Article

A Relationship between the Ordinary Maximum Entropy Method and the Method of Maximum Entropy in the Mean

Henryk Gzyl [1],* and Enrique ter Horst [1,2]

[1] Centro de Finanzas, IESA, Ave. Iesa, San Bernardino, Caracas 1010, Venezuela;
   E-Mail: enrique.terhorst@cesa.edu.co
[2] Institute 2, CESA, Calle 35 #6-16, Bogotá, Colombia

* Author to whom correspondence should be addressed; E-Mail: henryk.gzyl@iesa.edu.ve;
   Tel.: +58-212-5554385; Fax: +58-212-554466.

Abstract: There are two entropy-based methods to deal with linear inverse problems, which we shall call the ordinary method of maximum entropy (OME) and the method of maximum entropy in the mean (MEM). Not only does MEM use OME as a stepping stone, it also allows for greater generality. First, because it allows to include convex constraints in a natural way, and second, because it allows to incorporate and to estimate (additive) measurement errors from the data. Here we shall see both methods in action in a specific example. We shall solve the discretized version of the problem by two variants of MEM and directly with OME. We shall see that OME is actually a particular instance of MEM, when the reference measure is a Poisson Measure.

Keywords: maximum entropy; maximum entropy in the mean; constrained linear inverse problems

---

1. Introduction

During the last quarter of the XIX-th century Boltzmann proposed a way to study convergence to equilibrium in a system of interacting particles through a quantity that was that was a Lyapunov functional for the dynamics of the system, and increased as the system tended to equilibrium. A related idea was used at the beginning of the XX-th century by Gibbs to propose a theory of equilibrium statistical mechanics. The difference between the approaches was in the nature of the microscopic

description. In the late 1950s, Jaynes in [1] turned the idea into a variational method to determine a probability distribution given the expected value of a few random variables (observables to use the physical terminology). This procedure is called the method of maximum entropy. This methodology has proven useful in a variety of problems well removed from the standard statistical physics setup. See Kapur (1989) [2] for example, or the Kluwer Academic Press collection of Maximum Entropy and Bayesian Methods or the volume by Jaynes (2003) [3].

As it turns out, similar procedures had come up in the actuarial and statistical literature, see for example the works by Esscher (1932) [4] and by Kullback (1957) [5]. Jaynes's procedure was further extended in Decarreau *et al.* (1992) [6], and Dacunha-Castelle and Gamboa (1990) [7]. Such extension has proven a powerful tool to deal with linear inverse problems with convex constraints. See Gzyl and Velásquez (2011) [8] for example. This method uses the standard variational technique as a stepping stone in a peculiar way.

Besides providing a more general type of solutions to the OME problem, we shall verify in two different ways that the standard solution of the OME is actually a particular case of the more general MEM approach to solve linear inverse problems with convex constraints.

The paper is organized as follows. In the remainder of this section we shall state the two problems whose solutions we want to relate. These consists in obtaining a positive, continuous function satisfying some integral constraints. In the next section we shall recall the basics of MEM. In section three we continue with the same theme and examine specific choices of set up to implement the method. Section 4 is devoted to the issue of obtaining the problem by the OME from the solution by the MEM.

In section five we implement both approaches numerically to compare their performance in one simple example. The idea of using two different choices of prior is to emphasize the flexibility of the MEM.

1.1. Statement of the First Problem

Even though the problem considered is not in its most general form, it is enough for our purposes and can be readily extended. We want to find a continuous positive function $x(t) : [0, 1] \rightarrow [0, \infty)$ such that

$$\int_0^1 k_i(t)x(t)dt = m_i; \ \ \text{for} \ \ i = 1, ..., M \tag{1}$$

Typically $\{k_i(t) : i = 1, ..., M\}$ a collection of measurable functions defined on $[0, 1]$ describing some sort of observations made on a random variable whose density we want to estimate. These could be ordinary moments t^{n_i} for some collection $\{n_1, ..., n_M\}$ of integers. Or they could be fractional powers t^{a_i} for some collection $\{a_1, ..., a_M\}$ of reals. This problem appears when our information consists of the values of a Laplace transform at points $\{a_1, ..., a_M\}$ and we map our problem onto $[0, 1]$ by means of the change of variables $s \rightarrow t = e^{-s}$. Or the $k_i(t)$ could be trigonometric polynomials. In such case we refer to Equation (1) as a generalized moment problem. When $x(t)$ is required to be a probability density, we shall consider $k_1(t) \equiv 1$ on $[0, 1]$. It is also apparent that the convex constraint is the positivity constraint on $x(t)$.

