A GENERAL MAXIMUM ENTROPY THEORY: CONSTRAINT OPTIMIZATION IN PROBABILITY

BHASKAR BHATTACHARYA* AND JOHN GREGORY
DEPARTMENT OF MATHEMATICS, SOUTHERN ILLINOIS UNIVERSITY,
CARBONDALE, IL 62901

Abstract. A new theory and methods of solving variational problems of probability are given and applied to generalized maximum entropy problems. The cost functional and "moments" are generalized and the solution is efficiently found by now established analytical techniques or efficient and accurate numerical methods. It is shown that the classical maximum entropy problem is an immediate example of this theory.

To illustrate the type of results for the general setting, the results for the classical problem are developed. This will give the reader a good idea as to what to expect for more general problems.

In our chosen context with two moments, a simple second order differential equation in five dependent variables with appropriate boundary conditions, is obtained, which leads to the unique solution. In addition, previous work by the second author gives an efficient numerical algorithm with a priori, global, error of $O(h^2)$, where h is the node size.

Alternately, it is shown that this problem has a less familiar solution of a second order ODE with two parameters, each of which is associated with a moment isoperimetric constraint. Finally, because these problems fit within the second author's theory of constraint optimization, we can easily add additional reasonable constraints, inequality constraints, etc. to these problems.

1. Introduction

In his delightful book where he uses the ideas of probability to "cut a broad swath through the physical sciences", Coles [6] highlights the maximum energy principle for continuous parameters as a lack of information. Indeed, these ideas are a major part of the modern day sciences.

In Cole's notation, he has

$$S = -\int p(x) \log \frac{p(x)}{m(x)}\, dx \tag{1.1}$$

where m "normalizes" S under the expected moment constraints such as

$$\int x\, p(x)\, dx = \mu \text{ and } \int (x-\mu)^2 p(x)\, dx = \sigma^2. \tag{1.2}$$

2000 *Mathematics Subject Classification.* Primary 49K15. Secondary 49S05, 65K10.
Key words and phrases. Maximum energy, constraint optimization, entropy.
*Supported in part by NSF grant DMS - 0706041.

Even more complete in the discussion of this topic is the current edition of Wikipedia, the free encyclopedia on the web. They refer to [7], but clearly the importance of the problem can be found throughout the literature of applied mathematics.

The purpose of this paper is to give the theory for a broader interpretation to Problem (1.1) which includes this problem as a specific example. We will replace the integrand with $f(t, x_1, x_1', x_2, x_2')$ of two dependent variables and (1.2) by more generalized "moments". This problem will then be in a classical calculus of variations setting or, more specifically, a constraint optimization setting [11] and [13]. To aid the reader we will provide results and comments for Problem (1.1) as an example.

It is first worthwhile to consider even the classical role of "entropy". The role of entropy in statistical theory and information processing is well documented by Karlin and Rinott [18], Kullback [20],Marshall and Olkin [21], Rao [23], Rényi [24]. For applications of entropy in other areas such as thermodynamic systems, ecological structures, see Kapur [17] and Karlin and Rinott [19].

When selecting a model for a given situation it is often appropriate to express the prior information in terms of constraints. However, one must be careful so that no information other than these specified constraints is used in model selection. That is, other than the constraints that we have, the uncertainty associated with the probability distribution to be selected should be kept at its maximum. This is the 'principle of maximum entropy' advocated by Jaynes [15], and later treated axiomatically by Shore and Johnson [25].

Consider the set of constraints

$$\mathcal{C} = \{p(\boldsymbol{x}) : E_p[T_i(\boldsymbol{X})] = t_i, i = 0, \cdots, n\}$$

where T_i are integrable functions, t_i are known constants and $T_0(\boldsymbol{x}) = t_0 = 1$. The maximum entropy principle finds the unknown probability density function $p^*(\boldsymbol{x})$ which maximizes the entropy subject to the constraints in $\mathcal{C}$. This procedure has been shown to characterize most well known univariate probability distributions, e.g., see Kagan *et al.* [16], Kapur [17], Guiasu [9], Preda [22], and the references therein. Although, literature is significantly less for the multivariate distributions, Kapur [17] considered several usual multivariate distributions, Zografos [26] considered the cases of Pearson's type II and VII multivariate distributions, Bhattacharya [4] considered characterization for the multivariate Liouville distribution, and Aulogiaris and Zografos [3] considered symmetric Kotz type and Burr multivariate distributions. Expressions for entropies for several known and relatively unknown multivariate distributions can be found in Zografos and Nadarajah [27], Ahmed and Gokhale [1] and Darbellay and Vajda [8].

