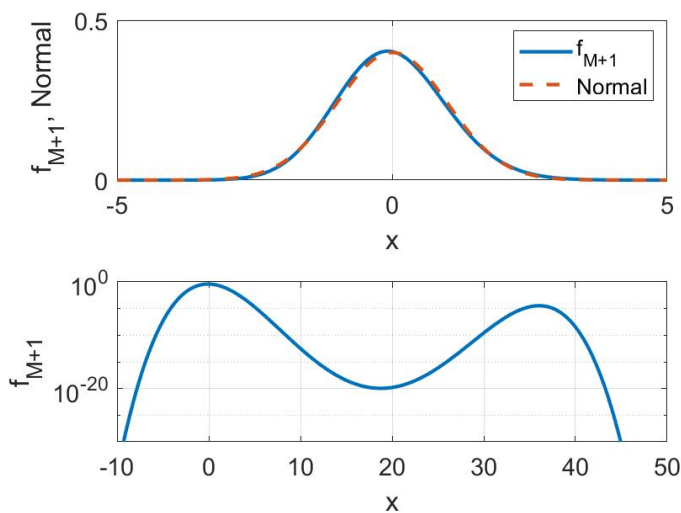

Figure 2. Hamburger case, $M = 3$. $f_{M+1}(\tilde{\mu}_{M+1})$ and Normal (top) . $f_{M+1}(\tilde{\mu}_{M+1})$ in logarithmic scale (bottom).

Remark 1. *It is worth to note that the nonsymmetric Hamburger case with $M = 3$ has been discussed in [13], pp. 413–415, Equation (12.32), but solely on the basis of a simple heuristic reasoning; they use a tricky problem to observe that even if the Lagrange multipliers cannot be chosen to satisfy the given constraints, the "maximum" entropy can be found and it is equal to*

$$\sup_{f \in \mathcal{C}_M} H_f = H_{f_{M-1}}$$

concluding that in this situation the entropy may only be ϵ-achievable. Just to give a simple example of it, but not a formal justification, the authors consider the case in which a Normal distribution be contaminated with a small "wiggle" at a very high value of x; consequently the moments of new distribution are almost the same as those of the non contaminated Normal, the biggest change being in the third moment (the new distribution is not any more symmetric). However, adding new wiggles in opportune positions to balance the changes caused by the original wiggle we can bring the first and the second moments back to their original values and also get any value of the third moment without reducing the entropy significantly below that of the associated non contaminated Normal (from this the conclusion about the ϵ-achievability of the entropy).

Just above heuristic procedure is displayed in Figure 2, and interpreted saying $\tilde{f}_M$ may be identified with the Normal distribution on which some wiggles are superimposed.

This result is a particular case of the more general result covered by this paper and coincides with the above (9)–(11) when, in this case, f_{M-1} is the density function of a Normal distribution.

Lastly, all above heuristics agrees with the mathematical general result that two continuous density functions having the same first $M + 1$ moments (including $\mu_0 = 1$) cross each other in at least $M + 1$ points ([14], Vol.1, No. 140, p. 83). In our case f_{M+1} and the Normal density plotted in Figure 2, share the first $M + 1 = 3$ moments and they cross each other at three points as the inspection of the previous figure suggests.

Example 3. *Symmetric Hamburger case with $M = 4$, prescribed $(\mu_2^* = 1, \mu_4^* = 4)$ and MaxEnt density f_M are considered. f_{M-2} is the Normal distribution with $\mu_{M,f_{M-2}} = 3$ and entropy $H_{f_{M-2}} = \frac{1}{2} \ln[2\pi e \mu_2^*] \simeq 1.4189385$. f_M does not exists, being its existence condition $\mu_M^* \leq \mu_{M,f_{M-2}} = 3$ not verified. A further even moment μ_{M+2} with increasing values is added and $f_{M+2}(\mu_{M+2})$ has to be calculated.*

