ENTROPY AND DENSITY APPROXIMATION FROM LAPLACE TRANSFORMS

Henryk Gzyl[1], Pierluigi Novi Inverardi[2], Aldo Tagliani[2,*]
[1]Centro de Finanzas IESA, DF, CARACAS (Venezuela)
E-mail: henryk.gzyl@iesa.edu.ve
[2] Department of Economy and Management
University of Trento - 38100 TRENTO (Italy)
E-mail: pierluigi.noviinverardi@unitn.it, aldo.tagliani@unitn.it

Abstract

How much information does the Laplace Transforms on the real line carry about an unknown, absolutely continuous distribution? If we measure that information by the Boltzmann-Gibbs-Shannon entropy, the original question becomes: How to determine the information in a probability density from the given values of its Laplace transform. We prove that a reliable evaluation both of the entropy and density can be done by exploiting some theoretical results about entropy convergence, that involve only finitely many real values of the Laplace transform, without having to invert the Laplace transform.

We provide a bound for the approximation error of in terms of the Kullback-Leibler distance and a method for calculating the density to arbitrary accuracy.

Keywords: Entropy convergence, Fractional moments, Kullback-Leibler distance, Laplace Transform, Maximum Entropy.

*Corresponding author

1 Introduction

Let X be a positive continuous random variable having an (unknown) probability density function (pdf) $f_X(x)$ with respect to the Lebesgue measure on $[0,\infty)$, and let us suppose that its Laplace Transform $L(\alpha) = E[e^{-\alpha X}] = \int_0^\infty e^{-\alpha x} f_X(x)dx$, is known.

Since the Laplace transform is not a continuously invertible mapping, (i.e., the inverse Laplace transform exists, but it is not continuous), the inverse problem consisting of numerically determining $f_X(x)$ is be set with difficulties. This is an important consideration if for example, $L(\alpha)$ is to be estimated numerically. The lack of continuity may cause the errors in the determination of the Laplace transform to be amplified in the inversion process.

But when we actually do not need to know the exact or true $f_X(x)$ but only some quantities related to it, like perhaps, expected values of some given functions of X, or as in many applications in statistical information theory, we may only want to estimate the entropy of f_X. In such case, one would not attempt to invert the Laplace transform, but to estimate the quantity of interest directly from the available data, which may consist of the values of the Laplace transform at finitely many points.

We will propose a way to use directly real values of Laplace Transform to estimate the Boltzmann-Gibbs-Shannon entropy (entropy for short) $H[f_X] = -\int_0^\infty f_X(x)\ln f_X(x)dx$ without having to determine f_X exactly. This generates an interesting mathematical problem, namely, to determine the conditions upon which the entropy of the estimates based on partial data converge to the entropy of (the unknown) f_X. See the work of Piera and Parada (2009) and of Silva and Parada (2012) for interesting results and further references to this problem.

Our task is similar to that previously described in the literature for the case in which X has support $[0,1]$ and a few of its integer moments are known. As far as the reconstruction of the density goes, Gavriliadis and Athanassoulis (2009) and Gavriliadis (2008) obtain some results about the separation of the main mass interval, the tail interval and the position of the mode. In some recent papers Mnatsakanov (2008a,b) provides a procedure to recover a probability density function f_X (and the associated distribution function F_X) directly from a finite but large number of integer moments and he estimates the nature of the convergence of the approximants to the true functions.

When the available information consists of integer moments, Tagliani (2002) provides an upper bound of $H[f_X]$ directly in terms of such moments, and Novi-Inverardi et al. (2012) estimate the entropy $H[f_X]$ by solving a linear system.

All the previous results come from the fact that, when X has bounded support, the underlying moment problem is determinate and the information content about a distribution is spread over the infinite sequence of its moments. Whereas, when X has unbounded support, the underlying moment problem may be determinate or indeterminate, and integer moments may prove to be unsuitable to estimate the above mentioned quantities. In this case, we may rely upon Laplace Transform rather than on the integer moments. And to transform the unbounded domain onto a bounded domain, we shall replace X by an appropriate Y so that the available information provided by the Laplace transform of X becomes information provided by fractional moments of Y.

So, to be specific, let f_X have support $[0,+\infty)$ and consider the auxiliary random variable $Y = e^{-X}$, with support $[0,1]$. As said, the Laplace transform $L(\alpha)$ of f_X can be thought of as the moment curve of Y, that is

$$L(\alpha) = E[e^{-\alpha X}] = E[Y^{\alpha}] \equiv \mu_Y(\alpha) = \int_0^1 y^{\alpha} f_Y(y)dy. \tag{1.1}$$

Certainly, once the probability density f_Y is determined, $f_X(x) = e^{-x} f_Y(e^{-x})$ is obtained by a simple change of variables. Thus the question becomes: Can we use Laplace transform based techniques to numerically approximate f_Y from the knowledge of a finite collection of M real values $L(\alpha_j \geq 0)$, $j = 0, ..., M$? The answer is yes under a restrictive hypothesis on f_X: As we shall see in what follows, the entropies $H[f_X]$ and $H[f_Y]$ of X and Y respectively are related by relationship $H[f_X] = H[f_Y] - L'(0)$, that requires that f_X has a finite first integer moment, i.e., $\mu_1(f_X) = -L'(0) < \infty$. The latter condition is direct consequence of introducing the auxiliary random variable $Y = e^{-X}$. We shall furthermore see how the approximants to f_Y may be used to estimate the entropy of the unknown f_Y, or that of f_X. The latter task requires both $H[f_X]$ and $H[f_Y]$ are finite, from which $\mu_1(f_X) = -L'(0)$ finite too.

The remainder of the paper is organized as follows. In the next section we briefly recall the result of applying the standard entropy method to estimate the density from a few values of its Laplace transform. In section three we provide bound and estimate for the entropy that involves only finitely many real values of the Laplace transform. In section four we present an efficient method to carry out the estimation of f_Y from a few values of its Laplace transform, as well as to find the optimal model approximates f_X with an prefixed error in terms of Kullback-Leibler distance. We devote section five to a numerical examples and then we round up with some concluding remarks.

2 The method of maximum entropy

The following problem is rather common in a variety of fields. Consider a random variable taking values in $[0,1]$, and suppose all that is known about it is the value of a few of its "generalized" moments $(\mu_0, ..., \mu_M)$, given by

$$\mu_j = E[a_j(Y)] = \int_0^1 a_j(y) f_Y(y) dy, \quad j = 0, ..., M \tag{2.1}$$

where the $a_j : [0,1] \to \mathbb{R}$ are given measurable functions, such that $a_0 \equiv 1$ and $\mu_0 = 1$ is the normalization condition upon f_Y. For example, we may consider $a_j(y) = y^j$ and be in the realm of the standard moments problem, or $a_j(y) = y^{\alpha_j}$ and be in the realm of the fractional moments problem, or the can be trigonometric functions $a_j(y) = e^{2i\pi j}$ and we shall have a trigonometric moment problem in our hands.

As the set of probability densities on $[0,1]$ satisfying (2.1) is a convex set in $L_1([0,1], dy)$, a simple way of picking a point from that set is by maximizing a concave function defined over it. This is a standard variational method procedure, known as the Maximum Entropy (MaxEnt) Principle (Jaynes, 1957).