In Section 2 we develop the major mathematical results for the general problem using the second author's theory of constrained optimization [11] and [13]. In Section 3 we consider the "classical" maximum entropy problem as a special example of Section 2. We get a second order ODE with two parameters, which are associated with the isoperimetric moment constraints (1.2). This implies a unique solution to our boundary value problem as we have four conditions. We also "obtain" the well-known solution $p(t) = e^{-t^2/2}$ for the interval $(-\infty, \infty)$ by an easy extension to $a = -\infty$ and $b = \infty$.

Hopefully, the interested reader can now solve his/her own problem whatever the conditions on f or the number of "moments".

In Section 4 we indicate how to efficiently solve these problems numerically with a global, a priori error of $O(h^2)$. The perceptive reader may be surprised to see that for the classical problem in Section 3 we obtain the values of the two isoperimetric constraints along with the numerical solution. This is because our numerical problem involves five dependent variables and not just one.

To reiterate ... As an example of our ideas for these general problems, the solutions to the classical problem can be found as a five dependent variable problem with 10 boundary values or as a two-point second order boundary value problem in one dependent variable with two parameters. For a quick introduction to the subject, we recommend the reader to be familiar with [11]. This would make it easier for anyone to go through the subject matter of this paper.

2. The Mathematical Development

The purpose of this section is to develop the key mathematical ideas. We will first state our problem and then obtain our second order ODE as a boundary value problem.

Our problem is

$$\left.\begin{array}{l} \min \int_a^b f\big(t, x_1(t), x_1'(t), x_2(t), x_2'(t)\big)\, dt, \quad f_{x_1' x_1'} > 0 \\ \text{s.t. } x_2' = g(t, x_1, x_1');\ x_1 \text{ in BC.} \end{array}\right\} \tag{2.1}$$

and f, g are continuous functions of their arguments, x_1, x_2 are continuous with derivatives which are piecewise smooth. In the above, BC means that x_1 is given or unspecified at either endpoint and we assume there is a unique minimum solution.

We will call x_2 a *moment function* for x_1 if the equality in (2.1) holds and $x_2(a) = 0$. This clearly generalizes the usual situation as in (1.2). If we choose to add additional moment constraints, it will be immediate, as in Section 3.

To obtain a solution for (2.1) we proceed as follows:

Following the procedures in [11] and [13], we define $X(t) = \big(x_1(t), x_2(t), x_3(t)\big)^T$ where $x_1(t) = x(t)$, $x_2(t) = \int_a^t g\big(s, x_1(s), x_1'(s)\big)\, ds$, $x_3'(t)$ is a multiplier as indicated below,

$$\left.\begin{array}{c} F\big(t, X(t), X'(t)\big) = f + x_3'[x_2' - g], \text{ with} \\ X(a) = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \quad \text{and} \quad X(b) = \begin{pmatrix} 1 \\ M \\ * \end{pmatrix} \end{array}\right\} \tag{2.2}$$

The perceptive reader will note that we "associate" $x_1(t)$ with the cumulative distribution function and $x_1'(t)$ with the probability density function. The first two components of $X(a)$ and $X(b)$ are now immediate. The third component of $X(a)$ is to "normalize" x_3, in that we only need the unique values of x_3' but can/will find $x_3(t)$. Finally, "$*$" refers to the fact that we can not specify $x_3(b)$ instead must use the transversality procedure, below.

We next make the following calculations.

$$F(t, X, X') = f(t, x_1, x_2, x_1', x_2') + x_3'\big[x_2' - g(t, x_1, x_1')\big],$$

so that

$$\left.\begin{aligned} F_X &= \begin{pmatrix} f_{x_1} - g_{x_1} x_3' \\ f_{x_2} \\ 0 \end{pmatrix}, \quad F_{X'} = \begin{pmatrix} f_{x_1'} - g_{x_1'} x_3' \\ f_{x_2'} + x_3' \\ x_2' - g(t, x_1, x_1') \end{pmatrix}, \\ \text{and} \quad F_{X'X'} &= \begin{pmatrix} f_{x_1'x_1'} - g_{x_1'x_1'} x_3' & 0 & -g_{x_1'} \\ 0 & f_{x_2'x_2'} & 1 \\ -g_{x_1'} & 1 & 0 \end{pmatrix}. \end{aligned}\right\} \tag{2.3}$$

We note that $F_{X'X'}$ is, of course, symmetric and that det $F_{X'X'} = -f_{x_2'x_2'} g_{x_1'}^2 - f_{x_1'x_1'} + g_{x_1'x_1'} x_3'$, which we assume is not identically zero on any subinterval.

