Stieltjes moment problem via fractional moments

Pierluigi Novi Inverardi a, Alberto Petri b, Giorgio Pontuale b, Aldo Tagliani a,*,1

a Faculty of Economics, Trento University, Trento 38100, Italy
b CNR, Istituto di Sistemi Complessi, Roma 00133, Italy

Abstract

Stieltjes moment problem is considered to recover a probability density function from the knowledge of its infinite sequence of ordinary moments. The approximate density is obtained through maximum entropy technique, under the constraint of few fractional moments. The latter are numerically obtained from the infinite sequence of ordinary moments and are chosen in such a way as to convey the maximum information content carried by the ordinary moments. As a consequence a model with few parameters is obtained and intrinsic numerical instability is avoided. It is proved that the approximate density is useful for calculating expected values and some other interesting probabilistic quantities.

© 2004 Elsevier Inc. All rights reserved.

Keywords: Entropy; Fractional moments; Hankel matrix; Laplace transform; Maximum Entropy; Ordinary moments

* Corresponding author.
E-mail address: aldo.tagliani@economia.unitn.it (A. Tagliani).

1 Work supported by COFIN MIUR 2002.
1. Introduction

There is no unique way to perform density reconstruction and many different methods have been proposed in literature. The techniques of density reconstruction are distinguished by the different functional choice of the approximant function \( f_{\text{appr}}(x) \) of the true but unknown density \( f(x) \), although many of them rely on exploitation of the moment sequence \( \{\mu_j\}_{j=1}^{\infty} \) to obtain the building blocks (of the unknown parameters) of \( f(x) \). The wide use of integer moments in many distribution reconstruction procedures relies on three main motivations:

(a) the sequence \( \{\mu_j\}_{j=1}^{\infty} \) characterizes the distribution in many cases of interest (but not always; the lognormal distribution, for example, is an exception);
(b) results are available on convergence of the reconstructed distribution via integer moments to the real but unknown distribution;
(c) the first few integer moments have a physical interpretation in terms of some descriptive characteristics of the distribution like center, dispersion, symmetry, tail behaviour.

Taking (a) into account, it is a well known fact that the sequence of integer moments carries all the information about the distribution; in other words, the moments share all the available information on the distribution; hence it may be that the moments of high order also contain a considerable part. Many density reconstruction procedures based on (a) and (b) involve only the moments up to order \( M \), where \( M \) is usually small for the sake of parsimony and to avoid numerical instability as a consequence of ill-conditioning of Hankel matrices involved. This implies that the moments of order bigger than \( M \) are neglected inside the reconstruction process and the quantity of information associated with them is irretrievably lost with consequences that we may easily imagine on the performance of the reconstruction. Furthermore, if the first few moments do not say much about the distribution, the situation is even worse with the reconstructed \( f_{\text{appr}}(x) \) being a bad approximant of \( f(x) \).

In the case of distributions with finite positive support (Hausdorff case) [1], the authors adopt a conceptually different approach to the density reconstruction. In fact they show that it is possible to express the fractional moment of order \( \alpha \), \( \alpha \in \mathbb{R}^+ \), as a weighted sum of all the integer moments and to squeeze the most relevant part of the information carried by the sequence of integer moments in a few fractional moments, losing in the “squeezing process” only a negligible quantity of information. The natural next step consists in utilizing these few fractional moments to recover the density \( f(x) \) via the Maximum Entropy (ME) approach. The proposed procedure fulfills the parsimony principle, avoids numerical instability problems, guarantees both the positiveness of the
approximant \( f_{\text{appr}}(x) \) and, last but not least, a good fit of \( f_{\text{appr}}(x) \) to \( f(x) \) as ME ensures the convergence of \( f_{\text{appr}}(x) \) to \( f(x) \). In fact, in Section 3, it is recalled that the ME reconstruction of \( f(x) \), \( f_{\text{appr}}(x) \), converges in entropy to \( f(x) \); this implies the convergence in directed divergence that entails the convergence in \( L_1 \)-norm that implies \( f_{\text{appr}}(x) \) converges in distribution to \( f(x) \). In other words, the reconstructed distribution \( f_{\text{appr}}(x) \) has the same information content as the true unknown distribution \( f(x) \), that is, expected values involving \( f(x) \) can be accurately evaluated through the corresponding expected values expressed in terms of \( f_{\text{appr}}(x) \).