It consists of maximizing the entropy functional defined over the class of probability densities by

$$H[f_Y] = -\int_0^1 f_Y(y) \ln f_Y(y) dy \tag{2.2}$$

subject to (2.1) as constraints. The procedure is rather standard. For a given set $(\mu_0, ..., \mu_M)$ of moments, when the solution f_M exists, it is an approximant to f_Y given by

$$f_M(y) = \exp\Big(-\sum_{j=0}^{M} \lambda_j a_j(y) \Big) \tag{2.3}$$

where $(\lambda_0, ..., \lambda_M)$ are the Lagrange's multipliers, that appear as part of the minimization procedure as solutions to a dual problem. Actually, the $\{\lambda_j,\ j = 1, ..., M\}$ are obtained minimizing the dual entropy function

$$H(\lambda, \mu) = \ln Z(\lambda) + \langle \lambda, \mu \rangle. \tag{2.4}$$

where

$$Z(\lambda) = \int_0^1 e^{-\sum_{i=1}^{M} \lambda_i a_i(y)} dy$$

At the minimum, $e^{-\lambda_0} = (Z(\lambda))^{-1}$, from which

$$H[f_M] = \lambda_0 + \langle \lambda, \mu \rangle \tag{2.5}$$

3 Bound and estimate for the entropy from integer moments

3.1 An entropy bound

Here we obtain an upper bound on the entropy. In the case of a probability density on a bounded interval, this measures how far (in entropy) the density is from a uniform density. For a density on the positive half-line, it measure how far it is from an exponential density. So, we begin by considering the uniform density $f^0(y) = 1$ on $[0, 1]$, the moments of which are given by

$$\mu_j^0 = \int_0^1 a_j(y)dy, \quad j = 1, ..., M. \tag{3.1}$$

In this case, it is easy to see that, in the notation of the previous section, $\lambda_j = 0$ minimizes the corresponding version of (2.4), and that the corresponding entropy is $H[f^0] = H(\mu^0) = 0$. It takes a few simple calculations, using (3.1) as starting point to notice that, as the first two terms of the Taylor expansion of $H(\mu)$ about μ^0 vanish

$$H[f_M] = \frac{1}{2}\sum_{i,j=0}^{M}(\mu_j - \mu_j^0)\frac{\partial^2 H[f_M](\nu)}{\partial\mu_i\partial\mu_j}(\mu_i - \mu_i^0).$$

Above, the Hessian matrix is evualuated at ν whose entries $\nu_j \in [\min\{\mu_j, \mu_j^0\}, \max\{\mu_j, \mu_j^0\}]$, $j = 0, ..., M$. Let's consider MaxEnt density f_M^ν constrained by $\{\nu_j\}_{j=0}^M$ and we consider $\{\nu_j(f_M^\nu)\}_{j=M+1}^{2M}$ its higher moments. Collecting together previous results (see Tagliani (2002) formulas (3.3)-(3.7) for details, as well Kesavan-Kapur (2002), pag. 46-48, since
$H[f_M] = \lambda_0 + \langle\lambda, \mu\rangle$;
$\frac{\partial H[f_M]}{\partial\mu_k} = \lambda_k - \delta_{0k}$ (δ-Kronecker);
we can compute the Hessian matrix explicitly as follows. $\{\frac{\partial^2 H[f_M]}{\partial\mu_j\partial\mu_k}\}_{j,k=0}^M = \{\frac{\partial\lambda_k}{\partial\mu_j}\}_{j,k=0}^M = -\Delta_{2M}^{-1}$, where Δ_{2M} is $(M+1)$-th order Hankel matrix generated by the vector $\nu = \{\nu_j(f_M^\nu)\}_{j=0}^{2M}$. Therefore

$$-\frac{2H[f_M]}{\| \mu - \mu^0 \|_2^2} = -\frac{\sum_{i,j=0}^M(\mu_j - \mu_j^0)\frac{\partial^2 H[f_M](\nu)}{\partial\mu_i\partial\mu_j}(\mu_i - \mu_i^0)}{\| \mu - \mu^0 \|_2^2} =$$
$$= \frac{(\mu - \mu^0)\Delta_{2M}^{-1}(\nu)(\mu - \mu^0)^T}{\| \mu - \mu^0 \|_2^2} > \lambda_{min}(\Delta_{2M}^{-1}(\nu)) = \frac{1}{\lambda_{max}(\Delta_{2M}(\nu))} >$$
$$> \frac{1}{\| \Delta_{2M}(\nu) \|_1} = \frac{1}{\sum_{j=0}^M \nu_j} \geq \frac{1}{\sum_{j=0}^M \max\{\mu_j, \mu_j^0\}} = \frac{1}{\| \max\{\mu, \mu^0\} \|_1}$$

where λ_{min} and λ_{max} denote smallest and greatest eigenvalues. Collecting together, we find bound similar to the one provided in (Tagliani, (2002) eq.(3.8)), namely

$$H[f_M] \leq -\frac{\| \mu - \mu^0 \|_2^2}{2 \| \max\{\mu, \mu^0\} \|_1}.$$

Since clearly $H[f_Y] \leq H[f_M]$, $\forall M \geq 0$ we obtain an upper bound for $H[f_Y]$ for any density f_Y having moments, given by $\mu = \{\mu_j(f_Y)\}_{j=0}^M$

$$H[f_Y] \leq \inf_M \Big[-\frac{\| \mu - \mu^0 \|_2^2}{2 \| \max\{\mu, \mu^0\} \|_1} \Big]. \tag{3.2}$$

The previous computations, coupled with the simple relationship $H[f_X] = H[f_Y] - L'(0)$, provide us with the following estimate of the entropy of f_X in terms of Laplace transform values $L(\alpha_j = j)$.

Proposition 3.1 *Let f_X be a pdf having Laplace transform $L(\alpha)$ and finite first moment $\mu_1(f_X) = -L'(0)$. Then its entropy is bounded as follows*

$$H[f_X] = H[f_Y] - L'(0) \leq \inf_M \Big[-\frac{|| \mu - \mu^0 ||_2^2}{2 || \max\{\mu, \mu^0\} ||_1} \Big] - L'(0) \tag{3.3}$$

3.2 An approximate computation of the entropy

Now we present a procedure to compute an approximate value for $H[f_Y]$ using integer moments. We combine previous results from [Novi Inverardi et al., 2012] concerning entropy estimate $H[f_Y]$ from its integer moments (equivalently from Laplace transform values $L(\alpha_j = j)$). If f_M denotes the MaxEnt approximation to f_Y constrained by first $M+1$ integer moments $\{\mu_j = L(\alpha_j = j)\}_0^M$, we know that the MaxEnt approximations converge in entropy, as M increases, according to