Lemma 2.1. If det $F_{X'X'} \not\equiv 0$ on any subinterval, then the Euler equation $\frac{d}{dt} F_{X'} = F_X$ is a regular, second order differential equation. It, along with the boundary conditions in (2.2), has a unique solution. □

In the classical case(s) where $f = x' \ln x'$ and $g = a(t)x'$, a cofactor expansion implies $\det F_{X'X'} = \frac{1}{x_1'}$ and the stated results.

We next note that under this assumption, the Bliss Multiplier Rule ideas hold, in that this problem is nonsingular in the sense of Bliss [5]. By [11] and [13], the basic necessary conditions for (2.2) hold. That is

Theorem 2.2. *There exists a unique solution to* (2.1) *such that*

$$\frac{d}{dt} F_{X'} = F_x\,, \tag{2.4a}$$

$$F_{X'}^T\, X\big|_a^b = 0, \text{ and} \tag{2.4b}$$

$$F_{X'} \text{ is continuous at points where } X'(t) \text{ is not continuous.} \quad \square \tag{2.4c}$$

We note that these three conditions are the conditions which give the solution when one exists.

If we regard this as a problem in three dependent variables, it is “complete”. However, it is instructive to note that these general problems can be reduced to a second order differential equation in one variable which is not ordinary. Thus, using the results of this theorem, we have

(i) $x_3' - g(t, x_1, x_1') \equiv 0$ by the use of the third component of (2.4a), the fact it is piecewise continuous by (2.4c), and it is 0 at b by (2.4b) and the $(*)$ condition. Thus, $x_3(t) = \int_a^t g(s, x, x')\, ds$ by use of the $X(a)$ third component of in (2.2) and hence $\int_a^b g(t, x, x')\, dt = M$. (2.5a)

From the second component we have

$$\text{(ii)} \qquad x_3' = \int_a^t f_{x_2} ds - f_{x_2'} \tag{2.5b}$$

and finally we have the second order equation for $x = x_1$,

$$\text{(iii)} \qquad \frac{d}{dt}\left[f_{x_1'} - g_{x_1'} \int_a^t f_{x_2}\, ds + g_{x_1'} f_{x_2'} \right] = f_{x_1} - g_{x_1} \int_a^t f_{x_2}\, ds + g_{x_1} f. \tag{2.5c}$$

Theorem 2.3. *The solution to* (2.1) *is as follows: As a problem with three dependent variables, its solution is given by the second order differential equation*

$$\frac{d}{dt} f_{X'} = f_X$$

in (2.3) *with boundary conditions* (2.2). *In addition, in one dependent variable it is given by* (2.5c) *with* x *in BC.* □

As expected, the one variable solution causes additional problems. In three variables the differential equation is ordinary and there are efficient and accurate numerical methods with an a priori global error of $O(h^2)$ as described in [11] and [12]. The arguments in one variable involve the antiderivative so at a minimum, (2.5c) will require additional differentiation to easily solve that equation.

3. The Classical Problems

In this section we consider the classical problem:

$$\begin{aligned} &\min \int_a^b x'(t) \ln x'(t)\, dt \\ &\text{s.t. } \int_a^b t\, x'(t)\, dt = M_1 \text{ and } \int_a^b t^2 x'(t)\, dt = M_2, \text{ and } x(a) = 0,\ x(b) = 1. \end{aligned} \tag{3.1}$$

We note that f is only a function of x_1' and that there are now two constraints or moments but they are of the form $x_2' = g_1(t, x_1') = t x_1'$ and $x_3' = g_2(t, x_1') = t^2 x_1'$, that is of the form $a(t)x_1'$.

Following the procedures given above, we define $X(t) = \big(x_1(t), x_2(t), x_3(t), x_4(t), x_5(t)\big)^T$ where $x_1(t) = x(t)$, $x_2(t) = \int_a^t s\, x_1'(s)\, ds$, $x_3(t) = \int_a^t s^2 x_1'(s)\, ds$, $x_4'(t)$, $x_5'(t)$ are multipliers we have,

$$F\big(t, X(t), X'(t)\big) = x_1'(t) \ln x_1'(t) + x_4'(t)\big[x_2'(t) - t\, x_1'(t)\big] + x_5'(t)\big[x_3'(t) - t^2 x_1'(t)\big]$$

$$X(a) = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \quad \text{and} \quad X(b) = \begin{pmatrix} 1 \\ M_1 \\ M_2 \\ * \\ * \end{pmatrix}. \tag{3.2}$$

We again note that we "associate" $x_1(t)$ with the cumulative distribution function, $x_1'(t)$ with the probability density function while $x_2' - t\, x_1' = 0$ and $x_3' - t^2\, x_1' = 0$ are associated with the first and second "moment integrands", respectively. The first three components of $X(a)$ and $X(b)$ are now immediate. The fourth and fifth components of $X(a)$ are to "normalize" x_4 and x_5, in that we only seek the unique values of x_4' and x_5' but can/will find $x_4(t)$ and $x_5(t)$. Finally, "$*$" refers to the fact that we can not specify $x_4(b)$ and $x_5(b)$ and instead must use the transversality procedure, below.