In this paper the authors extend the above density reconstruction procedure to the case where the distribution has \([0, \infty) \) support (Stieltjes case), exploiting a result of [2] to express the fractional moments of order \( \alpha, \alpha > 0 \), through the Laplace transform of \( f(x) \), which can be obtained through the infinite sequence of integer moments. Once the (first \( M \)) fractional moments have been obtained, by means of the usual ME machinery, it will be possible reconstruct the unknown density \( f(x) \) with any order of precision, showing that the chain of convergences above listed is preserved also in Stieltjes case and the consequences on expected values reconstruction are still valid. Further it is important to note that, as in the Hausdorff case, the use of fractional moments permits ME to choose exponents which satisfy some restriction which reconstruction must fulfill; in this case, Lin’s theorem [3] guarantees that the fractional moments corresponding to the posed restrictions on the exponents still characterize the underlying distribution. For example, in the presence of an asymptotically constant Hazard rate, the freedom to choose fractional moments with exponents that satisfy the restriction to be posed for preserving such a behaviour, ensures an asymptotically constant reconstructed Hazard rate. Note that there is no analog result when integer moments are involved, since they provide an increasing Hazard rate as a consequence of the ME machinery.

2. Fractional moments from moments

Let \( X \) a continuous positive random variable with probability density function (pdf) \( f(x) \) and \( \{\mu_j\}_{j=0}^{\infty} \) its infinite sequence of ordinary moments which characterize \( X \). The following result [2] expresses the fractional moment \( \mathbb{E}(X^\alpha), \alpha > 0 \) by means of the Laplace transform

\[
L(s) = \int_0^\infty e^{-sx} f(x) \, dx
\]

(with \( s > -s_0 \) and \( -s_0 \leq 0 \) the abscissa of convergence) and the ordinary moments
with \( r \in (0, 1) \), \( M = 1, 2, \ldots \). It remains to calculate \( L(s) \) from \( \{\mu_j\}_{j=0}^\infty \) for real \( s \) values only.

As far as the abscissa of convergence \( -s_0 \) is concerned, which is given by the well-known formula

\[
\frac{1}{s_0} = \lim_{j \to \infty} \text{Sup} \left( \frac{\mu_j}{j!} \right)^{1/j},
\]

the following three cases have to be distinguished, since \( f(x) \) is a pdf:

(i) \( -s_0 = -\infty \);
(ii) \( -s_0 < 0 \) finite;
(iii) \( -s_0 = 0 \) (even in this case the moment problem could be determined).

Case (i) In such a case \( L(s) \) is yielded by the well-known series

\[
\hat{L}(s) = \sum_{j=0}^\infty (-1)^j \frac{\mu_j s^j}{j!}, \quad s \in \mathbb{R}
\]

whose radius of convergence \( R = +\infty \). Then (2.1) provides \( E(X^\alpha) \), \( \alpha > 0 \).

Case (ii) In such a case \( L(s) = \sum_{j=0}^\infty (-1)^j \frac{\mu_j s^j}{j!} \), \( -s_0 < s < s_0 \) holds. It remains to obtain \( L(s) \) from the sequence \( \{\mu_j\}_{j=0}^\infty \) when \( s \geq s_0 \). To this end we report some results obtained in [4].

\( L(s) \) is a completely monotonic function on \([0, \infty)\), i.e. \((-1)^m L^m(s) > 0\), \( \forall s > 0, m = 0, 1, \ldots \). Since \( L(s) \) may be written as \( L(s) = \int_0^\infty e^{-sx} dF(x) \), with \( dF(x) = f(x) dx \) and \( F(x) \) non-decreasing continuous bounded function. Then \( L(s) \) may be uniformly approximated on \([0, \infty)\) by the following exponential sum:

\[
Y_n(s) = \sum_{j=1}^n a_j e^{-\lambda_j s}
\]