Theorem 3.1 [Tagliani, 1999, Th. 4.1] *Let $\{\mu_j\}_0^\infty$ the sequence of moments of an unknown density f_Y with finite entropy $H[f_Y]$, MaxEnt approximate f_M converge in entropy to f_Y, i.e.*

$$\lim_{M \to \infty} H[f_M] = H[f_Y]. \tag{3.4}$$

(Entropy-convergence had been proved for a bounded function f_Y [Forte et al. 1989, Th. 2]. In the generalized Hausdorff moment problem [Borwein-Lewis, 1991] the authors proved how some properties of Shannon entropy, i.e. strict convexity, essential smoothness and coercivity lead to various distinct types of convergence. In particular, if f_Y is twice continuously differentiable and strictly positive, the sequences f_M converge uniformly to f_Y [Borwein-Lewis, 1991, Corollary 5.3]). Denote the Kullback-Leibler distance between f_Y and f_M

$$K(f_Y, f_M) = \int_0^1 f_Y(y) \ln \frac{f_Y(y)}{f_M(y)} dy \tag{3.5}$$

since f_Y and f_M share same first $M+1$ integer moments $\{\mu_j\}_{j=0}^{M}$, it follows

$$\int_0^1 f_Y(y)\ln\frac{f_Y(y)}{f_M(y)}dy = H[f_M]-H[f_Y]. \tag{3.6}$$

Combining entropy convergence theorem with (3.6) it follows the entropy convergence result

$$\lim_{M\to\infty} H[f_M]-H[f_Y] = \lim_{M\to\infty}\int_0^1 f_Y(y)\ln\frac{f_Y(y)}{f_M(y)}dy = 0 \tag{3.7}$$

and then f_M converges to f_Y a.e. Here we recall the inequality (Csiszár and Shields (2004) or Kullback and Leibler (1951)), which asserts that

$$\int_0^1 \mid f_M(y)-f_Y(y)\mid dy \le \sqrt{2(H[f_M]-H[f_Y])} \tag{3.8}$$

It follows that f_M converges to f_Y in L_1-norm. Calling $\Delta\mu_j = \mu_j(f_M)-\mu_j(f_Y)$ and taking into account (3.8) , one has

$$\begin{aligned} \mid \Delta\mu_j \mid=\mid \mu_j(f_M)-\mu_j(f_Y)\mid=\mid \int_0^1 y^j\Big(f_M(y)-f_Y(y)\Big)dy \mid\le \int_0^1 y^j \mid f_M(y)-f_Y(y)\mid dy \le \\ \le \int_0^1 \mid f_M(y)-f_Y(y)\mid dy \le \sqrt{2(H[f_M]-H[f_Y])}\to 0. \end{aligned} \tag{3.9}$$

$\mu_j(f_M)$ closely approximate $\mu_j(f_Y)$, $j=M+1,...,2M$, as M increases. As a consequence, we may identify $\mu_j(f_M)=\int_0^1 y^j f_M(y)dy$ with $\mu_j(f_Y)=\int_0^1 y^j f_Y(y)dy$ for each fixed j and sufficiently large $> M$. Collecting together just above results the following approximate procedure to calculating Lagrange multipliers may be formulated.

Integrating by parts $\int_0^1 y^j f_M(y)dy = \mu_j$, $j=1,...,M$ the relationship relating $(\lambda_1,...,\lambda_M)$ with $\{\mu_j=\mu_j(f_M)\}_{j=1}^{2M}$ is obtained

$$\sum_{k=1}^{M} k\lambda_k(\mu_k-\mu_{k+j}) = 1-(j+1)\mu_j, \qquad j=1,...,M \tag{3.10}$$

For sufficiently large M, thanks to (3.9), the moments $\{\mu_j(f_M)\}_{j=M+1}^{2M}$ are identified with moments $\{\mu_j(f_Y)\}_{j=M+1}^{2M}$ (which are known). Then (3.10) may be considered a linear system with unknown $(\lambda_1,...,\lambda_M)$ which admits an unique solution being an identity relating $(\lambda_1,...,\lambda_M)$ with $\{\mu_j(f_M)\}_{j=1}^{2M}$. Here comes the ansatz: We shall suppose that the solution to (3.10) coincides with that obtained by minimizing (2.4). This brings this section close to experimental mathematics. The necessary analysis to compute the error in this approximation is hard, and the verification comes in a posteriori as the numerical results based on it make good sense. And finally, to compute λ_0, note that asf_M integrates to one, $\lambda_0=\ln\int_0^1\exp\Big(-\sum_{j=1}^M \lambda_j y^j\Big)dy$. All this taken together, allows us to obtain $H[f_M] = \lambda_0 + \langle \lambda, \mu \rangle$ solving a linear system solely without solving the underlying moment problem. We gather these comments as

Proposition 3.2 (Plausible) *Let f_X be a pdf having Laplace transform $L(\alpha)$ and having a first moment finite $\mu_1(f_X) = -L'(0)$. For high M values the entropy $H[f_X]$ is given by*

$$H[f_X] = H[f_Y] - L'(0) \simeq H[f_M] - L'(0)$$

where $H[f_M]$ is obtained by solving a linear system solely.

This, for high M values, it allows us to identify $H[f_Y]$ with $H[f_M]$ (numerical evidence in Figures 1,3,5 below).

The remainder of this section is to make the ansatz plausible.

The above procedure of calculating $\lambda' s$ through a linear system is meaningful when M takes large values and the convergence of $H[f_M]$ to $H[f_Y]$ is fast.

To examine the crucial numerical issue, note that the matrix in (3.10) becomes severly ill-conditioned and we explain next. Let us consider (3.10) and call Δ_{2M} the corresponding matrix

$$\begin{aligned} \Delta_{2M} &= \begin{bmatrix} \mu_1 - \mu_2 & \cdots & \mu_M - \mu_{M+1} \\ \vdots & \vdots & \vdots \\ \mu_1 - \mu_{M+1} & \cdots & \mu_M - \mu_{2M} \end{bmatrix} \\ &= \begin{bmatrix} 1 & \cdots & 1 \\ \vdots & \vdots & \vdots \\ 1 & \cdots & 1 \end{bmatrix} \mathrm{diag}(\mu_1, ..., \mu_M) - \begin{bmatrix} \mu_2 & \cdots & \mu_{M+1} \\ \vdots & \vdots & \vdots \\ \mu_{M+1} & \cdots & \mu_{2M} \end{bmatrix} \\ &= \Delta_{2M}^{(1)} - \Delta_{2M}^{(2)} \end{aligned} \tag{3.11}$$

that is, we split Δ_{2M} into two matrices, $\Delta_{2M}^{(1)}$ and $\Delta_{2M}^{(2)}$, where the last one is an Hankel matrix. Now Δ_{2M} is a moments matrix relating the Lagrange multipliers $\boldsymbol{\lambda}$ wto the moments $\{\mu_j(f_M)\}_{j=1}^{2M}$, which, as expected, is ill-conditioned. A previous result [Fasino, (1995), Th. 3.2] states Hankel matrices $\Delta_{2M}^{(2)}$ generated by moments of a strictly positive weight function with support on the interval $[0, 1]$, is asymptotically conditioned in the same way as Hilbert matrices H_M of the same size, i.e.

$$\lim_{M\to\infty} \Big(\mathrm{Cond}(\Delta_{2M}^{(2)})\Big)^{1/M} = \lim_{M\to\infty} \Big(\mathrm{Cond}(H_M)\Big)^{1/M} \simeq e^{3.525} \tag{3.12}$$

