



# Entropy Estimate of Probability Densities Having Assigned Moments: Hausdorff Case

A. TAGLIANI

Faculty of Economics, Trento University, 38100 Trento, Italy

(Received June 2000; revised and accepted May 2001)

**Abstract**—Considered here are absolutely continuous probability distributions, concentrated on the interval  $[0, 1]$ , and with the first  $M$  algebraic moments assigned. Lower and upper bounds for entropy are provided solely in terms of assigned moments. © 2002 Elsevier Science Ltd. All rights reserved.

**Keywords**—Convex hull, Entropy, Hankel matrix, Moment problem.

## 1. INTRODUCTION

Every probability distribution has some uncertainty associated with it, and its entropy provides a quantitative measure of this uncertainty. Partial information given, for instance, in terms of averages about a random variate decreases its entropy. It thus appears interesting to provide an entropy estimate when partial information is given.

A viable approach consists in use of the maximum entropy (ME) principle [1], according to which, out of all the probability distributions consistent with a given set of constraints, the one that has maximum entropy should be chosen. Such a value of entropy, obtained by the ME principle, therefore, represents an upper bound on the entropy of the underlying distribution.

In general, however, the following drawback arises: the entropy is not provided directly in terms of the given averages, but it includes the parameters of the maximum entropy distribution (equations (2.1),(2.2)). Consequently, it is desirable to obtain an entropy estimate in terms of the given averages only.

This paper attempts, under special hypotheses, to accomplish that goal by providing an upper and lower bound for absolutely continuous distributions, having assigned the first  $M$  algebraic moments  $(\mu_1, \dots, \mu_M)$ . The upper bound (equation (3.8)) will be stated under the most general hypothesis on the underlying distribution, whereas the lower bound (equation 3.14) will be provided under restrictive hypotheses. Indeed, a very sharply peaked distribution has a very low entropy, whereas if the distribution is widely spread, the entropy is higher, owing to the fact that the entropy measures the “uniformity” of a distribution.

The absolutely continuous distributions considered here are concentrated on the interval  $[0, 1]$  and have their first  $M$  moments known. This is the classical reduced Hausdorff moment problem [2] consisting of recovering an unknown probability density in  $[0, 1]$  whose first  $M$  algebraic moments are known to match the given moments.

In studies on maximum entropy, there has been a great deal of interest in this problem. Csiszar [3] stated the existence conditions, while Borwein-Lewis [4] furnished several convergence criteria, according to different hypotheses concerning the underlying density  $f(x)$ . Tagliani [5] has used a unified approach to obtain the same results of existence, convergence, and also provided a stability analysis.

As regards the entropy estimate of a distribution in terms of given moments, only a few results are known in the literature. Jardas [6] considered discrete distributions with countable range and provided an upper bound of the entropy in terms of mean value only. Cover and Thomas [7] extended this result involving the first two moments.

The present paper is both a natural continuation of [5] and an extension of previous results [6,7] to absolutely continuous distributions, taking higher moments into account. The technique used consists of a Taylor expansion of the entropy of the maximum entropy distribution (equation (2.3)) around the moments vector of the uniform distribution. Thus, the obtained entropy estimate grows increasingly sharp as the given vector of moments approach the moment vector of the uniform distribution.

The possibility of extending this technique to absolutely continuous distributions defined within an unbounded domain is not remote. In several papers, the present writer has provided the existence and convergence conditions (see [5] for references) for the corresponding ME distributions when the first algebraic moments are assigned. Hence, it should be possible to generalize the technique to an unbounded domain. Some difficulties may be raised by the paucity of results on the spectral properties of the Hankel matrices involved (equation (2.5)) [8]. Furthermore, in the case of unbounded domains, Theorem 2.2 quoted below does not hold.

## 2. SOME BACKGROUND

### 2.1. The Maximum Entropy Technique

Let  $X$  be an absolutely continuous random variable with density function  $f(x)$  whose first  $M$  moments  $(\mu_1, \dots, \mu_M)$ ,  $\mu_j = \int_0^1 x^j f(x) dx$ ,  $j = 1, \dots, M$ ,  $\mu_0 = 1$ , are assigned and the Shannon entropy is  $H[f] = - \int_0^1 f(x) \ln f(x) dx$ . Once  $(\mu_0, \dots, \mu_M)$  is given, the ME principle yields the following approximate [9] of  $f(x)$ :