having parameters satisfying the constraints \( 0 < \lambda_1 < \lambda_2 < \ldots < \lambda_n \), \( a_i > 0 \), \( i = 1, \ldots, n \). \( Y_n(s) \), which is also completely monotonic, interpolates \( L(s) \) at the \( 2n \) equally spaced points \( s_j \geq 0, j = 1, \ldots, 2n \), with \( L(s) - Y_n(s) > 0 \) on \((s_{2n}, +\infty)\). Such a choice of equispaced points allows us to calculate \( Y_n(s) \) through the Prony method. The Prony method is, in general, well suited to numerical computation only when \( n \) is fairly small, due to ill-conditioning of the Hankel matrices involved [5].
As far as the convergence of $Y_n(s)$ to $L(s)$ is concerned, $L(s)$ being completely monotonic on the extended interval $(-s_0, \infty)$ for some $s_0 > 0$ we have

$$\|L(s) - Y_n(s)\|_\infty = O(q^n), \quad n \to \infty, \quad 0 \leq s \leq s_{2n} \simeq s_0,$$

(2.3)

for some $q \in (0, 1)$. For $s > s_{2n}$, $Y_n(s)$ converges to $L(s)$ uniformly, even if quantitative results are not available.

As a consequence, for practical purposes, $Y_n(s)$ replaces $L(s)$ in (2.1).

The calculation of fractional moments, replacing $L(s)$ with $Y_n(s)$ in (2.1), is stable, as we shall now prove.

Let $L^\text{ex}(s)$ be the exact Laplace transform; $L^\text{appr}(s) = Y_n(s)$ be its approximation, such that $|L^\text{ex}(s) - L^\text{appr}(s)| < \epsilon \forall s \geq 0$; $E^\text{ex}(X^{r+M-1})$ and $E^\text{appr}(X^{r+M-1})$ be the exact and approximate fractional moments from (2.1) inserting $L^\text{ex}(s)$ and $L^\text{appr}(s)$ respectively. Then it holds

$$|E^\text{ex}(X^{r+M-1}) - E^\text{appr}(X^{r+M-1})| \leq \epsilon - S_{0}^{r-M+1} \frac{1}{r + M - 1} \prod_{j=0}^{M-1} (r + j).$$

(2.4)

**Proof.** From

$$E^\text{ex}(X^{r+M-1}) = (-1)^{M} \prod_{j=0}^{M-1} \frac{(r + j)}{\Gamma(1 - r)} \int_{0}^{\infty} s^{r-M} L^\text{ex}(s) - \sum_{j=0}^{M-1} (-1)^{j} \frac{\mu_{j} s^{j}}{j!} ds,$$

$$E^\text{appr}(X^{r+M-1}) = (-1)^{M} \prod_{j=0}^{M-1} \frac{(r + j)}{\Gamma(1 - r)} \int_{0}^{\infty} s^{r-M} L^\text{appr}(s) - \sum_{j=0}^{M-1} (-1)^{j} \frac{\mu_{j} s^{j}}{j!} ds,$$

and taking into account (2.3) we have

$$|E^\text{ex}(X^{r+M-1}) - E^\text{appr}(X^{r+M-1})| \leq \prod_{j=0}^{M-1} \frac{(r + j)}{\Gamma(1 - r)} \int_{0}^{\infty} s^{r-M} |L^\text{ex}(s) - L^\text{appr}(s)| ds$$

$$\simeq \frac{\prod_{j=0}^{M-1} (r + j)}{\Gamma(1 - r)} \int_{s_{0}}^{\infty} s^{r-M} |L^\text{ex}(s) - L^\text{appr}(s)| ds$$

$$\leq \epsilon - \frac{\prod_{j=0}^{M-1} (r + j)}{\Gamma(1 - r)} \int_{s_{0}}^{\infty} s^{r-M} ds = \epsilon \frac{S_{0}^{r-M+1}}{r + M - 1} \frac{1}{\Gamma(1 - r)} \prod_{j=0}^{M-1} (r + j). \quad \square$$

Case (iii) $-s_0 = 0$. This case remains an unsolved problem, since

$L(s) = \sum_{j=0}^{\infty} (-1)^{j} \frac{\mu_{j} s^{j}}{j!}$

has a radius of convergence $R = 0$. 
3. Recovering \( f(x) \) from fractional moments

Let be \( X \) a positive r.v. with density \( f(x) \), Shannon-entropy \( H[f] = -\int_0^\infty f(x) \ln f(x) \, dx \) and moments \( \{\mu_j\}_{j=0}^\infty \), from which positive fractional moments \( \mathbb{E}(X^z) \) may be obtained from (2.1).