Taking $\varepsilon = 10^{-3}$, Equation (11) is satisfied starting from $\tilde{\mu}_{M+2} = 160$. Then $\tilde{f}_M$, which is identified with $f_{M+2}(\tilde{\mu}_{M+2})$, jointly with f_{M-2} are displayed in Figure 3 (top). In Figure 3 (bottom) the same $f_{M+2}(\tilde{\mu}_{M+2})$, in logarithmic scale and on extended x-axis scale, is reported to highlight the presence of two symmetric wiggles travelling in opposite direction and illustrated too in same Figure (bottom). The moments $\mu_{2,\tilde{f}_M}$, $\mu_{4,\tilde{f}_M}$ satisfy $| \mu_2^* - \mu_{2,\tilde{f}_M} | \sim 10^{-8}$, $| \mu_4^* - \mu_{4,\tilde{f}_M} | \sim 10^{-6}$, respectively.
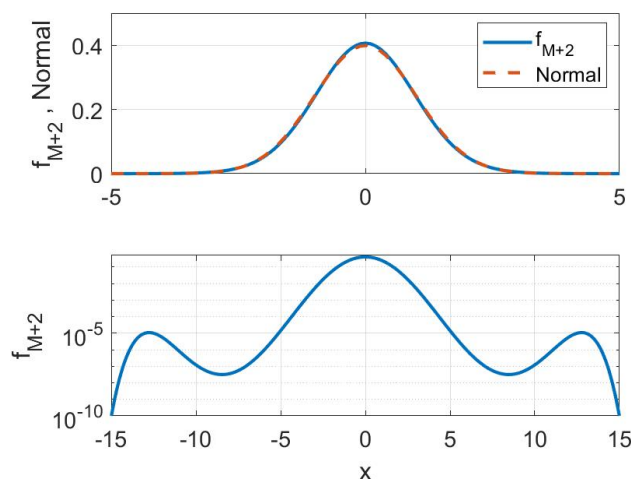

Figure 3. Symmetric Hamburger case with $M = 4$. $f_{M+2}(\tilde{\mu}_{M+2})$ and Normal (top). $f_{M+2}(\tilde{\mu}_{M+2})$ in logarithmic scale (bottom).

In each of the previous three examples we have assumed that μ_M^* and $\mu_{M,f_{M-1}}$ differ from a small amount and this to avoid the detrimental effect due to the wiggle; consequently, the difference between $\tilde{f}_M$ and f_{M-1} becomes insignificant too. As a result,

1. *The convergence of $H_{f_{M+1}}$ to $H_{f_{M-1}}$ is fast and avoids the formation of small evanescent wiggles at a great distance;*
2. *The rise of numerical quadrature problems.*

As a consequence the two densities f_{M+1} and f_{M-1} are almost superimposed precluding the possibility to evaluate the effect produced in f_{M-1} from having discarded μ_M^*. It would be interesting to be able to assess how high values $\mu_M^* - \mu_{M,f_{M-1}}$ may affect the difference $f_{M+1} - f_{M-1}$. The goal could be achieved by a suitable numerical quadrature method.

5. Discussion and Conclusions

In the present paper, we have discussed the case which arises when, in presence of a prefixed moment set $(\mu_1^*, \mu_2^*, \ldots, \mu_M^*)$ representing the available information, the (reduced) moment problem admits solution but the MaxEnt density as a solution of the regularization problem does not exist. In the previous sections, we have given the conditions under which a solution of the Stieltjes and Hamburger (reduced) moment problems may be found in the genuine Jaynes' spirit by finding the overall largest entropy distribution which is compatible with the available information and showing that this is the best approximant of the underlying true but unknown distribution. The substitute of the missing MaxEnt solution is found using solely the usual MaxEnt machinery.