1.2. Statement of the Second Problem

Clearly Equation (1) is a particular case of the following more general problem: Let $k(s,t) : [a,b] \times [0,1] \to \mathbb{R}$. Let $\mathcal{K} \subset \mathcal{C}([0,1])$ be a cone contained in the class of continuous functions, and let $m(s) : [a,b] \to \mathbb{R}$ be some continuous function. We want to find $x(t) \in \mathcal{K}$ satisfying the integral constraints

$$\int_0^1 k(s,t)x(t)dt = m(s), \quad s \in [a,b] \tag{2}$$

We remark that when $x(t)$ is a density and $k(s,t) \equiv 1$, them $m(s) \equiv 1$. Clearly the integral constraints could be incorporated into $\mathcal{K}$, but it is convenient to keep both separated. For what comes below, and to relate to the first problem, we shall restrict ourselves to the convex set of continuous density functions. Such type of problems were considered for example in [9] or in much grater generality in [10] or and more recently in [11] where applications and further references to related work are collected. As mentioned above, the setup can be relaxed considerable at the expense of technicalities. For example, one can consider the kernel k to be defined on the product $S_1 \times S_2$ of two locally compact, separable metric spaces, and dt could be replaced by some $\sigma-$finite measure $\nu(dt)$ on $(S_2, \mathcal{B}(S_2))$. But let us keep it as simple as possible.

2. The Maximum Entropy in the Mean Approach

The basic intuition behind the MEM goes as follows. We search for a stochastic process with independent increments $\{X(t)|t \in [0,1]\}$ defined on some auxiliary probability space $(\Omega, \mathcal{F}, Q)$ such that

$$dX(t) = X(t+\delta) - X(t) \in \mathcal{K}, \quad \text{for } t \in [0,1] \text{ and } \delta > 0 \tag{3}$$

$$\frac{dE_P[X(t)]}{dt} = x(t); \quad \text{for some } P << Q \tag{4}$$

$$E_P[\int_0^1 k(s,t)dX(t)] = \int_0^1 k(s,t)x(t)dt = m(t) \tag{5}$$

Here the measure P on $(\Omega, \mathcal{F})$ is yet to be determined. If it exists, notice that $x(t) \in \mathcal{K}$ automatically. The integral with respect to $dX(t)$ is to be understood in the Itô sense.

As we want to implement the scheme numerically, it is more convenient to discretize Equation (2) and then to bring in the MEM. It is at this point where the regularity properties of $k(s,t)$ and $x(t)$ come in to make life easier. Consider a partition of $[a,b]$ into M equal adjacent intervals and a partition of $[0,1]$ into N adjacent intervals. Let $\{s_i|i = 1, ..., M\}$ and respectively $\{t_j|j = 1, ..., N\}$ be the center points of those intervals. Let us set $A_{ij} = A(i,j) = k(s_i, t_j)$. Also, set $x_j \equiv x(j) = x(t_j)/N$ and finally $m_i \equiv m(i) = m(s_i)$.

Comment We chose $x_j = x(t_j)/N$ because when $x(t)$ is a density, we want its discretized version to satisfy $\sum_j x_j = 1$.

With these changes, the discretized version of the second problem becomes: Given M real numbers $m_i : i = 1, ..., M$, determine positive numbers $x_j : j = 1, ..., N$ such that

$$\sum_{j=1}^N A_{ij}x_j = m_i, \quad \text{for } i = 1, ..., M \tag{6}$$

To assemble the model, consider $\Omega = [0, \infty)^N$ with $\mathcal{F} = \mathcal{B}(\Omega)$ the usual Borel sets. To make things really simple, let $q_j(d\xi_j)$ be N copies of a Measure $q(d\xi)$ on $([0, \infty), \mathcal{B}([0, \infty)))$ and let

$$Q(d\boldsymbol{\xi}) = \prod_{j=1}^{N} q_j(d\xi_j)$$

be the reference measure. Note that with respect to Q the coordinate maps $X_j \equiv X(j) : \Omega \to [0, \infty)$ defined by $X_j(\boldsymbol{\xi}) = \xi_j$ satisfy the positivity constraints and are independent. With this notation, the original discretized problem (6) is transformed

$$\text{Determine a probability measure } P << Q \ \text{ such that } \ E_P[\mathbf{AX}] = \mathbf{m} \tag{7}$$

Note that if such a measure P is found, then $x_j = E_P[X_j]$ satisfies Equation (6). It is to determine P where the OME comes in as a stepping stone.