From [6], we know that the Shannon-entropy maximizing density function \( f_M(x) \), which has the same \( M \) fractional moments \( \mathbb{E}(X^z) \), of \( f(x) \), \( j = 0, \ldots, M \), is

\[
f_M(x) = \exp \left( -\sum_{j=0}^M \lambda_j x^{z_j} \right). \tag{3.1}\]

Here \( (\lambda_0, \ldots, \lambda_M) \) are Lagrangian multipliers, which must be supplemented by the condition that first \( M \) fractional moments of \( f_M(x) \) coincide with \( \mathbb{E}(X^z) \), i.e.,

\[
\mathbb{E}(X^z) = \int_0^\infty x^z f_M(x) \, dx, \quad j = 0, \ldots, M, \quad z_0 = 1. \tag{3.2}\]

The Shannon-entropy \( H[f_M] \) of \( f_M(x) \) is given as

\[
H[f_M] = -\int_0^\infty f_M(x) \ln f_M(x) \, dx = \sum_{j=0}^M \lambda_j \mathbb{E}(X^{z_j}). \tag{3.3}\]

The choice of fractional moments relies upon two recent theorems concerning the existence of \( f(x) \) and the convergence of \( f_M(x) \) to \( f(x) \).

**Theorem 3.1** [3]. Let \( \{z_j\}_{j=0}^\infty \) be an infinite sequence, with \( z_0 = 0 \) and \( z_j \in (0, z^*) \) and \( \mathbb{E}(X^{z^*}) < +\infty \), then the sequence of expected values \( \{\mathbb{E}(X^{z_j})\}_{j=0}^\infty \) guarantees the existence of a unique \( f(x) \).

**Theorem 3.2** [7]. If \( \{z_j\}_{j=0}^M \) are equispaced, i.e., \( z_j = j \frac{z^*}{M}, j = 0, \ldots, M \), for some step \( z^* \), then the ME approximant \( f_M(x) \) converges in entropy to \( f(x) \), i.e.,

\[
\lim_{M \to \infty} H[f_M] = -\int_0^\infty f_M(x) \ln f_M(x) \, dx = \sum_{j=0}^M \lambda_j \mathbb{E}(X^{z_j}) = H[f] \tag{3.4}\]

\[
= -\int_0^\infty f(x) \ln f(x) \, dx. \tag{3.5}\]

3.1. The convergence

Given two probability densities \( f(x) \) and \( f_M(x) \), there are two well-known measures of the distance between \( f(x) \) and \( f_M(x) \). Namely,
the divergence measure $D(f, f_M) = \int_0^\infty f(x) \ln \frac{f(x)}{f_M(x)} \, dx$ and
the variation measure $V(f, f_M) = \int_0^\infty |f_M(x) - f(x)| \, dx$.

If $f(x)$ and $f_M(x)$ have the same fractional moments $\mathbb{E}(X^{a_j}), j = 1, \ldots, M$, then
\[ D(f, f_M) = H[f_M] - H[f] \]  
holds. In fact
\[
D(f, f_M) = \int_0^\infty f(x) \ln \frac{f(x)}{f_M(x)} \, dx = -H[f] + \sum_{j=0}^M \lambda_j \int_0^\infty x^{a_j} f(x) \, dx
\]
\[ = -H[f] + \sum_{j=0}^M \lambda_j \mathbb{E}(X^{a_j}) = H[f_M] - H[f]. \]

In the literature, several lower bounds for the divergence measure $D$, based on $V$, are available. We shall however use the following Pinsker inequality [8]:
\[ D \geq V^2 \cdot \frac{1}{2}. \]  
(3.5)
allows us to translate results from Information Theory (results involving $D$) to results in Probability Theory (results involving $V$) and vice versa.