Now we look at the issue from a different point of view. Suppose $(\mu_1^*, \mu_2^*, \ldots, \mu_M^*)$ represent all and only the available information. Two cases 1. and 2. may present:

1. Only the first M moments may be measured but additionally the f_M exists. In this situation the traditional MaxEnt machinery will produce the usual solution f_M which has a well known analytical form corresponding to the Jaynes' non committal approximant (MaxEnt) of the underlying f (see Equation (2));
2. Only the first M moments may be measured but additionally the f_M does not exist. Here any information about the analytical form of the substitute of the missing MaxEnt solution is lacking. If only the first M moments may be measured, it is reasonable to assume the underlying f admits the first M moments solely. Then it can to be restated $(\mu_1^*, \mu_2^*, \ldots, \mu_M^*, \mu_{M+1} = +\infty)$ to represent all and only the available information. Next, assuming μ_{M+1} takes finite value, MaxEnt machinery may be invoked, from which f_{M+1} as above and the consequent Theorem 1. To find a genuine minimal committal approximant in the MaxEnt spirit of the underlying f just

$$\lim_{\mu_{M+1} \to \infty} H_{f_{M+1}}(\mu_{M+1})$$

is taken, so that, from the monotonicity of $H_{f_{M+1}}$, the spurious information represented by μ_{M+1} has a minimum effect on the approximant (in other terms, to guarantee to be minimal committal).

The solution we have proposed in this paper for case 2. offers an alternative and exhaustive answer to the common empirical "forced" practices consisting in

(a) Replacing an unbounded support with an arbitrarily large interval, or
(b) Neglecting the prescribed higher moment so that the reduced number of moments allows the existence of MaxEnt solution.

As we have widely said before (see Introduction), solutions like (a) and (b) imply a forced pseudo-solution of the original problem which conflicts with MaxEnt rationale.

The above conflict is not merely theoretical and it has some practical consequences. This leads us to distinguish theoretical and practical aspects of the procedure we proposed. MaxEnt technique is invoked because one reputes the found distribution to be "the best" and the obtained results are "the best". Essentially this is the practitioner's main concern. More specifically, since the MaxEnt distribution constrained by first M moments does not exist, we are inclined to turn to f_{M-1}. Depending on whether μ_M^* is considered or not considered, f_{M-1} or $\tilde{f}_M$ will be used to approximate the unknown underlying density f. It may happen that some summarizing quantities based on different approximations of f as f_{M-1} or $\tilde{f}_M$, remain unaltered as we illustrate in next few rows.

If g is a bounded function of X, $\tilde{f}_M$ and f_{M-1} lead to similar values, as Pinsker's inequality ([15], p. 390) and (11) yield,

$$\mid \mathbb{E}_{f_{M-1}}[g(X)] - \mathbb{E}_{\tilde{f}_M}[g(X)] \mid \leq \int_{\mathcal{S}_X} \mid g(x) \mid \cdot \mid f_{M-1}(x) - \tilde{f}_M(x) \mid dx$$

$$\leq \parallel g \parallel_\infty \sqrt{2(H_{f_{M-1}} - H_{\tilde{f}_M})} \leq \parallel g \parallel_\infty \sqrt{2 \cdot \varepsilon}$$

As a consequence, although we settle for a density constrained by fewer moments, and then conceptually in contrast with the MaxEnt spirit, nevertheless the results remain unaltered.

The matter runs similarly whether quantiles have to be calculated. They may be configured as expected values of proper bounded functions: indeed, for fixed x, $F(x) = \mathbb{E}[g(t)]$ with $g(t) = 1$ if $t \in [0,x]$ and $g(t) = 0$ if $t \in (x,\infty)$. Then, if $\tilde{F}_M$ and F_{M-1} denote the distribution functions corresponding to $\tilde{f}_M$ and f_{M-1}, respectively, we have in Stieltjes case (and mutatis mutandis equivalently holds for Hamburger case)

$$\mid F_{M-1}(x) - \tilde{F}_M(x) \mid \leq \int_0^x \mid f_{M-1}(t) - \tilde{f}_M(t) \mid dt$$

$$\leq \int_0^\infty \mid f_{M-1}(t) - \tilde{f}_M(t) \mid dt$$

$$\leq \sqrt{2(H_{f_{M-1}} - H_{\tilde{f}_M})} \leq \sqrt{2 \cdot \varepsilon}.$$

Again, although we settle for a density constrained by fewer moments, and this goes conceptually against the spirit of Jaynes, nevertheless the results concerning expected values of g remain unaltered. However, if g is an arbitrary unbounded function of X then the sequence of the above inequalities does not hold and the calculation of expected values of g could lead to different results, i.e., $\mathbb{E}_{f_{M-1}}[g(X)] \neq \mathbb{E}_{\tilde{f}_M}[g(X)]$.