Then Hilbert matrix H_M is a good preconditioner for this class of matrices ([Fasino, (1995), Th. 3.3), as the conditioning of the preconditioned matrix grows at most linearly with M. Now let's conjecture ill-conditioning of Δ_{2M} is mainly due to $\Delta_{2M}^{(2)}$ and then $\Delta_{2M}^{(2)}$ and Δ_{2M} have condition

number comparable (for a graphical comparison see Novi et al. (2012)). As a final result the Hilbert matrix H_M is chosen (by rule of thumb) as preconditioner for Δ_{2M}. Thus we can indirectly solve (3.10), written as $\Delta_{2M}\boldsymbol{\lambda} = \boldsymbol{b}$, by solving the preconditioned system $H_M^{-1}\Delta_{2M}\boldsymbol{\lambda} = H_M^{-1}\boldsymbol{b}$ or, equivalently $C_M\Delta_{2M}C'_M\boldsymbol{y} = C_M\boldsymbol{b}$ and $\boldsymbol{\lambda} = C'_M\boldsymbol{y}$. Here $C'_M C_M = H_M^{-1}$ and C' stands for the transpose of C; $C_M = [c_{ij}]$ is the inverse of the lower triangular Cholesky factor of the Hilbert matrix H_M with entries $c_{ij} = (2i+1)^{1/2}(-1)^j\binom{i}{j}\binom{i+j}{j}$, $i, j = 0, 1, ..., M-1$ (see Talenti, (1987) for details). By solving the above preconditioned system $C_M\Delta_{2M}C'_M\boldsymbol{y} = C_M\boldsymbol{b}$ from which $\boldsymbol{\lambda} = C'_M\boldsymbol{y}$, higher M values may be taken into account with consequent improved estimate of $H[f_Y]$.
As a final remark, the computation of products like $C_M\Delta_{2M}C'_M$ and $C_M\boldsymbol{b}$ is very unstable due to presence of large entries, both positive and negative, in the matrix C_M. Only for moderate M, say, up to 12-13, this computation is feasible in double precision arithmetics, since for larger dimensions the error in last entries of the computed matrix and vector grow exponentially, so that high precision arithmetics is needed.

4 Density and entropy estimation from fractional moments

This section is devoted to the two related problems: one hand, there is the need of the estimation of the entropy of a pdf f_X when the only information available consists of the values of the Laplace transform along the positive real axis, and on the other, the estimation of a density from a few given values of its Laplace transform. But the first issue to take care of is to decide whether a pdf f_X is determined by an appropriate collection of his values on the real axis. That this is so, is explained by the following Theorem by Lin (1992), the proof of which relies on the fact that an analytic function is determined by its values on a countable set having an accumulation point in the domain of analyticity.

Theorem 4.1 [Lin, 1992] *Let F_Y be the distribution function of a positive random variable Y. Let $\{\alpha_n\}_{n\geq 0}$ be a sequence of positive and distinct numbers in $(0, a)$ for some $a > 0$, satisfying $lim_{n\to\infty}\alpha_n = \alpha_0 < a$. If $E[Y^a] < \infty$, the sequence of moments $E[Y^{\alpha_n}]$ characterizes F_Y.*

As the Laplace transform $L(\alpha)$ of f_X can be thought of as the moment curve $\mu_Y(\alpha)$ of Y, the former problem translates into a similar problem for f_Y. The MaxEnt technique uses the information contained in the moment curve to determine an approximate density and an approximate value of the entropy of the unknown density. And by using a sufficiently large number of fractional moments an approximation to the entropy of the true density can be obtained. With these remarks in mind, our problem consists of determining which values of the parameter α, or which moments $\mu_Y(\alpha)$ will at the same time yield a value of $H[f_M]$ as large (as informative) as possible, and as close as possible to $H[f_Y]$. The answer is provided by the following procedure, devised by Novi-Inverardi and Tagliani (2003). Note that given a finite collection of fractional moments $\{\mu_Y(\alpha_k)\}_0^M$, the corresponding MaxEnt solution to the fractional moment problem is given by

$$f_M(y) = \exp\left(-\sum_{k=0}^{M} \lambda_k y^{\alpha_k}\right) \tag{4.1}$$

where the λ_k are such that the constraints

$$\int_0^1 e^{-\sum_{k=1}^{M} \lambda_k y^{\alpha_k}} dy = \mu_Y(\alpha_k), \qquad k = 0, ..., M. \tag{4.2}$$

are satisfied, and $f_M(y)$ depends on the $\mu_Y(\alpha_k)$ (thus on the α_k) through the λ_k. Actually, the λ's are obtained by minimizing the "dual" entropy

$$\ln\Big(\int_0^1 \exp(-\sum_{j=1}^{M} \lambda_j y^{\alpha_j}) dy\Big) + \sum_{j=1}^{M} \lambda_j \mu_Y(\alpha_j) \tag{4.3}$$

At the minimum, the value of this function coincides with the value of $H[f_M]$. This is the key idea behind the nested minimization procedure proposed below. Therefore, minimizing (3.5) amounts to

$$\arg\min\{\int_0^1 f_Y(y) \ln \frac{f_Y(y)}{f_M(y)} dy \,|\, \alpha_1, ..., \alpha_M\} = \arg\min\{H[f_M] \,|\, \alpha_1, ..., \alpha_M\}. \tag{4.4}$$

Indeed, the quantity to be minimized ($\min K(f_Y, f_M)$) depends on f_Y but, the fact that f_Y is unknown is unnecessary if we keep (3.5) in mind. Consequently, the knowledge of f_Y is superfluous at the end to solve the minimization problem (4.4) and the choice of $(\alpha_1, ...\alpha_M)$ depends on the minimization with respect to α's of $H[f_M]$. The latter depends uniquely from the available information $\{\mu_Y(\alpha_j)\}_{j=1}^M$, according to the MaxEnt procedure.

That is, f_M is obtained through a nested minimization procedure, namely

$$\begin{aligned} (\alpha_1, ..., \alpha_M, \lambda_1, ..., \lambda_M) : \quad & \min_{\alpha_1,...,\alpha_M} \min_{\lambda_1,...,\lambda_M} H[f_M(\boldsymbol{\lambda}, \boldsymbol{\alpha})] = \\ = \min_{\alpha_1,...,\alpha_M} \min_{\lambda_1,...,\lambda_M} \Big[\ln\Big(\int_0^1 \exp(-\sum_{j=1}^{M} \lambda_j y^{\alpha_j}) dy\Big) & + \sum_{j=1}^{M} \lambda_j \mu_Y(\alpha_j)\Big], \quad M = 1, 2, ... \end{aligned} \tag{4.5}$$

The crucial issue solving the above nested minimization consists of finding the optimal M value. We will consider that in next section. Now we prove the above procedure (4.5) arises approximate

f_M converging in entropy to f_Y, so that we identify $H[f_Y]$ with $H[f_M]$. Indeed, whenever nodes at $\alpha_j = j\,a/M,\ j = 0, ..., M$ for some $a > 0$, are selected, Novi and Tagliani (2003) proved f_M converge in entropy to the underlying unknown density f_Y. As a consequence, entropy convergence is guaranteed and accelerated when nodes $\alpha_j = ja/M$ are replaced by optimal nodes obtained in (4.5). Loosely speaking, the $\alpha_j = j\,a/M,\ j = 0, ..., M$ yield a constrained minimum, whilst yields (4.5) an unconstrained minimum.