Taking into account (3.4), (3.5) then entropy convergence, as in Theorem 3.2, entails convergence in directed divergence and in $L_1$-norm and then in distribution. The latter convergence is equivalent to
\[ \lim_{M \to \infty} \int_0^\infty g(x) f_M(x) \, dx = \int_0^\infty g(x) f(x) \, dx \]  
for each bounded function $g(x)$.

If $g(x)$ is a bounded function, taking into account (3.4) and (3.5) we have
\[ \| \mathbb{E}_f(g) - \mathbb{E}_{f_M}(g) \| = \left| \int_0^\infty g(x)(f(x) - f_M(x)) \, dx \right| \]
\[ \leq \|g\|_\infty \int_0^\infty |f(x) - f_M(x)| \, dx \leq \|g\|_\infty \sqrt{2(H[f_M] - H[f])}. \]  
(3.7)

### 3.2. On the choice $(\alpha_1, \ldots, \alpha_M)$

According to Theorem 3.2 we are able to formulate the choice criterion of $(\alpha_1, \ldots, \alpha_M)$, with $0 = \alpha_0 < \alpha_1 < \cdots < \alpha_M$. The optimal exponents are obtained as
\[ \{\alpha_j\}_{j=1}^M : H[f_M] = \text{minimum} \]  
(3.8)
or explicitly for practical purposes

$$\inf \min_{z_j} \left\{ \ln \int_0^\infty \exp \left( - \sum_{j=1}^M \lambda_j x_j \right) \, dx + \sum_{j=1}^M \lambda_j \mathbb{E}(X_j) \right\}. \quad (3.9)$$

The sequence is optimal in the sense that it accelerates the convergence of $H[f_M]$ to $H[f]$. Equivalently, it uses a minimum number of fractional moments to reach a prefixed (even if unknown) gap $H[f_M] - H[f]$. The choice (3.8) reflects a principle of parsimony, selecting a model with the lowest number of parameters and, in the meantime, avoiding the drawbacks of numerical instability which affect any inverse problem. The choice (3.8) agrees with Theorem 3.2, since the convergence in entropy is accelerated whenever equispaced nodes $x_j = \frac{j}{M}$ are replaced by optimal entropy nodes (3.8).

As we saw, the gap $H[f_M] - H[f]$ governs the error in the expected values calculation. As a consequence, the approximating pdf $f_M(x)$, with $(x_1, \ldots, x_M, \lambda_1, \ldots, \lambda_M)$ calculated through (3.9), is suitable for the accurate calculation of expected values, as required in Applied Probability.

Nevertheless, if our goal consists, for instance, in calculating the hazard rate function $\psi(x) \simeq \psi_M(x) = \frac{f_M(x)}{1 - \int_0^x f_M(u) \, du}$ then the optimal choice of $(x_1, \ldots, x_M)$ is closely related to a priori information about the asymptotic behaviour of $\psi(x)$. Indeed, if $\lim_{M \to \infty} \psi(x) = \text{constant}$, for instance, then the approximating $f_M(x)$ reflects properly such an asymptotic behaviour by imposing $x_M = 1$ in (3.9) as from elementary considerations. On the contrary, if $\lim_{M \to \infty} \psi(x) = +\infty$ then $x_M > 1$ has to be chosen. It is evident how the choice of integer moments $x_j = j$ in the ME approach will be fruitful only in the latter assumption concerning the asymptotic behaviour of $\psi(x)$. In other terms, by choosing integer moments, entropy convergence is guaranteed [9], so that expected values are accurately calculated. Nevertheless, an accurate calculation of other quantities is closely related to a priori information about such quantities.

Fractional moments are more flexible than integer moments, of course, and allow an accurate calculation of a large number of interesting probabilistic quantities.

4. Sensitivity analysis

We prove that the calculation of expected values is stable replacing $\mathbb{E}^{ex}(X^2)$ with $\mathbb{E}^{appr}(X^2)$, where $|\mathbb{E}^{ex}(X^2) - \mathbb{E}^{appr}(X^2)| \ll 1$. Lagrange multipliers have been calculated replacing $\mathbb{E}^{ex}(X^2)$ with $\mathbb{E}^{appr}(X^2)$. Let call $f_M(x, \lambda)$ the approximation of $f(x)$ obtained using $\mathbb{E}^{ex}(X^2)$, $f_M(x, \lambda + \Delta \lambda)$ the one obtained using $\mathbb{E}^{appr}(X^2)$. In both cases we assume the optimal exponents $(z_1, \ldots, z_M)$ are
coincident, being $E^{\text{appr}}(X^g) \simeq E^x(X^g)$. We want to estimate the difference $|E_{f_M(x, \lambda + \Delta \lambda)}(g) - E_{f_M(x, \lambda)}(g)|$ in the expected values calculation.