$$f_M(x) = \exp \left( - \sum_{j=0}^M \lambda_j x^j \right), \quad (2.1)$$

where  $\lambda_0, \dots, \lambda_M$  are Lagrange multipliers, to be supplemented by the condition that its first  $M + 1$  moments are given by  $\mu_j$ ,  $j = 0, \dots, M$ ,

$$\mu_j = \int_0^1 x^j f_M(x) dx, \quad j = 0, \dots, M, \quad (2.2)$$

and whose entropy is

$$H[f_M] = \sum_{j=0}^M \lambda_j \mu_j. \quad (2.3)$$

### 2.2. A Differential Relationship

Varying only one moment  $\mu_i$ ,  $i = 0, \dots, M$ , while the remaining ones are held fixed, so that  $\lambda_j = \lambda_j(\mu_i)$ ,  $j = 0, \dots, M$ , differentiation of (2.2) provides

$$\Delta_{2M} \cdot \begin{bmatrix} \frac{d\lambda_0}{d\mu_i} \\ \vdots \\ \frac{d\lambda_M}{d\mu_i} \end{bmatrix} = -e_{i+1}. \quad (2.4)$$

Here  $e_{i+1}$  is the canonical unit vector  $\in \mathbb{R}^{M+1}$ , while  $\Delta_{2M}$  is the  $(M+1)$ -order symmetric definite positive Hankel matrix

$$\Delta_{2M} = \begin{bmatrix} \mu_0 & \cdots & \mu_M \\ \vdots & \ddots & \vdots \\ \mu_M & \cdots & \mu_{2M} \end{bmatrix}, \quad M = 1, 2, \dots \tag{2.5}$$

The entries of  $\Delta_{2M}$  satisfy the relationship  $\mu_0 > \mu_1 > \mu_2 > \dots$  essential in the subsequent procedure for entropy estimate.

### 2.3. The Moment Space

The moment space  $D^M \subset \mathbb{R}_+^M$  whose points are the  $n$ -ple  $(\mu_1, \dots, \mu_M)$  is the convex hull of the curve  $\{x, x^2, \dots, x^M\}$ ,  $x \in [0, 1]$ . The existence conditions of  $f_M(x)$  involve  $D^M$ , more precisely if

- (a) the point  $\mu = (\mu_1, \dots, \mu_M)$  is outside  $D^M$ , the corresponding finite Hausdorff moment problem does not admit any solution,
- (b)  $\mu \in \partial D^M$  ( $\partial D^M$  is the boundary of  $D^M$ ), the only distribution having  $(\mu_1, \dots, \mu_M)$  as its first moments is a (uniquely determined) convex combination of Dirac's delta,
- (c)  $\mu$  belongs to the interior of  $D^M$  many infinitely distributions exist, one of them being  $f_M(x)$  [10].

### 2.4. Some Known Theorems

Let  $(\mu_1, \dots, \mu_{M-1}) \in D^{M-1}$  be assigned. For a density  $f(x)$  with the same moments  $(\mu_1, \dots, \mu_{M-1})$ , let  $\mu_M^- = \inf_{f(x)} = \int_0^1 x^M f(x) dx$  and  $\mu_M^+ = \sup_{f(x)} = \int_0^1 x^M f(x) dx$  where the minimum and maximum are taken over those density functions for which the moments up to degree  $M-1$  coincide with the assigned ones. Therefore,  $\mu_M^-$  and  $\mu_M^+$  are the extremes of the  $M^{\text{th}}$  moment, and they assume a finite value because the moment space  $D^M$  is compact [10]. Likewise,  $(\mu_M^+ - \mu_M^-)$  is the width of  $D^M$  in the  $\mu_M$  direction.

The following theorems will be used.

**THEOREM 2.1.** (See [4].) Let  $(\mu_1, \dots, \mu_M)$  be assigned and let  $f_M(x)$  be the corresponding ME density. Under the sole hypothesis of the existence of  $f_M(x)$ ,  $\lim_{M \rightarrow \infty} H[f_M] = H[f]$  holds.