3. Solution of Equation (7) by MEM

The notation will be as at the end of the previous section. For the purpose of comparison, we shall solve Equation (7) using two different measures. First we shall consider a product of exponential distributions on Ω and then we shall consider a product of Poisson distributions. Let us first develop the generic procedure, and then particularize for each choice of reference measure.

At this point we mention that the only requirement on the reference measure $Q(d\boldsymbol{\xi}) = \prod_{j=1}^{N} q_j(d\xi_j)$ is the following:

Assumption We shall require the closure of the convex hull generated by the support of Q to be exactly Ω.

Let us consider the convex set $\mathcal{P}(Q) = \{P \text{ measure on} (\Omega, \mathcal{F}), P << Q\}$ on which we define the following concave functional

$$S_Q(P) = -\int_\Omega \ln\left(\frac{dP}{dQ}\right) dP \tag{8}$$

whenever $\ln(dP/dQ)$ is $P-$integrable and equal $-\infty$ otherwise. This is the negative of the Kullback-Leibler divergence between P and Q. It is a standard result that $S_Q(P)$ is concave in P and we have

Lemma 3.1. *Suppose that P, Q and R are probability measures on $(\Omega, \mathcal{F})$ such that $P << Q, P << R$ and $R << Q$, then $S_R(P) \leq 0$, and*

$$S_Q(P) = S_R(P) - \int dP \ln\left(\frac{dR}{dQ}\right) \tag{9}$$

Proof. The verification of the first assertion is easy invoking Jensens' inequality. The second follows readily from the fact that $\frac{dP}{dQ} = \frac{dP}{dR} / \frac{dQ}{dR}$. □

We want to consider the following consequence of this lemma. Let us define R by $dR_{\boldsymbol{\lambda}}(\boldsymbol{\xi}) = \rho_{\boldsymbol{\lambda}}(\boldsymbol{\xi}) dQ(\boldsymbol{\xi})$ where

$$\rho_{\boldsymbol{\lambda}}(\boldsymbol{\xi}) = \frac{e^{-<\boldsymbol{\lambda}, \mathbf{A}\boldsymbol{\xi}>}}{Z(\boldsymbol{\lambda})} \tag{10}$$

where $Z(\boldsymbol{\lambda})$ is the obvious normalization factor, which is given by

$$Z(\boldsymbol{\lambda}) = \int e^{-<\boldsymbol{\lambda}, A\boldsymbol{\xi}>} dQ\boldsymbol{\xi}$$

The idea behind the maximum entropy method, comes from the realization that for such R, when P satisfies the constraints, substituting Equation (10) in the integral term in Equation (9), we obtain that

$$\Sigma(\boldsymbol{\lambda}) \equiv \ln Z(\boldsymbol{\lambda}) + <\boldsymbol{\lambda}, \mathbf{m}> \geq S_Q(P) \text{ for any } \boldsymbol{\lambda} \in \mathbb{R}^M \tag{11}$$

Thus, whenever $\{\boldsymbol{\lambda} \in \mathbb{R}^M | Z(\boldsymbol{\lambda}) < \infty\}$ we expect the problem to have a solution, and whenever the class of P's satisfying the constraints Equation (7) is non-empty, a minimizer of the convex function $\Sigma(\boldsymbol{\lambda})$ is expected to exist. Actually, Csiszar in [12,13] proved the existence of such P's. Actually a different (a perhaps more physicist oriented) proof was provided more recently in [14]. It is actually a simple exercise to verify the following