We also prove explicitly that the sequence $\{H[f_M]\}$ coming from (4.5) is strictly monotonic decreasing. Indeed, let's consider (4.5), fix M and calculate $(\alpha_1, ..., \alpha_M)$ from which $H[f_M]$. Next put $M + 1$ in (4.5). As a first step, take the special set $(\alpha_1, ..., \alpha_{M+1})$ where the first entries $(\alpha_1, ..., \alpha_M)$ coincide with the just above found and $\alpha_{M+1} > \alpha_M$ is kept arbitrarily (that is constrained minimization running on α_{M+1} only, whilst $(\alpha_1, ..., \alpha_M)$ is held fixed). Calculate $H[f_{M+1}]$ and call it $H^*[f_{M+1}]$, with $H^*[f_{M+1}] < H[f_M]$. As a second step take $M + 1$ in (4.5), where the minimum runs on $(\alpha_1, ..., \alpha_{M+1})$, from which $H[f_{M+1}]$ (that is unconstrained minimization). It follows $H[f_{M+1}] < H^*[f_{M+1}] < H[f_M]$. The sequence $\{H[f_M]\}$ is strictly monotonic decreasing and convergent to $H[f_Y]$.

4.1 Optimal model

The true densities are related by $f_X(x) = e^{-x} f_Y(e^{-x})$, whilst f_X is approximated is given by $e^{-x} f_M(e^{-x})$; here f_M is from (4.1), with parameters estimated by (4.5). The optimal model is found by requiring that the Kullback-Leibler distance satisfies $K(f_X, e^{-x} f_M(e^{-x})) < Tol$, where Tol indicates a prefixed error. It holds

$$K(f_X, e^{-x} f_M(e^{-x})) = \int_0^\infty e^{-x} f_Y(e^{-x}) \ln\Big[\frac{e^{-x} f_Y(e^{-x})}{e^{-x} f_M(e^{-x})}\Big] dx =$$
$$= \int_0^1 f_Y(y) \ln \frac{f_Y(y)}{f_M(y)} dy = H[f_M] - H[f_Y]$$

According to our previous results

1. $H[f_Y]$ is calculated by using finitely many real values $L(\alpha_j = j)$ of the Laplace transform, or integer moments of f_Y. (It must not be confused with the $H[f_M]$ in point 2, coming up next, where $H[f_M]$ is calculated by fractional moments according to (4.5));

2. $H[f_M]$ is calculated from (4.5) assigning increasing values to M.

Optimal model in terms of Kullback-Leibler distance is identified by

Proposition 4.1 *If f_Y has finite entropy $H[f_Y]$, optimal M comes from*

$$\min_M \Big[K(f_X, e^{-x} f_M(e^{-x})) = H[f_M] - H[f_Y] \Big] < Tol \tag{4.6}$$

4.2 Cumulative distribution function

Let F_M and F_X denote the cumulative distribution functions of $e^{-x} f_M(e^{-x})$ and f_X respectively. Taking into account (3.8) one has

$$\begin{aligned} \sup_{x\in[0,+\infty)} \mid F_M(x) - F_X(x) \mid \leq \sup_{x\in[0,+\infty)} \int_0^x e^{-t} \mid f_M(e^{-t}) - f_Y(e^{-t}) \mid dt = \\ = \sup_{s\in(0,1]} \int_s^1 \mid f_M(t) - f_Y(t) \mid dt \leq \\ \leq \int_0^1 \mid f_M(t) - f_Y(t) \mid dt \leq \sqrt{2(H[f_M] - H[f_Y])} \end{aligned}$$

Optimal model in terms of distribution function F_M is identified by

Proposition 4.2 *If f_Y has finite entropy $H[f_Y]$, optimal M comes from*

$$\sup_{x\in[0,+\infty)} \mid F_M(x) - F_X(x) \mid \leq \sqrt{2(H[f_M] - H[f_Y])} < Tol$$

Remark. Analogous error estimate is found in [Mnatsakanov-Sarkisian (2013), Th. 1]. Using equispaced real valued $L(\alpha_j = j)$ only, the authors construct an approximate cumulative distribution function F_M converging uniformly to F_X. Nevertheless the error $\sup_{x\in[0,+\infty)} \mid F_M(x) - F_X(x) \mid$ requires bounded f_X and f'_X and which are unknown.

4.3 Entropy estimation

Entropies $H[f_X]$ and $H[e^{-x} f_M(e^{-x})]$ of f_X and $e^{-x} f_M(e^{-x})$ respectively are compared. With easy computation, and recalling both (4.1) and $\mu'_Y(0) = L'(0)$, one has

$$\begin{aligned} H[e^{-x} f_M(e^{-x})] - H[f_X] = (H[f_M] - \mu'_{f_M}(0)) - (H[f_Y] - \mu'_Y(0)) = \\ = H[f_M] - H[f_Y] - \int_0^1 \ln(y) \Big(f_M(y) - f_Y(y) \Big) dy \end{aligned} \tag{4.7}$$

(here $\mu_{f_M}(\alpha)$ denotes the moment curve of f_M). Taking into account $\lim_{M\to\infty} H[f_M] = H[f_Y]$ from (4.7) it follows

$$\lim_{M\to\infty} H[e^{-x} f_M(e^{-x})] - H[f_X] = \lim_{M\to\infty} \int_0^1 \ln(y)(f_Y(y) - f_M(y))dy = \lim_{M\to\infty} \mu'_Y(0) - \mu'_{f_M}(0) \tag{4.8}$$

In general, $(\mu'_Y(0) - \mu'_{f_M}(0))$ may not converge to zero. The hypothesis that $H[f_Y]$ finite, is not enough to guarantee that $e^{-x} f_M(e^{-x})$ converges to f_X in entropy. Additional hypotheses, like for instance, that f_Y being twice continuously differentiable and strictly positive, imply the sequences f_M converge uniformly to f_Y (see [Borwein-Lewis, 1991, Corollary 5.3]). It is not evident how only the knowledge of $L(\alpha)$ is enough to guarantee such hypotheses are fulfilled). In this case, in (4.8) we have that $\lim_{M\to\infty} \int_0^1 \ln(y)(f_Y(y) - f_M(y))dy = 0$ and then $\lim_{M\to\infty} \mu'_{f_M}(0) - \mu'_Y(0) = 0$. Under the above restrictive hypotheses on f_Y the following conclusions can be drawn

1. $\lim_{M\to\infty} H[e^{-x} f_M(e^{-x})] - H[f_X] = 0$, i.e. $e^{-x} f_M(e^{-x})$ converges to f_X in entropy.
2. The moment curves $\mu_{f_Y}(\alpha)$ and $\mu_{f_M}(\alpha)$ are tangent at $\alpha = 0$.

This last result allows us to extend a previous one about geometric meaning of the moment curves $\mu_{f_M}(\alpha)$ and $\mu_{f_Y}(\alpha)$ presented in Gzyl et al., (2014), and and formulated as follows: the two moment curves generated by the unknown density and its MaxEnt approximation are interpolating at the nodes $(\alpha_0 = 1, \alpha_1, ..., \alpha_M)$ and tangent at the nodes $(\alpha_1, ..., \alpha_M)$. Therefore, the geometrical meaning of the MaxEnt procedure can be reformulated as

Proposition 4.3 *Supposing that f_Y is twice continuously differentiable and strictly positive, the moment curves $\mu_{f_Y}(\alpha)$ and $\mu_{f_M}(\alpha)$ are interpolating in the Hermite-Birkoff sense; that is, they are both interpolating and tangent at the selected nodes $(\alpha_0 = 1, \alpha_1, ..., \alpha_M)$.*

By summarizing, if $\mu_Y(\alpha) = \mathrm{L}(\alpha)$ is the moment curve with pdf f_Y and finite entropy $H[f_Y]$, and f_M denoting its MaxEnt approximant constrained by fractional moments, we have

a) The sequence $\{H[f_M]\}$ converges to $H[f_Y]$;

b) And $e^{-x} f_M(e^{-x})$ converges to $f_X = e^{-x} f_Y(e^{-x})$ in directed Kullback-Leibler divergence, i.e., $\lim_{M\to\infty} K[f_X, e^{-x} f_M(e^{-x})] = 0$.