**THEOREM 2.2.** (See [11].) Assuming  $f(x) \geq \gamma > 0$ , with  $\gamma$  arbitrary constant, and with moments  $\mu_1, \mu_2, \dots$ , and  $\Delta_{2M}$  as the obtained Hankel matrices, then

$$\lim_{M \rightarrow \infty} (\text{Cond}_2(\Delta_{2M}))^{1/(M+1)} = \lim_{M \rightarrow \infty} (\text{Cond}_2(H_{M+1}))^{1/(M+1)} \simeq e^{3.525} \tag{2.6}$$

holds. Here, the matrix  $H_{M+1} \equiv \{h_{ij}\}$ ,  $h_{ij} = 1/(i+j+1)$ ,  $i, j = 0, \dots, M$ , is the  $(M+1)$ -order Hilbert matrix, and  $\text{Cond}_2(\cdot)$  denotes the spectral condition number, i.e., for a given matrix  $(\cdot)$ ,  $\text{Cond}_2(\cdot) = \|(\cdot)\|_2 \cdot \|(\cdot)^{-1}\|_2$ . Since  $\Delta_{2M}$  (and also,  $H_{M+1}$ ) are symmetric definite positive,  $\text{Cond}_2(\Delta_{2M}) = \|(\Delta_{2M})\|_2 \cdot \|(\Delta_{2M})^{-1}\|_2 = \lambda_{\max}(\Delta_{2M})/\lambda_{\min}(\Delta_{2M})$  holds, where  $\lambda_{\max}$  and  $\lambda_{\min}$  denote the highest and lowest eigenvalue, respectively. Theorem 2.2 essentially states that each Hankel matrix, raised by a strictly positive weight  $f(x)$ , is asymptotically conditioned like a Hilbert matrix. In the moment space  $D^M$ , the assumption  $f(x) \geq \gamma > 0$  means that point  $(\mu_1, \dots, \mu_M)$  is far from moment space boundary, or equivalently,  $f(x)$  is far from a Dirac delta-type configuration. In this case,  $H[f] \rightarrow -\infty$  holds, since  $H[f]$  is a measure of uniformity.

### 3. ENTROPY ESTIMATE

#### 3.1. An Upper Bound

I now provide an estimate of  $H[f_M]$  solely in terms of  $\mu_0, \dots, \mu_M$ . Consider the vector  $\mu^H = \{1/(j+1)\}$ ,  $j = 0, \dots, M$ , whose entries are the Hilbert matrix entries and  $\mu = (\mu_0, \dots, \mu_M)$  is the vector of given moments. When  $\mu = \mu^H$ , the Hankel matrix and the Hilbert matrix coincide, since the corresponding ME density is  $f_M^{\mu^H}(x) \equiv 1$ ,  $\forall M$ , so that  $\lambda_j = 0$ ,  $j = 0, \dots, M$ , and  $H[f_M^{\mu^H}] = 0$ . Let us examine the vector  $\xi = (\xi_0, \dots, \xi_M)$  belonging to the segment joining  $\mu^H$  to  $\mu$ . Once  $\xi$  is given, we consider the corresponding ME density  $f_M^\xi(x)$  from which the next moments  $\xi_j$ ,  $j > M$ , can be obtained. From (2.3) and (2.4), we have

$$\frac{\partial H[f_M]}{\partial \mu_k} = \sum_{j=0}^M \mu_j \frac{\partial \lambda_j}{\partial \mu_k} + \lambda_k = \lambda_k - \delta_{0k}, \quad k = 0, \dots, M, \quad (3.1)$$

where  $\delta_{0k}$  denotes  $\delta$ -Kronecker. When  $\mu = \mu^H$ , from (3.1), we have

$$\left. \frac{\partial H[f_M]}{\partial \mu_k} \right|_{\mu^H} = -\delta_{0k}, \quad k = 0, \dots, M. \quad (3.2)$$

From (3.1), and taking (2.4) into account, the Hessian matrix, evaluated at  $\xi = (\xi_0, \dots, \xi_M)$ , is obtained:

$$\begin{aligned} \left. \frac{\partial^2 H[f_M]}{\partial \mu_j \partial \mu_k} \right|_{\mu=\xi} &= \left. \frac{\partial \lambda_k}{\partial \mu_j} \right|_{\mu=\xi} = -\frac{1}{|\Delta_{2M}|_{\mu=\xi}} \\ &\times \begin{vmatrix} \xi_0 & \cdots & \xi_{k-1} & \vdots & \xi_{k+1} & \cdots & \xi_M \\ \vdots & \vdots & \vdots & e_{j+1} & \vdots & \vdots & \vdots \\ \xi_M & \cdots & \xi_{k-1+M} & \vdots & \xi_{k+1+M} & \cdots & \xi_{2M} \end{vmatrix}, \end{aligned} \quad (3.3)$$