In conclusion: if the maximum entropy distribution does not exist, being guided by the spirit of maximum entropy could always turn out to be the best choice.

Author Contributions: Conceptualization, P.L.N.I. and A.T.; Methodology, P.L.N.I.; Software, A.T.; Writing original draft, A.T.; Writing, review and editing, P.L.N.I. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

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

Appendix A. Existence of MaxEnt Distributions

Symmetric positive definite Hankel matrices $\Delta_M = \parallel \mu_{i+j} \parallel_{i,j=0}^N$ and $\Delta_{M,1} = \parallel \mu_{i+j+1} \parallel_{i,j=0}^N$, $M = 2N$ and $M = 1,2,...$ are recalled. Necessary condition for the existence of MaxEnt distributions is the positivity of determinants $\mid \Delta_M \mid > 0$ for Hamburger case and $\mid \Delta_M \mid > 0$, $\mid \Delta_{M,1} \mid > 0$ for Stieltjes case ([1], Theorem 1.2, p. 5 and Theorem 1.3, p. 6). The existence sufficient conditions in both cases are quoted below. Once the first moments $(\mu_1,...,\mu_M)$ are assigned

1. let us define the Mth moment space $\mathcal{C}_M$ as the convex hull of the curve $\{(x^1,...,x^M), x \in S\}$. $\mathcal{C}_M$ is convex and closed with boundary $\partial \mathcal{C}_M$. Then we mean $\mathcal{C}_M = \text{Int}\mathcal{C}_M \cup \partial \mathcal{C}_M$
2. let us call μ_{M+1}^- the value of μ_{M+1} such that $\mid \Delta_N \mid = 0$ or $\mid \Delta_{N,1} \mid = 0$, with $M = 2N$ or $M = 2N + 1$, so that the point $(\mu_1,...,\mu_M,\mu_{M+1}^-) \in \partial \mathcal{C}_{M+1}$
3. we recall too the moment space $\mathcal{E}_M$ relative to the MaxEnt densities, where $\mathcal{E}_M \subseteq \mathcal{C}_M$. Here we mean $\mathcal{E}_M = \text{Int}\mathcal{E}_M \cup \partial \mathcal{E}_M$. Take note $\partial \mathcal{E}_M$ includes both points $\in \partial \mathcal{C}_M$ and

points $\in \mathrm{Int}\mathcal{C}_M$, so that $\mathcal{E}_M = \mathrm{Int}\mathcal{E}_M \cup (\partial\mathcal{E}_M \cap \partial\mathcal{C}_M) \cup (\partial\mathcal{E}_M \cap \mathrm{Int}\mathcal{C}_M)$. Additional boundary points $(\partial\mathcal{E}_M \cap \mathrm{Int}\mathcal{C}_M)$ arise from the conditions

(i) $\lambda_M = 0$ (see Equation (2)) in Stieltjes case. For the special case $M = 2$, see [5], Theorem 2 or Example 2.

(ii) $\lambda_{M-1} = \lambda_M = 0$ (see Equation (2)) in symmetric Hamburger case.

Appendix A.1. Stieltjes Case

The moment set $(\mu_1^*, ..., \mu_M^*)$ is prescribed. The existence of f_M has been investigated ([7], Theorem 4 and [8], Theorem 3). Here the results are summarized in next two items.

1.   Let us suppose f_M exists with its $(M+1)$th moment μ_{M+1,f_M}. If $\mu_{M+1}^* \le \mu_{M+1,f_M}$, then f_{M+1} exists; conversely if $\mu_{M+1}^* > \mu_{M+1,f_M}$, then f_{M+1} does not exist. The existence of f_M is iteratively and numerically determined, starting from f_1 which exists.

2.   If f_M does not exist, both f_{M-1} and f_{M+1} exist for every $\mu_{M-1} > \mu_{M-1}^-$ and $\mu_{M+1} > \mu_{M+1}^-$ respectively;

3.   If $(\mu_1^*, \ldots, \mu_{M-1}^*)$ are fixed, whilst $\mu_M > \mu_M^-$ varies continuously, the entropy $H_{f_M}(\mu_M)$ of f_M is monotonic increasing function ([4], p. 59, Equation (2.73)).