Proposition 3.1. *Suppose that a measure P satisfying (7) exists and that the minimum of $\Sigma(\boldsymbol{\lambda})$ is reached at $\boldsymbol{\lambda}^*$ in the interior of $\{\boldsymbol{\lambda} \in \mathbb{R}^M | Z(\boldsymbol{\lambda}) < \infty\}$, then the probability P^* that maximizes $S_Q(P)$ and satisfies Equation (7) is given by*

$$dP^*(\boldsymbol{\xi}) = \frac{e^{-<\boldsymbol{A}^\dagger\boldsymbol{\lambda}^*, \boldsymbol{\xi}>}}{Z(\boldsymbol{\lambda}^*)} dQ(\boldsymbol{\xi}) \tag{12}$$

We are using $\mathbf{A}^\dagger$ to denote the transpose of $\mathbf{A}$ With that result, the next step consists of computing

$$x^*(j) = \int_\Omega \xi_i dP^*(\boldsymbol{\xi}) \tag{13}$$

Let us now examine two possible choices for Q.

3.1. Exponential Reference Measure

Since we want positive x_j, we shall first try all factor $q(d\xi) = \mu e^{-\mu\xi} d\xi$, which has $[0, \infty)$ as support. A simple computation yields that

$$Z(\boldsymbol{\lambda}) = \prod_{j=1}^N \frac{\mu}{\mu + (\boldsymbol{A}^\dagger\boldsymbol{\lambda})_j}$$

For the purpose of modeling one has to chose μ large enough such that $\mu + (\boldsymbol{A}^\dagger\boldsymbol{\lambda})_j$ is positive. Another simple computation (as in the verification of the preceding proposition), yields

$$x_j^* = \frac{1}{\mu + (\boldsymbol{A}^\dagger\boldsymbol{\lambda}^*)_j} \tag{14}$$

Observe that the method has just shifted the mean of each exponential to a new value. We let the reader to write down P^* explicitly to verify this assertion.

3.2. Poisson Reference Measure

This time, instead of a product of exponentials, we shall consider a product of Poisson measures, *i.e.*, we take

$$q(d\xi) = e^{-\mu} \sum_{k \geq 0} \frac{\mu^k}{k!} \epsilon_{\{k\}}(d\xi)$$

Here we use $\epsilon_{\{a\}}(d\xi)$ to denote the unit point mass (Dirac delta) at a. Certainly the convex hull of the non-negative integers is $[0, \infty)$. Notice that now

$$Z(\boldsymbol{\lambda}) = \prod_{j \geq 0}^{N} \exp\left(-\mu(1 - e^{-(\boldsymbol{A}^{\dagger}\boldsymbol{\lambda})_j})\right)$$

from which we obtain

$$\Sigma(\boldsymbol{\lambda}) = -\mu \sum_{j=1}^{N} \left(1 - e^{-(\boldsymbol{A}^{\dagger}\boldsymbol{\lambda})_j}\right) + <\boldsymbol{\lambda}, \boldsymbol{m}>$$

Notice now that if $\boldsymbol{\lambda}^*$ minimizes that expression, then the estimated solution to Equation (7) is

$$x_j^* = e^{-(\boldsymbol{A}^{\dagger}\boldsymbol{\lambda}^*)_j} \tag{15}$$

3.3. The MEM Approach to the Original Problem

Consider Equation (1) again. This time we shall consider a Poisson point process on $([0, 1], \mathcal{B}([0, 1])$ with intensity dt. By this we mean a base probability space $(\Omega, \mathcal{F}, Q)$ on which a collection of random measures $\{N(A) : A \in \mathcal{B}([0, 1])\}$ (the point process) is given, which has the following properties:

(1) N(A) is a Poisson random variable with intensity (mean)$|A|$,

(2) $Q-$ almost everywhere $A \to N(A)$ is an integer valued measure

(3) For any disjoint $A_1, ..., A_k$ the $N(A_1), ..., N(A_k)$ are independent

From these, it is clear that for any $\boldsymbol{\lambda} \in \mathbb{R}^M$ the random variable $\int_0^1 <\boldsymbol{\lambda}, \boldsymbol{k}(t)> N(dt)$ satisfies

$$E_Q[e^{-\int_0^1 <\boldsymbol{\lambda}, \boldsymbol{k}(t)>N(dt)}] = \exp(-\int_0^1 dt(\exp(-<\boldsymbol{\lambda}, \boldsymbol{k}(t)>) - 1))$$

