Applied Mathematics and Computation 265 (2015) 225–236

Contents lists available at ScienceDirect

Applied Mathematics and Computation

journal homepage: www.elsevier.com/locate/amc

Entropy and density approximation from Laplace transforms

Henryk Gzyl[a], Pierluigi Novi Inverardi[b], Aldo Tagliani[b,*]

[a] Centro de Finanzas IESA, DF, Caracas, Venezuela
[b] Department of Economy and Management, University of Trento, 38100 Trento, Italy

ARTICLE INFO

Keywords:
Entropy convergence
Fractional moments
Kullback–Leibler distance
Laplace transform
Maximum entropy

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.

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 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 [18] and of Silva and Parada [19] 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 [7] and Gavriliadis [6] 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 [13,14] provides a procedure to recover a probability density function f_X (and the associated distribution

* Corresponding author. Tel.: +39 0461282116; fax: +39 0461282124.
E-mail addresses: henryk.gzyl@iesa.edu.ve (H. Gzyl), pierluigi.noviinverardi@unitn.it (P. Novi Inverardi), aldo.tagliani@unitn.it (A. Tagliani).

http://dx.doi.org/10.1016/j.amc.2015.05.020
0096-3003/

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 [21] provides an upper bound of $H[f_X]$ directly in terms of such moments, and Novi-Inverardi et al. [17] 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 $L(\alpha_j \geq 0)$, $j = 0, \dots, M$? 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 Section 2 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 3 we provide bound and estimate for the entropy that involves only finitely many real values of the Laplace transform. In Section 4 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 a prefixed error in terms of Kullback–Leibler distance. We devote Section 5 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 $(\mu_0, \dots, \mu_M)$, given by

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

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 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 [11].

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 $(\mu_0, \dots, \mu_M)$ of moments, when the solution f_M exists, it is an approximant to f_Y given by

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

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

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

where

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