With additional hypotheses f_Y twice continuously differentiable and strictly positive, $f_X = e^{-x} f_Y(e^{-x})$ and its approximate $e^{-x} f_M(e^{-x})$ satisfy both the above conditions a), b) and the following c), d) where

c) f_X and $e^{-x} f_M(e^{-x})$ converge in entropy, i.e. $\lim_{M\to\infty} H[e^{-x} f_M(e^{-x})] = H[f_X]$;

d) moment curves $\mu_{f_Y}(\alpha)$ and $\mu_{f_M}(\alpha)$ are interpolating in Hermite-Birkoff sense; that is, they are both interpolating and tangent at the selected nodes $(\alpha_0 = 1, \alpha_1, ..., \alpha_M)$.

The latter are found by means of (4.5). Equivalently, the Laplace transforms of f_X and its approximant $e^{-x} f_M(e^{-x})$ interpolate at selected nodes $(\alpha_0 = 1, \alpha_1, ..., \alpha_M)$. In other words, with a proper choice of nodes (the ones minimizing $H[f_M]$), the Laplace Transforms of f_X and $e^{-x} f_M(e^{-x})$ become tangent too at such nodes.

5 Numerical example

We illustrate the procedures described in the previous sections taking bounded and unbounded densities f_Y with fast or slow entropy convergence rate, so that general criteria cannot be drawn.

Example 1. For that, let us suppose that the Laplace transform that we have to start with, is given by

$$L(\alpha) = \mu_Y(\alpha) = \frac{\Gamma(p+q)\Gamma(p+\alpha)}{\Gamma(p+q+\alpha)\Gamma(p)}$$

The corresponding density f_Y comes from Beta distribution

$$f_Y(y) = \frac{1}{B(p,q)} y^{p-1}(1-y)^{q-1}$$

We chose $p = 4$ and $q = 2$. The density f_Y has Shannon entropy given in terms of Beta and Digamma function

$$\begin{aligned} H[f_Y] &= \ln B(p,q) - (p-1)[\psi(p) - \psi(p+q)] - (q-1)[\psi(q) - \psi(p+q)] \\ &= \ln 20 + 79/30 \simeq -0.3623989402206575, \end{aligned}$$

and $L'(0) = -0.45$. $L(\alpha_j = j)$, $j = 1, ..., M$ easy to obtain. Now, from (4.3) the upper bound $H[f_Y] \leq -0.011933$, and therefore $H[f_X] = H[f_Y] - L'(0) \leq 0.43807$. Note that latter bound is not tight, since $H[f_X] = H[f_Y] - L'(0) = 0.087601059$ is the correct value. Next, the linear system (3.10) with an increasing number M of integer moments $\mu_j = L(j)$ is solved. Using that, the entropy $H[f_M] = \lambda_0 + \langle \lambda, \mu \rangle$ is calculated and then, for high M values, $H[f_Y]$ may be replaced with $H[f_M]$. The approximate values of $H[f_M]$ (constrained by integer moments) are reported in Figure 1 and compared with exact $H[f_Y]$ (several examples are illustrated in Novi et al., 2012). As a final step, MaxEnt density f_M constrained by fractional moments $\mu_Y(\alpha_j) = L(\alpha_j)$, $\alpha_j \in I\!R^+$ is considered. The parameters of f_M are calculated according to (4.5) from which $H[f_M]$ reported in Table 1 for some values of M (here exact $H[f_Y] = -0.3623989402206575$ is used). The latter allows us the choice of the optimal model, i.e. the one with assigned Kullback-Leibler distance $K(f_X, e^{-x} f_M(e^{-x})) = H[f_M] - H[f_Y] < Tol$ according to (4.6). In Figure 2 the comparison between f_X and its approximate $e^{-x} f_M(e^{-x})$ as $M = 4$, where $Tol = 10^{-4}$ is chosen.

Example 2. In analogy with Example 1 the same Laplace Transform

$$L(\alpha) = \mu_Y(\alpha) = \frac{\Gamma(p+q)\Gamma(p+\alpha)}{\Gamma(p+q+\alpha)\Gamma(p)}$$

is considered, but this time we chose $p = q = 0.5$. The corresponding densities $f_Y(y) = \frac{1}{\pi\sqrt{y(1-y)}}$ and f_X are unbounded with $H[f_Y] \simeq -0.24156427$ (analytical expression of $H[f_Y]$ and moments see Example 1) and $L'(0) = -1.38629436112$, from which $H[f_X] = H[f_Y] - L'(0) = 1.14473$. From (4.3) the upper bound $H[f_Y] \leq -0.015175$ and then $H[f_X] = H[f_Y] - L'(0) \leq 1.3711$. Note that latter bound is more reliable if compared with Example 1. The comparison of $H[f_M]$ (constrained by fractional moments $\mu_j = L(\alpha_j)$) and $H[f_Y]$ is reported in Table 2. From Figure 3 entropy convergence is visibly much slower than Example 1. In Figure 4 the comparison between unbounded f_X and its approximate $e^{-x} f_M(e^{-x})$ as $M = 12$, where $Tol = 5 \cdot 10^{-3}$ is chosen.

Example 3. The following Laplace Transform is considered

$$L(\alpha) = \frac{1}{\ln(b/a)} \cdot \ln \frac{\alpha + b}{\alpha + a}$$

with

$$f_X(x) = \begin{cases} \frac{1}{x}(e^{-ax} - e^{-bx})/\ln(b/a) & \text{if} \quad x > 0 \\ \frac{b-a}{\ln(b/a)} & \text{if} \quad x = 0 \end{cases}$$

for $a = 1$ and $b = 2$. The corresponding densities f_Y and f_X are bounded, and $H[f_Y] \simeq -0.04868969144$ and $L'(0) = -\frac{1}{2\ln 2} \simeq -0.72134752$, from which $H[f_X] = H[f_Y] - L'(0) = 0.67265782899$. From (4.3) the upper bound $H[f_Y] \leq -0.0048739$ and then the bound $H[f_X] = H[f_Y] - L'(0) \leq 0.71647352$ are obtained. The comparison of $H[f_M]$ (constrained by fractional moments) and $H[f_Y]$ is reported in Table 3. From Figure 5 entropy convergence is visibly fast. In Figure 6 we display the comparison between f_X and its approximant $e^{-x} f_M(e^{-x})$ for $M = 3$, when $Tol = 2 \cdot 10^{-7}$ is chosen.

6 Conclusions

We considered the problem of estimating Boltzmann-Gibbs-Shannon entropy of a distribution with unbounded support on the positive real line. The entropy is to be obtained from the values of the Laplace transform without having to extend the Laplace transform to the complex plane to apply the Fourier based inversion. We only make use of a few well chosen values of the parameter of the Laplace transform to determine the probability density. The underlying methodology amounts to transform the unbounded domain onto a bounded domain, so that one replaces X with an appropriate Y. Then available information provided by the Laplace transform of X becomes information provided by fractional moments of Y. The results available for random variables with bounded support may be therefore applied. The methodology that we develop allows us to obtain both an estimate of the entropy of the unknown distribution and a good approximation to that density. Entropy and density estimates lead to a nested minimization procedures involving fractional moments.