Appendix A.2. Hamburger Case

The moment set $(\mu_1^*, ..., \mu_M^*)$ with M even, is considered. The existence of f_M has been investigated ([7], Theorem 4 and [9], Theorem 2). Here the results are summarized in next three items:

1.   In the non-symmetric case, the existence of f_M is guaranteed except for a special set of moments which is unknown a priori. So that, excluding the latter ones, the positivity of the Hankel determinants, which is necessary condition of the solvability, guarantees the existence of f_M.

2.   In the symmetric case (i.e., $\mu_{2j-1}^* = 0$) the condition of the solvability of f_M is analogous to Stieltjes case, being f_M symmetric. Thus the existence of f_M, $M \ge 4$, is iteratively and numerically determined, starting from f_2 which exists (being the Normal distribution); if f_M does not exist, both f_{M-2} and f_{M+2} exist for every $\mu_{M-2} > \mu_{M-2}^-$ and $\mu_{M+2} > \mu_{M+2}^-$ respectively;

3.   if $(\mu_1^*, \ldots, \mu_{M-1}^*)$ are fixed, whilst $\mu_M > \mu_M^-$ varies continuously, thanks to MaxEnt machinery, the entropy $H_{f_M}(\mu_M)$ of f_M is monotonic increasing function ([4], p. 59, Equation (2.73)). In MaxEnt setup this guarantees procedures and results valid for Stieltjes case can be equally extended to Hamburger one.

References

1.   Shohat J.A.; Tamarkin J.D. Math. Surveys and Monograph. In *The Problem of Moments*; American Mathematical Society: Providence, RI, USA, 1943; Volume I.

2.   Jaynes E.T. Information Theory and Statistical Mechanics. *Phys. Rev.* **1957**, *106*, 620–630. [CrossRef]

3.   Jaynes E.T. Prior Probabilities. *IEEE Trans. Syst. Sci. Cybern.* **1968**, *4*, 227–241. [CrossRef]

4.   Kesavan H.K.; Kapur J.N. *Entropy Optimization Principles with Applications*; Academic Press: London, UK, 1992.

5.   Dowson D.C.; Wragg A. Maximum-entropy distributions having prescribed first and second moments (Corresp.). *IEEE Trans. Inf. Theory* **1973**, *19*, 689–693. [CrossRef]

6.   Kociszewski, A. The existence conditions for maximum entropy distributions, having prescribed the first three moments. *J. Phys. A Math. Gen.* **1986**, *19*, L823–L827. [CrossRef]

7.   Junk M. Maximum entropy for reduced moment problems. *Math. Models Methods Appl. Sci.* **2000**, *10*, 1001–1025. [CrossRef]

8.   Tagliani A. Maximum entropy solutions and moment problem in unbounded domains. *Appl. Math. Lett.* **2003**, *16*, 519–524. [CrossRef]

9.   Tagliani A. Hamburger moment problem and maximum entropy: on the existence conditions. *Appl. Math. Comp.* **2014**, *231*, 111–116. [CrossRef]

10.   Milev M.; Tagliani A. Entropy convergence of finite moment approximations in Hamburger and Stieltjes problems. *Stat. Probab. Lett.* **2017**, *120*, 114–117. [CrossRef]

11.   Junk M. Domain of Definition of Levermore's Five-Moment System. *J. Stat. Phys.* **1998**, *93*, 1143–1167. [CrossRef]

12. Summy D.P.; Pullin D.I. On the Five-Moment Hamburger Maximum Entropy Reconstruction. *J. Stat. Phys.* **2018**, *172*, 854–879. [CrossRef]
13. Cover T.M.; Thomas J.A. *Elements of Information Theory*; Wiley-Interscience: Hoboken, NJ, USA, 2006.
14. Pólya G.; Szegö G. *Problems and Theorems in Analysis I*; Springer: Berlin, Germany, 1972.
15. Kullback S. *Information Theory and Statistics*; John Wiley: New York, NY, USA, 1959.