where $\mathbf{k}(t)$ is the vector of generalized moments appearing in Equation (1). Clearly, here we again denote the previous quantity by $Z_Q(\boldsymbol{\lambda}, \boldsymbol{k})$, and again $\Sigma(\boldsymbol{\lambda}) = \ln\left(Z_Q(\boldsymbol{\lambda})\right) + <\boldsymbol{\lambda}, \boldsymbol{m}>$. This function is convex on $\{\boldsymbol{\lambda} \in \mathbb{R}^M | Z_Q(\boldsymbol{\lambda}) < \infty\}$. When a minimizer $\boldsymbol{\lambda}^*$ exists in the interior of that domain, then P^* with density

$$\frac{dP^*}{dQ} = \frac{e^{-\int_0^1 <\boldsymbol{\lambda}^*, \boldsymbol{k}(t)>N(dt)}}{Z_Q(\boldsymbol{\lambda}^*)}$$

is such that

$$x_\infty^*(t) \equiv \frac{dE_{P^*}[N[0, t]]}{dt} = e^{-<\boldsymbol{\lambda}^*, \boldsymbol{k}(t)>} \tag{16}$$

solves Equation (1).

4. OME from MEM

4.1. Discrete Case

We shall now relate the last result to the standard (ordinary) method of maximum entropy. Suppose that the unknown quantities x_j in Equation (3) are indeed probabilities, and that $m_1 = 1$ and $A_{1j} = 1$ for all $j = 1, ..., N$. It is easy to verify using the first equation of the set Equation (3) that $\exp(-\lambda_1) = 1/\zeta(\boldsymbol{\lambda}_r^*)$ where $\boldsymbol{\lambda}_r = (\lambda_2, ..., \lambda_M)$ and

$$\zeta(\boldsymbol{\lambda}_r) = \sum_{j=1}^{N} e^{-\sum_{i=2}^{M} \lambda_i A_{i,j}}$$

From this, Equation (15) becomes

$$x_j^* = \frac{e^{-\sum_{i=2}^{M} \lambda_i^* A_{i,j}}}{\zeta(\boldsymbol{\lambda}_r^*)}$$

which is the solution to Equation (3) by the OME method.

4.2. Continuous Case as Limit of the Discrete Case

This is the second place in which our discretization procedure enters. First rewrite Equation (15) x_j^* as $x^*(t_j)/N$, from which Equation (15) becomes

$$x^*(t_j) = \frac{e^{-\sum_{i=2}^{M} \lambda_i^* A_{i,j}}}{\frac{1}{N}\zeta(\boldsymbol{\lambda}_r^*)} \tag{17}$$

One may want to argue as follows: Notice as well, that given any $t \in [0,1]$ as $N \to \infty$, there is a sequence $t_{j(N)}$ converging to t. In addition, it is clear that

$$\frac{1}{N}\zeta(\boldsymbol{\lambda}_r^*)) \to \int_0^1 e^{-\sum_{i=2}^{M} \lambda_i k_i(t)} dt$$

and therefore

$$x^*(t) = \frac{e^{-\sum_{i=2}^{M} \lambda_i^* k_i(t)}}{\int_0^1 e^{-\sum_{i=2}^{M} \lambda_i^* k_i(s)} ds}$$

which we would like to identify as the solution (1) provided by the OME method. The problem with the procedure is that the $\boldsymbol{\lambda}^*$ depends on N and changes along the way. Let us indicate a possible way to overcome this issue.

For each N denote by $A_j(N)$ the blocks of the partition of $[0, 1]$, and suppose that the partitions refine each other as N increases (consider dyadic partitions for example). For each N denote the maxentropic solution described in Equation (17) by $x_N^*(t_j)$ and define the piecewise constant (continuous) density

$$\tilde{x}_N(t) = \sum_j x_N^*(t_j) I_{A_j(N)}(t)$$

Clearly, $\tilde{x}_N$ satisfies Equation (1), but it is not the density that maximizes the entropy. Actually, one can rapidly verify that

$$S(\tilde{x}_N) \leq S(\tilde{x}_{N+1}) \leq S(\tilde{x}_\infty)$$

We shall relate $\tilde{x}_\infty$ to the x^*_∞ displayed in Equation (16) below. The remaining part of the argument is to verify that $\boldsymbol{\lambda}_N \to \boldsymbol{\lambda}_\infty$ (in an obvious notation) as $N \to \infty$. This is simple to say, but hard to prove. A way around the convergence of the $\boldsymbol{\lambda}_N$ issue is provided by [12].