Then the computational cost is comparable with the case of random variables having bounded support. Ill-conditioning plays a crucial role when high number of Laplace Transform values are considered; it arises just from ill-conditioned matrices involved in the linear system to be solved. Several authors investigated the question of whether or not the detrimental effects of ill-conditioning can be removed or mitigated by the use of special techniques. Here we adopt the preconditioning and the use of high-precision float arithmetics. Therefore, if $L(\alpha)$ is determined with small uncertainty, the uncertainty in the reconstruction may be large. What may save the situation is the fact that in many cases, we know that the function f_X we are looking for has no high-frequency components and, as a matter of fact, is actually quite a smooth function. This may help the user in choosing among possible solutions. Indeed, through the Maxent technique, we select a solution which is essentially smooth; in other terms, we use knowledge of the structural behavior of unknown density to obtain numerical values. In [Gautschi, 1969] the author concludes that there is no escape from ill-conditioning, which, after all, only reflects indirectly the fact the original inversion problem is not well posed.

References

[1] Borwein J.M., Lewis, A.S. (1991), Convergence of best entropy estimates, *SIAM J. Optimization*, **1**, 191-205.

[2] Csiszár I., Shields P.C. (2004) *Information Theory and Statistics: A Tutorial*, New Publishers Inc., Delft.

[3] Fasino D. (1995), Spectral properties of Hankel matrices and numerical solutions of finite moment problems, *J. Comput. Appl. Math.,* **65**, 145-155.

[4] Forte B., Hughes W., Pales Z.(1989), Maximum Entropy Estimators and the Problem of Moments, *Rendiconti di Matematica*, Serie VII, **9**, 689-699, (1989).

[5] Gautschi, W. (1969) On the Condition of a matrix arising in the Numerical Inversion of the Laplace Transform , *Math. of Computation*, **23**, n.105, 109-118.

[6] Gavriliadis P.N. (2008) Moment information for probability distributions, without solving the moment problem. I: Where is the mode?, *Comm. in Statistics- Theory and Methods* , 37, 671-681.

[7] Gavriliadis P.N., Athanassoulis G.A. (2009) , Moment information for probability distributions, without solving the moment problem. II:Main-mass, tails and shape approximation, *J. of Comput. and Appl. Math.*, **229**, 7-15.

[8] Gzyl H., Novi-Inverardi P.L., Tagliani A. (2014), Fractional moments and maximum entropy: geometric meaning, *Comm. in Statistics - Theory and Methods*, **43**, 3596-3601.

[9] Kesavan H.K., Kapur J.N. (1992), *Entropy Optimization Principles with Applications*, Academic Press, London.

[10] Kullback S., Leibler, R. (1951), On information and sufficiency , *The Annals of Math. Stats.*, **22**, 79-86.

[11] Jaynes, E.T. (1957) Information theory and statistical mechanics, *The Physical Review*, **106**, 620-630.

[12] Lin, G.D. (1992) Characterizations of distributions via moments, *Sankhya: The Indian Journal of Statistics*, **54**, 128-132.

[13] Mnatsakanov,R. (2008a) Hausdorff moment problems: Reconstruction of distributions. *Statistics& Probability Letters*, **78**, 1612-1618.

[14] Mnatsakanov,R. (2008b) Hausdorff moment problems: Reconstruction of probability density functions, *Statistics& Probability Letters*, **78**, 1869-1877.

[15] Mnatsakanov R., Sarkisian K. (2013) A note on recovering distributions from exponential moments. *Applied Math. and Comput.*, **219**, 8730-8737.

[16] Novi-Inverardi P.L., Tagliani (2003) A. Maximum entropy density estimation from fractional moments. *Comm. in Statistics - Theory and Methods*, **32**, 15-32.

[17] Novi-Inverardi P.L., Milev M., Tagliani A. (2012), Moment information and entropy evaluation probability density, *Applied Math. and Comput.*, **218**, 5782-5795.

[18] Piera F., Parada P. (2009), On the convergence properties of Shannon entropy, *Problems of Information Transmission*, **45**, 75-94.

[19] Silva, J.F. and Parada, P. (2012),On the convergence of Shannon differential entropy, and its connection with density and entropy estimation, *Jour. Statistical Planning and Inference*, **142**, 1716-1732.

[20] Tagliani A. (1999) Hausdorff moment problem and maximum entropy: a unified approach, *Applied Math. and Comput.*, **105**, 291-305.

[21] Tagliani A. (2002) Entropy estimate of probability densities having assigned moments: Hausdorff case , *Appl. Math. Letters*, **15**, 309-314.

[22] Talenti, G. (1987), Recovering a function from a finite number of moments. *Inverse Problems*, **3**, 501-517.

M	$H[f_M]$	$H[f_M] - H[f_Y]$
1	-0.24232990	0.12006904
2	-0.35963377	0.00276517
3	-0.36149744	0.00090150
4	-0.36235147	0.00004740
6	-0.36236427	0.00003466
8	-0.36238158	0.00001736

Table 1: Entropy for increasing number of fractional moments

M	$H[f_M]$	$H[f_M] - H[f_Y]$
1	-0.09403989	0.14752437
2	-0.20756106	0.03400320
3	-0.22698956	0.01457470
4	-0.22884011	0.01272415
5	-0.23102261	0.01054165
6	-0.23244454	0.00911972
7	-0.23469685	0.00686741
8	-0.23535919	0.00620507
10	-0.23569185	0.00587241
12	-0.23617640	0.00538786

Table 2: Entropy for increasing number of fractional moments

M	$H[f_M]$	$H[f_M] - H[f_Y]$
1	-0.048689391	0.0000002996
2	-0.048689451	0.0000002398
3	-0.048689493	0.0000001980

Table 3: Entropy for increasing number of fractional moments

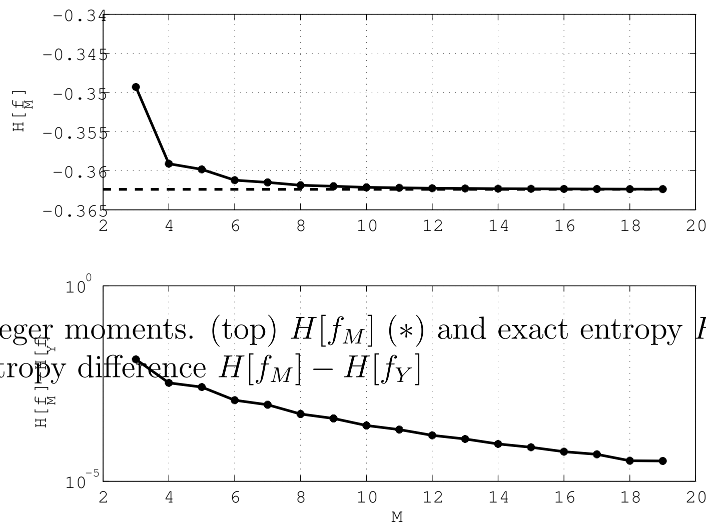

Figure 1: Integer moments. (top) $H[f_M]$ $(*)$ and exact entropy $H[f_Y]$ (dashed line); (bottom) Entropy difference $H[f_M] - H[f_Y]$