For notational convenience we set

$$
\mu_j = E(X^g) = \int_0^\infty x^g f_M(x) \, dx,
$$

$$
\mu_{ij} = (E(X^a), E(X^g)) = \int_0^\infty x^a x^g f_M(x) \, dx = \int_0^\infty x^{a+g} f_M(x) \, dx,
$$

$$
\Delta \mu_j = E^{\text{appr}}(X^g) - E(X^g).
$$

Let us now consider the measure $d\sigma(x) = f_M(x) \, dx$. Then $x^g \in L^2_{d\sigma}([0, \infty))$, where, as usual

$$
L^2_{d\sigma}([0, \infty)) = \left\{ x^g : \int_0^\infty x^{gj} \, d\sigma(x) = \int_0^\infty x^{gj} f_M(x) \, dx < +\infty \right\}.
$$

Thus

1. The matrix $G_M = [\mu_{ij}]$ is the Gram matrix.
2. The functions $\{x^{a1}, \ldots, x^{aM}\}$ being linearly independent on $[0, \infty)$ then $|G_M| > 0$.
3. The space $V^M$ of moments, given by the convex hull generated by the points $\{x^{a1}, \ldots, x^{aM}\}$, $x \in [0, \infty)$, has a non-empty interior.
4. If the sequence of prescribed moments $\{E(X_{a1}), \ldots, E(X_{aM})\}$ is an inner point of $V^M$ then there are uncountably many probability measure $d\sigma(x)$ having such prescribed moments, one of them being $d\sigma(x) = f_M(x) \, dx$.

If $\alpha_n$, with $0 \leq n \leq M$ is an arbitrary index, from

$$
\int_0^\infty x^n \exp \left( -\sum_{j=0}^M \lambda_j x^j \right) \, dx = E(X^n) = \mu_n, \quad n = 0, \ldots, M
$$

(4.1)

differentiating both sides with respect to $\mu_i = E(X^x)$, with $0 \leq i \leq M$, while the remaining expected values are fixed, we obtain

$$
G_M \cdot \begin{bmatrix} d\lambda_0 / d\mu_i \\ \vdots \\ d\lambda_M / d\mu_i \end{bmatrix} = -e_{i+1},
$$

(4.2)

where $e_{i+1}$ is the canonical unit column vector $\in \mathbb{R}^{M+1}$.

4.1. Relative error calculation

Let us consider the relative error

$$
\epsilon[f_M] = \frac{f_M(x, \lambda + \Delta \lambda) - f_M(x, \lambda)}{f_M(x, \lambda)}.
$$

(4.3)
We note that if all the expected values $\mu_i$ are changed to $\mu_i + \Delta \mu_i$ then the corresponding $\lambda_i$ becomes $\lambda_i + \Delta \lambda_i$, with $\Delta \lambda_i = \Delta \lambda_i(\Delta \mu_0, \ldots, \Delta \mu_M)$.

By Taylor expansion we have

$$
e[f_M] = \frac{f_M(x, \lambda + \Delta \lambda) - f_M(x, \lambda)}{f_M(x, \lambda)} = \exp \left( - \sum_{j=0}^{M} \Delta \lambda_j x^j \right) - 1$$

$$= \exp \left( - \sum_{j=0}^{M} x^j \sum_{i=0}^{M} \frac{\Delta \lambda_i}{\Delta \mu_i} \Delta \mu_i \right) - 1 \simeq - \sum_{j=0}^{M} x^j \sum_{i=0}^{M} \frac{\Delta \lambda_i}{\Delta \mu_i} \Delta \mu_i$$