so that the  $(j, k)$ -entry of the Hessian matrix coincides, up to sign, with the  $(j, k)$ -entry of the inverse  $\Delta_{2M}^{-1}|_\xi$  of  $\Delta_{2M}|_\xi$ . From Taylor expansion, at  $\mu = \mu^H$ , and taking (3.1)–(3.3) into account, we obtain

$$\begin{aligned} H[f_M] &= H[f_M]|_{\mu^H} + \sum_{j=0}^M \left( \mu_j - \frac{1}{1+j} \right) \left. \frac{\partial H[f_M]}{\partial \mu_j} \right|_{\mu^H} - \frac{1}{2} (\mu - \mu^H) \Delta_{2M}^{-1}|_\xi (\mu - \mu^H)^\top \\ &= -\frac{1}{2} (\mu - \mu^H) \Delta_{2M}^{-1}|_\xi (\mu - \mu^H)^\top. \end{aligned} \quad (3.4)$$

Since  $\Delta_{2M}^{-1}|_\xi$  is symmetric definite positive, then

$$\frac{1}{\lambda_{\max}(\Delta_{2M}|_\xi)} \leq \frac{(\mu - \mu^H) \Delta_{2M}^{-1}|_\xi (\mu - \mu^H)^\top}{\|\mu - \mu^H\|_2^2} \leq \frac{1}{\lambda_{\min}(\Delta_{2M}|_\xi)}, \quad (3.5)$$

and then

$$\lambda_{\max}(\Delta_{2M}|_\xi) \leq \|\Delta_{2M}|_\xi\|_1 = \sum_{j=0}^M \xi_j \leq \sum_{j=0}^M \max \left\{ \mu_j, \frac{1}{1+j} \right\} = \|\max\{\mu, \mu^H\}\|_1. \quad (3.6)$$

Combining (3.6), (3.5), and (3.4), we have

$$\frac{1}{\|\max\{\mu, \mu^H\}\|_1} \leq \frac{1}{\lambda_{\max}(\Delta_{2M}|_\xi)} \leq \frac{(\mu - \mu^H) \Delta_{2M}^{-1}|_\xi (\mu - \mu^H)^\top}{\|\mu - \mu^H\|_2^2} = -\frac{2H[f_M]}{\|\mu - \mu^H\|_2^2}. \quad (3.7)$$

From (3.7) and the ME principle,

$$H[f] \leq H[f_M] \leq -\frac{\|\mu - \mu^H\|_2^2}{2\|\max\{\mu, \mu^H\}\|_1} \quad (3.8)$$

holds for each  $M$ . Inequality (3.8) is the required upper bound of  $H[f]$  and provides an estimate of  $H[f]$  in terms of given moments only for each function  $f(x)$  with the same  $\mu = (\mu_0, \dots, \mu_M)$ .

### 3.2. An Asymptotic Lower Bound

With  $(\mu_0, \dots, \mu_{M-1})$  fixed and varying  $\mu_M$  continuously, from (2.3) and (2.4), we have

$$\frac{dH[f_M]}{d\mu_M} = \lambda_M, \tag{3.9}$$

and from (2.4),

$$\frac{d^2 H[f_M]}{d\mu_M^2} = \frac{d\lambda_M}{d\mu_M} = -\frac{|\Delta_{2(M-1)}|}{|\Delta_{2M}|} < 0, \tag{3.10}$$

so that  $H[f_M]$  is a concave differentiable function of  $\mu_M$ . When  $\mu_M \rightarrow \mu_M^\pm$ , we have from (2.4), with  $i = M$ ,  $\lambda_M \rightarrow \mp\infty$ , so that  $H[f_M] \rightarrow -\infty$ . Thus,  $H[f_M]$  may admit a lower bound if an additional hypothesis is imposed on  $f(x)$ . This lower bound will be obtained under the hypothesis  $f(x) \geq \gamma > 0$ , so that Theorem 2.2 can be used.

Since  $\Delta_{2M}|\xi$  is symmetric definite positive,

$$\text{Cond}_2(\Delta_{2M}|\xi) = \frac{\lambda_{\max}(\Delta_{2M}|\xi)}{\lambda_{\min}(\Delta_{2M}|\xi)},$$

and moreover,

$$\lambda_{\max}(\Delta_{2M}|\xi) = \|\Delta_{2M}|\xi\|_2 \geq \max_{i,j} [\Delta_{2M}|\xi] = \mu_0 = 1 \tag{3.11}$$