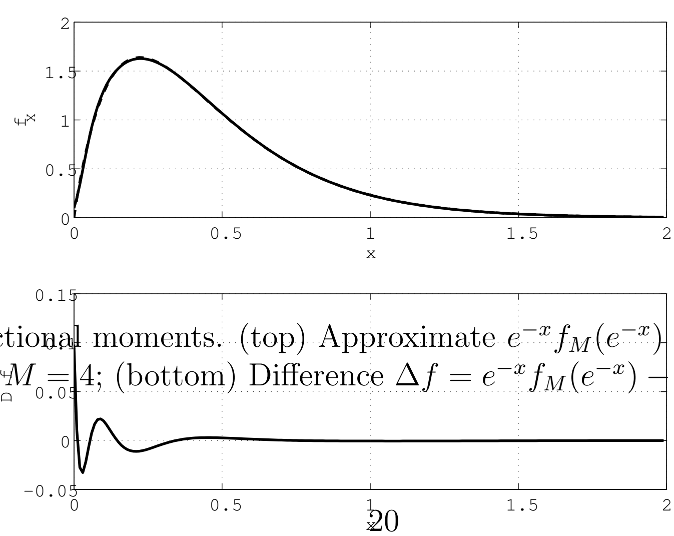

Figure 2: Fractional moments. (top) Approximate $e^{-x} f_M(e^{-x})$ (continuous line), f_X (dashed line), $M = 4$; (bottom) Difference $\Delta f = e^{-x} f_M(e^{-x}) - f_X$

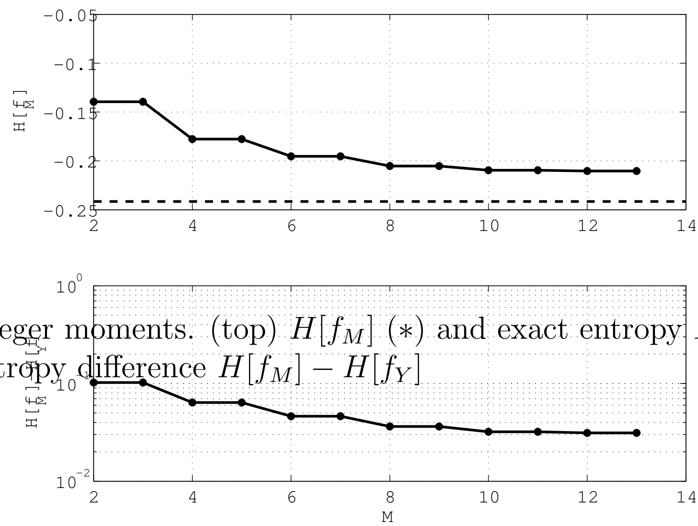

Figure 3: Integer moments. (top) $H[f_M]$ $(*)$ and exact entropy $H[f_Y]$ (dashed line); (bottom) Entropy difference $H[f_M] - H[f_Y]$

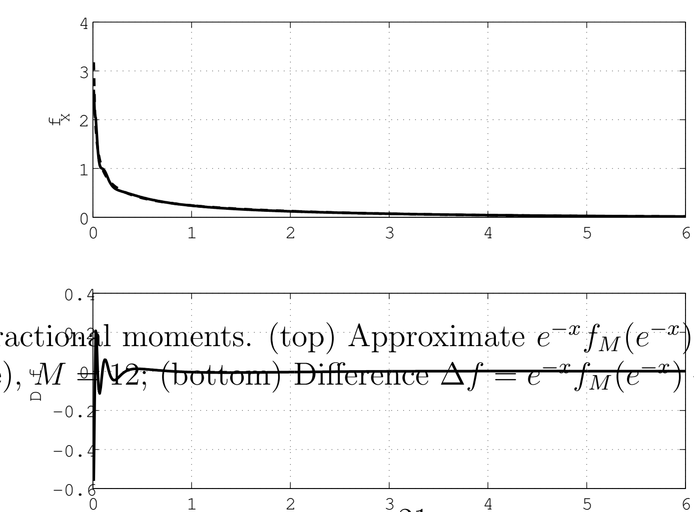

Figure 4: Fractional moments. (top) Approximate $e^{-x} f_M(e^{-x})$ (continuous line), f_X (dashed line), $M = 12$; (bottom) Difference $\Delta f = e^{-x} f_M(e^{-x}) - f_X$

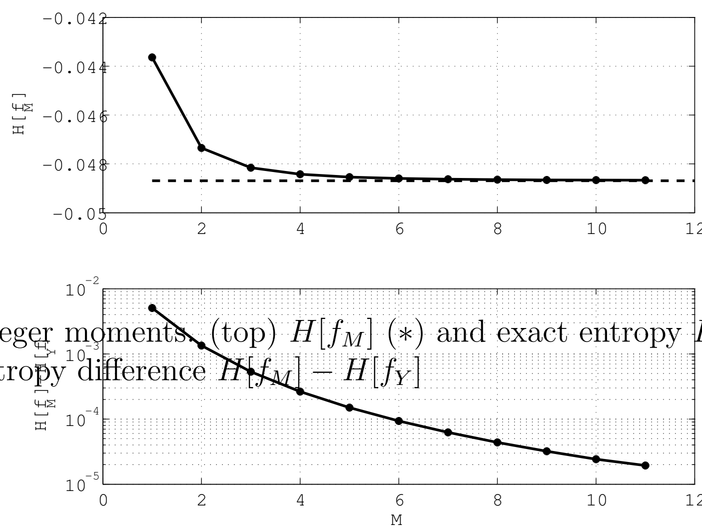

Figure 5: Integer moments. (top) $H[f_M]$ $(*)$ and exact entropy $H[f_Y]$ (dashed line); (bottom) Entropy difference $H[f_M] - H[f_Y]$

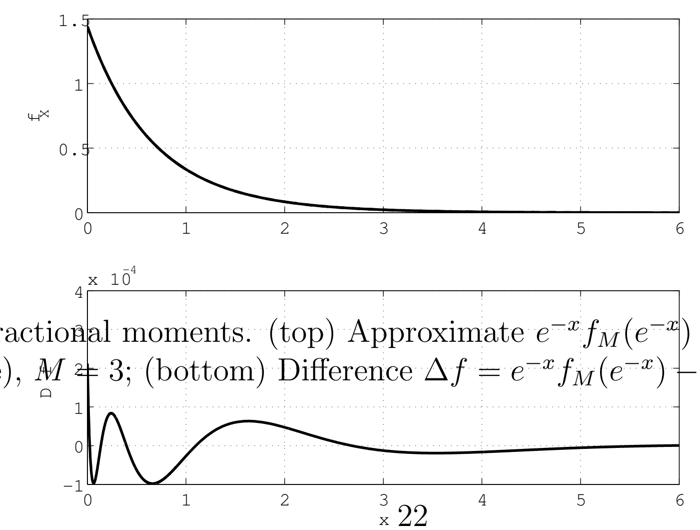

Figure 6: Fractional moments. (top) Approximate $e^{-x} f_M(e^{-x})$ (continuous line), f_X (dashed line), $M = 3$; (bottom) Difference $\Delta f = e^{-x} f_M(e^{-x}) - f_X$