4.3. The Full Continuous Case

Here we show how to obtain the OME solution to Equation (1) from the MEM solution Equation (16) without the labor described in the previous section. The argument is similar to the one mentioned above. As $k_1(t) = 1$, we can isolate λ^*_1 and rewrite $x^*_\infty(t)$ as

$$x^*(t) = \frac{e^{-\sum_{i=2}^M \lambda^*_i k_i(t)}}{\int_0^1 e^{-\sum_{i=2}^M \lambda^*_i k_i(s)} ds} \tag{18}$$

That this solves Equation (1) is due to the fact that the equations that determine $\boldsymbol{\lambda}^*$ in the full MEM and in the OME cases coincide. This happens because of the special form of $Z_Q(\boldsymbol{\lambda})$ when the underlying auxiliary process is the Poisson point process

5. Numerical Examples

To compare the output of the three methods, we consider a simple example in which the data consists in a few values of the Laplace transform of the density of a $\Gamma(a, b)$ density. Observe that if S denotes the original random variable, then $T = e^{-S}$ denotes the corresponding random variable with range mapped onto $[0, 1]$. The values of the Laplace transform of S are the fractional (non-necessarily integer) moments of T. The maxentropic methods yield the density $x(t)$ of T, from which the density of S is to be obtained by the change of variable $f_S(s) = e^{-s} x(e^{-s})$.

If we let $\{\alpha_1 = 0, \alpha_2, ..., \alpha_M\}$ and $k_i(t) = t^{\alpha_i}$ be M given powers of T, the corresponding moments to be used in Equation (1) are $m_i = \left(b/(\alpha_i + b)\right)^a$, with $m_1 = 1$. To be specific, let us consider $a = b = 1$, and $\alpha_2 = 1/5, \alpha_3 = 1/4, \alpha_4 = 1/3, \alpha_5 = 1/2, \alpha_6 = 5, \alpha_7 = 10, \alpha_8 = 15, \alpha_9 = 20$ from which we readily obtain the values of the 9 generalized moments m_i. To finish, we take $N = 100$ partition points of $[0, 1]$.
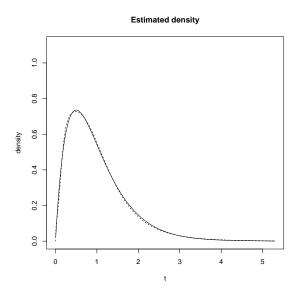
5.1. Exponential Reference Measure

We shall set $\mu = 10$ as a number high enough so that the positivity conditions mentioned in Section (3.1) holds. The function to be minimized to determine the λs is

$$\Sigma(\boldsymbol{\lambda}) = N \ln \mu - \sum_{j=1}^N \ln \left(\mu + (\boldsymbol{A}^\dagger \boldsymbol{\lambda})_j\right) + < \boldsymbol{\lambda}, \boldsymbol{m} >$$

To find the minimizer, we use the Barzilai-Borwein code available for R, see [15]. Once the optimal $\boldsymbol{\lambda}$ is obtained it is inserted in Equation (14). That is the density of T on $[0, 1]$. To plot the density on $[0, \infty)$ we perform the change of variables mentioned above and the result is plotted in the Figure 1.

We point out that the L_1-norm of the difference between the reconstructed and the original densities is 0.0283 rounding at the fourth decimal place.

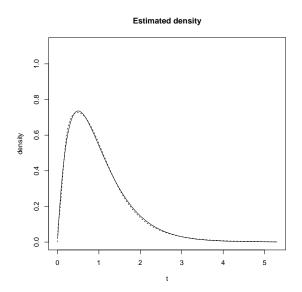**Figure 1.** Reconstruction by MEM with exponential reference measure.

5.2. The Poisson Reference Measure

In reference to the setup of Section 3.3 we set $\mu = 5$ this time. The function to be minimized this time is

$$\Sigma(\boldsymbol{\lambda}) = -\mu \sum_{j=1}^{N} \left(1 - e^{-(\boldsymbol{A}^{\dagger}\boldsymbol{\lambda})_j} \right) + < \boldsymbol{\lambda}, \boldsymbol{m} >$$

Once the minimizing $\boldsymbol{\lambda}^*$ has been found, the routine is as above: the density on $[0, 1]$ is mapped onto a density on $[0, \infty)$ by means of a change of variables. The result obtained is displayed on Figure 2.