(here  $\max_{i,j}$  runs over the entries of  $\Delta_{2M}|\xi$ ). Combining (3.5) and (3.11) yields

$$\begin{aligned} \frac{(\mu - \mu^H) \Delta_{2M}^{-1}|\xi (\mu - \mu^H)^\top}{\|\mu - \mu^H\|_2^2} &\leq \lambda_{\max}(\Delta_{2M}|\xi) \cdot \frac{(\mu - \mu^H) \Delta_{2M}^{-1}|\xi (\mu - \mu^H)^\top}{\|\mu - \mu^H\|_2^2} \\ &\leq \frac{\lambda_{\max}(\Delta_{2M}|\xi)}{\lambda_{\min}(\Delta_{2M}|\xi)} = \text{Cond}_2(\Delta_{2M}|\xi). \end{aligned} \tag{3.12}$$

We prove now that  $\Delta_{2M}|\xi$  satisfies the hypothesis of Theorem 2.2 when  $\xi$  runs along the segment joining  $\mu^H$  to  $\mu$  in the moment space  $D^M$ . Let us fix  $\xi$  and consider the corresponding ME density  $f_M^\xi(x) = \exp(-\sum_{j=0}^M \lambda_j x^j)$ . When  $\xi \equiv \mu^H$  or  $\xi \equiv \mu$ , the Lagrange multipliers  $\lambda_j$ ,  $j = 0, \dots, M$ , assume finite values. Varying  $\xi$  continuously along the segment joining  $\mu^H$  to  $\mu$ , all the Lagrange multipliers vary continuously and then assume finite values. Thus, the corresponding density  $f_M^\xi(x)$  satisfies the condition  $f_M^\xi(x) \geq \gamma > 0$ .

When  $M \rightarrow \infty$ , Theorem 2.2 yields  $\text{Cond}_2(\Delta_{2M}|\xi) = \text{Cond}_2(H_{M+1}) \simeq e^{3.525(M+1)}$  and Theorem 2.1 yields  $H[f_M] \simeq H[f]$ . Thus, from (3.4) and (3.12), we have

$$\begin{aligned} -\frac{2H[f]}{\|\mu - \mu^H\|_2^2} &\simeq -\frac{2H[f_M]}{\|\mu - \mu^H\|_2^2} = \frac{(\mu - \mu^H) \Delta_{2M}^{-1}|\xi (\mu - \mu^H)^\top}{\|\mu - \mu^H\|_2^2} \\ &\leq \text{Cond}_2(\Delta_{2M}|\xi) \simeq e^{3.525(M+1)}, \end{aligned} \tag{3.13}$$

from which,

$$H[f] > -\frac{e^{3.525(M+1)}}{2} \|\mu - \mu^H\|_2^2, \tag{3.14}$$

providing an asymptotic lower bound of  $H[f]$ . Equations (3.8) and (3.14) are the main results.

## REFERENCES

1. E.T. Jaynes, Where do we stand on maximum entropy, In *The Maximum Entropy Formalism*, (Edited by R.D. Levine and M. Tribus), pp. 15–118, MIT Press, Cambridge, MA, (1978).
2. J.A. Shohat and J.D. Tamarkin, *The Problem of Moments, Volume 1*, AMS Mathematical Survey, Providence, RI, (1943).
3. I. Csiszar,  $I$ -divergence geometry of probability distributions and minimization problems, *The Annals of Probability* **1**, 146–158 (1975).
4. J.M. Borwein and A.S. Lewis, Convergence of best entropy estimate, *SIAM J. Optimization* **1**, 191–205 (1991).
5. A. Tagliani, Hausdorff moment problem and maximum entropy: A unified approach, *Applied Math. and Computation* **105**, 291–305 (1999).
6. C. Jardas, J. Pecaric, R. Roki and N. Sarapa, On some inequalities for entropies of discrete probability distributions, *J. Australian Mathematical Society, Ser. B*, **40**, 535–541 (1999).
7. T.M. Cover and J.A. Thomas, *Elements of Information Theory*, John Wiley & Sons, (1991).
8. E.E. Tyrtysnikov, How bad are Hankel matrices?, *Numerische Mathematics* **67**, 261–269 (1994).
9. H.K. Kesavan and J.N. Kapur, *Entropy Optimization Principles with Applications*, Academic Press, (1992).
10. S. Karlin and L.S. Shapley, *Geometry of Moment Spaces, Volume 12*, AMS Memoirs, Providence, RI, (1953).
11. D. Fasino, The spectral properties of Hankel matrices and numerical solution of finite moment problem, *J. of Computational and Applied Mathematics* **65**, 145–155 (1995).