$$= \sum_{j=0}^{M} x^j e_j^T G^{-1}_M \begin{bmatrix} \Delta \mu_0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} + \sum_{j=0}^{M} x^j x^j e_j^T G^{-1}_M \begin{bmatrix} \Delta \mu_1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} + \cdots + \sum_{j=0}^{M} x^j \sum_{i=0}^{M} \frac{\Delta \lambda_i}{\Delta \mu_i} \Delta \mu_i$$

$$= \left[ [x^{x_0}, 0, \ldots, 0] + [0, x^{x_1}, 0, \ldots, 0] + \cdots + [0, \ldots, 0, x^{x_M}] \right] G^{-1}_M \begin{bmatrix} \Delta \mu_0 \\ \vdots \\ \Delta \mu_M \end{bmatrix}$$

$$= \left[ x^{x_0}, x^{x_1}, \ldots, x^{x_M} \right] G^{-1}_M \begin{bmatrix} \Delta \mu_0 \\ \vdots \\ \Delta \mu_M \end{bmatrix}$$

$$= \left| G_M \right|^{-1} \begin{vmatrix} 0 & x^{x_0} & x^{x_1} & \cdots & x^{x_M} \\ \Delta \mu_0 & \mu_{0,0} & \mu_{0,1} & \cdots & \mu_{0,M} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \Delta \mu_M & \mu_{M,0} & \mu_{M,1} & \cdots & \mu_{M,M} \end{vmatrix}.$$  \hfill (4.4)

**4.2. Calculation of expected values**

(4.4) may be evaluated under the following assumption:

$$\frac{\Delta \mu_j}{\mu_j} = \frac{\mathbb{E}^{\text{app}}(X^{x_j}) - \mathbb{E}^{\text{ex}}(X^{x_j})}{\mathbb{E}^{\text{ex}}(X^{x_j})} = \Delta \mu = \text{constant} \quad \forall j. \hfill (4.5)$$
Replacing (4.5) into (4.4) and recalling that $D_{l,j,0} = D_{l,j}$, we have

$$
\epsilon[f_M] = \frac{\Delta \mu}{|G_M|} \cdot |G_M| = \Delta \mu.
$$

(4.6)

As far as the expected value $E[g(x)]$ is concerned we have

$$
\|E_{f_M(x, \lambda + \Delta \lambda)}(g) - E_{f_M(x, \lambda)}(g)\| = \left| \int_0^\infty [f_M(x, \lambda + \Delta \lambda) - f_M(x, \lambda)]g(x)\,dx \right| \\
\leq \|g\|_\infty \int_0^\infty \frac{|f_M(x, \lambda + \Delta \lambda) - f_M(x, \lambda)|}{f_M(x, \lambda)}f_M(x, \lambda)\,dx \\
= \|g\|_\infty \Delta \mu \int_0^\infty f_M(x, \lambda)\,dx = \|g\|_\infty \Delta \mu.
$$

(4.7)

Hence, under hypothesis (4.5) the calculation of expected values is stable under small fluctuations of $E_{\text{ex}}(X^x)$.

5. Numerical results

Before illustrating the recovery of densities from fractional moments, some preliminary tests are in order. The first concerns the comparison between $L(s)$ and $Y_n(s)$. The Gamma distribution with density $f(x) = xe^{-x}$ is chosen, with $s_0 = 1$, $\mu_j = \Gamma(2 + j)$, $j \in \mathbb{N}$, so that $L(s) = \sum_{j=0}^\infty (-1)^j \frac{\mu_j^s}{j!}$, $0 < s < 1$ holds. $2n = 16$ equispaced interpolation nodes $s_j$ with $0 \leq s_j < 0.95$ (16 is the maximum allowed number due to conditioning of Hankel matrices involved) are chosen. Through the Prony method $Y_n(s)$ is obtained. The difference $L(s) - Y_n(s)$ is reported in Fig. 1.

The second test concerns (2.4), the comparison between exact and approximate fractional moments. The latter obtained through (2.1), with $L(s)$ replaced by $Y_n(s)$ given by (2.2). The chosen pdf coincides with the previous Gamma density having $E(X^x) = \Gamma(2 + x)$.