Figure 2. Reconstruction by MEM with Poisson reference measure.

The L_1-norm of the difference between the reconstructed and the original densities is 0.00524.
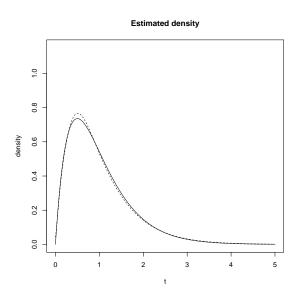
5.3. The OME Method

In this case, to determine $\boldsymbol{\lambda}^*$ we have to minimize

$$\Sigma(\boldsymbol{\lambda}) = \int_0^1 e^{-<\boldsymbol{\lambda},\,\boldsymbol{k}_{(t)}>}dt + <\boldsymbol{\lambda},\boldsymbol{m}>$$

which clearly is the same thing as minimizing

$$\Sigma(\boldsymbol{\lambda}) = -\int_0^1 \left(1 - e^{-<\boldsymbol{\lambda},\,\boldsymbol{k}_{(t)}>}\right)dt + <\boldsymbol{\lambda},\boldsymbol{m}>$$

as mentioned at the end of Section 4.3. The result, after the change of variables is displayed in Figure 3.

Figure 3. Reconstruction with OME.

The L_1-norm of the difference between the reconstructed and the original densities is 0.03479.

Acknowledgements

Conflicts of Interest

The authors declare no conflicts of interest.

Bibliographycal Comment

The literature on the OME method is very large, consider this journal for example. Even though the literature on the MME method is not that large, we cited only a few foundational papers and a very small sample of recent papers that apply the method to interesting problems.

References

1. Jaynes, E. Information theory and statistical physics. *Phys. Rev.* **1957**, *106*, 171–197.
2. Kapur, J. *Maximum Entropy Models in Science and Engineering*; Wiley Eastern Ltd.: New Delhi, India, 1989.
3. Jaynes, E. *Probability Theory: The Logic of Science*; Cambridge University Press: Cambridge, UK, 2003.
4. Escher, F. On the probability function in the collective theory of risk. *Scan. Aktuarietidskr.* **1932**, *15*, 175–195.
5. Kullback, S. *Information Theory and Statistics*; Dover Pubs: New York, NY, USA, 1968.
6. Decarreau, A.; Hilhorst, D.; Demarechal, C.; Navaza, J. Dual Methods in Entropy Maximization. Application to Some Problems in Crystallography. *SIAM J. Optim.* **1992**, *2*, 173–197.
7. Dacunha-Castelle, D.; Gamboa, F. Maximum d'entropie et problème des moments. *Ann. Inst. Henri Poincaré* **1990**, *26*, 567–596.
8. Gzyl, H.; Velásquez, Y. *Linear Inverse Problems: The Maximum entropy Connection*; World Scientific Pubs: Singapore, Singapore, 2011.
9. Gzyl, H. Maxentropic reconstruction of Fourier and Laplace transforms under non-linear constraints. *Appl. Math. Comput.* **1995**, *25*, 117–126.
10. Gamboa, F.; Gzyl, H. Maxentropic solutions of linear Fredholm equations. *Math. Comput. Model.* **1997**, *25*, 23–32.
11. Gallon, S.; Gamboa, F.; Loubes, M. Functional Calibration estimation by the maximum entropy in the mean principle. **2013**, arXiv: 1302.1158[math.ST].
12. Csiszar, I. I-divergence geometry of probability distributions and minimization problems. *Ann. Probab.* **1975**, *3*, 148–158.
13. Csiszar, I. Generalized I-projection and a conditional limit theorem. *Ann. Probab.* **1984**, *12*, 768–793.
14. Cherny, A.; Maslov, V. On minimization and maximization of entropy functionals in various disciplines. *Theory Probab. Appl.* **2003**, *17*, 447–464.
15. Varadhan, R.; Gilbert, P. An R package for solving large system of nonlinear equations and for optimizing a high dimensional nonlinear objective function. *J. Stat. Softw.* **2009**, *32*, 1–26.