In Fig. 2 we report (a) $E_{\text{ex}}(X^{r+M-1})$ and (b) the relative error $\frac{E_{\text{ex}}(X^{r+M-1}) - E_{\text{app}}(X^{r+M-1})}{E_{\text{ex}}(X^{r+M-1})}$. Such a relative error is quite small over the entire range involved in the minimization problem (3.9), except for some $r \approx 0$. As a consequence, in the following tests we may use indifferently $E_{\text{ex}}(X^x)$ or $E_{\text{app}}(X^x)$ for Lagrange multipliers $\lambda_j$ estimation in (3.9), according to sensitivity analysis of Section 4.

Example 1. The same previous Gamma density $f(x) = xe^{-x}$, with $E(X^x) = \Gamma(2 + x)$ and $H[f] \simeq 1.5767449292$ is used. The ME density is estimated according to (3.9) for an increasing number of optimal fractional moments. In Table 1 the entropy difference $H[f_M] - H[f]$ for an increasing number $M$ of fractional moments is reported. The entropy convergence is fast.
or, equivalently, 2 fractional moments capture approximately all the information content of the distribution. This might be a direct consequence of the fact that $f(x)$ has two characterizing moments $E(X)$ and $E(\ln X)$. Then $E(X^{\alpha_1})$ and $E(X^{\alpha_2})$, with $\alpha_1 \approx 0.04007, \alpha_2 \approx 0.94885$ through (3.9), might represent an accurate approximation of the characterizing moments $E(X)$ and $E(\ln X)$. However, the small difference $H[f_2] - H[f] \div 10^{-5}$, compared with $H[f_1] - H[f] \div 10^{-1}$, shown in Table 1, suggests the existence of two characterizing moments, emphasizing the flexibility of fractional moments in emulating the characterizing moments.

Table 1

<table>
<thead>
<tr>
<th>$M$</th>
<th>$H[f_M] - H[f]$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.4087E-1</td>
</tr>
<tr>
<td>2</td>
<td>0.6750E-5</td>
</tr>
<tr>
<td>3</td>
<td>0.4538E-5</td>
</tr>
</tbody>
</table>
Led by a mere curiosity, since only the difference $H[f_M] - H[f]$ may be estimated, in Fig. 3 we illustrate a graphical comparison of $f(x)$ and $f_M(x)$ when $M = 2$.

**Example 2.** The density $f(x) = 2xe^{-x^2}$ is chosen, with $\mu_j = \Gamma(1 + j/2), j \in \mathbb{N}$ and, in general, $E(X^2) = \Gamma(1 + \alpha/2)$ and $H[f] \approx 0.59544601848$. Here $\hat{L}(s) = \sum_{j=0}^{\infty} (-1)^j \frac{\mu_{2j}}{j!}$ has radius of convergence $R = +\infty$. $Y_n(s)$, with $2n = 16$ is obtained by choosing equispaced interpolation nodes $s_j \in [0, 5]$ (different choices lead to similar results). The comparison of $Y_n(s)$ with $\hat{L}(s)$ is accurate, and the comparison of $E^{ex}(X^2)$ with $E^{apr}(X^2)$ reflects the features reported in Fig. 2. The same quantities as in Example 1 are reported. Here $f(x)$ has two characterizing moments $E(x^2)$ and $E(\ln X)$. Also in this example the gap $H[f_2] - H[f] \div 10^{-3}$, in Table 2, may suggests the existence of two characterizing moments. The entropy convergence is fast or, equivalently, from Table 2, 3–4 fractional moments capture approximately all the information content of the distribution. Fig. 4 shows a graphical comparison of $f(x)$ and $f_M(x)$ when $M = 4$.

From both Examples 1 and 2 we may draw the conclusion that fractional moments are powerful tools in recovering a pdf when the available information consists in the infinite sequence of ordinary moments. Few fractional moments

<table>
<thead>
<tr>
<th>$M$</th>
<th>$H[f_M] - H[f]$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.7612E–1</td>
</tr>
<tr>
<td>2</td>
<td>0.2236E–3</td>
</tr>
<tr>
<td>3</td>
<td>0.3769E–5</td>
</tr>
<tr>
<td>4</td>
<td>0.6694E–6</td>
</tr>
</tbody>
</table>
contain approximately the same information as that carried by the infinite sequence of ordinary moments. Expected values are accurately calculated and instability problems are avoided.

References