We next make the following calculations.

$$F_X = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \quad F_{X'} = \begin{pmatrix} \ln x_1' + 1 - t\, x_4' - t^2 x_5' \\ x_4' \\ x_5' \\ x_2' - t\, x_1' \\ x_3' - t^2 x_1' \end{pmatrix},$$

$$\text{and} \quad F_{X'X'} = \begin{pmatrix} \frac{1}{x_1'} & 0 & 0 & -t & -t^2 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ -t & 1 & 0 & 0 & 0 \\ -t^2 & 0 & 1 & 0 & 0 \end{pmatrix}, \tag{3.3}$$

and have

Lemma 3.1. Det $F_{X'X'} \not\equiv 0$, except at isolated points, and hence, the Euler equation $\frac{d}{dt}F_{X'} = F_X$ is a regular, second order differential equation. It, along with the boundary conditions in (3.2), has a unique solution. □

A cofactor expansion implies det $f_{X'X'} = \frac{1}{x_1'}$ and the stated results. The reader is invited to observe why this result is much simpler than in the general problem of Section 2.

Theorem 3.2. *There exists a unique solution to* (3.1) *such that*

$$\frac{d}{dt}F_{X'} = F_x\,, \tag{3.4a}$$

$$F_{X'}^T\, X\big|_a^b = 0, \text{ and} \tag{3.4b}$$

$F_{X'}$ *is continuous at points where* $X'(t)$ *is not continuous.* □ (3.4c)

We note that these three conditions are the conditions which give the solution. To aid the reader in perhaps unfamiliar ground, we repeat ideas of the previous section. Otherwise, these specific results are immediate from Section 2.

Using the results of this theorem, we have

(i) $x_3' - t^2x_1' \equiv 0$ by the use of fifth component of (3.4a), the fact it is piecewise continuous by (3.4c), and it is 0 at b by (3.4b) and the $(*)$ condition. Thus, $x_3(t) = \int_a^t s^2x_1'(s)\,ds$ by use of the third component of $X(a)$ in (3.2) and the fact that $\int_a^b t^2x'(t)\,dt = M_2\,.$ (3.5a)

Similar reasoning yields

(ii) $x_2' = t\,x_1'$ and, hence, $x_2(t) = \int_a^t s\,x_1'(s)\,ds$ and $\int_a^b t\,x_1(t)\,dt = M_1\,.$ (3.5b)

We also have

(iii) $x_5'' = 0$ and $x_5(t) = c_2(t-a)$ by the fifth component of $X(a)$. (3.5c)

Similarly,

(iv) $x_4(t) = d_2(t-a)$. (3.5d)

Once again, we invite the reader to consider why the special form in (3.5c), (3.5d), and (3.5e) below

$$\text{(v)} \quad \frac{d}{dt}\left[(\ln x' + 1) - d_2 t - c_2 t^2\right] = 0\,. \tag{3.5e}$$

The fact that the multipliers x_4' and x_5' are constants agrees with the classical isoperimetric result [2]. Finally, (3.5e) is a 4-parameter boundary value problem including M_1 and M_2. We have

$$\frac{d}{dt}(\ln x') = d_2 + 2c_2 t \tag{3.6}$$

which implies that

$$\begin{aligned} \ln x' &= d_2 t + c_2 t^2 + d_1 \text{ and} \\ x(t) &= \int_a^t e^{d_1 + d_2 s + c_2 s^2}\, ds + c_1\,. \end{aligned} \tag{3.7}$$

By the boundary conditions, $c_1 = 0$. In addition, we use $x(a) = 0$ and $x(b) = 1$, which are the familiar results from continuous probability and M_1 and M_2 to uniquely determine the unique values of c_2, d_1 and d_2.

An easy argument extends these results to the interval $(-\infty, \infty)$. Thus, $x'(t) = e^{c_2(t-c_3)^2}$ where $c_2 < 0$. But the moment constraints in (3.1) immediately imply that $c_3 = 0$ and $c_2 = -\frac{1}{2}$. This is the known, classical result that $p(t) = e^{-t^2/2}$ in this case.

4. The Numerical Problem

The purpose of this section is to consider numerical solutions for (2.1) The numerical extension to a finite number of moments and any interval $[a, b]$ will be immediate and left to the reader.

Our first task is to give the relevant theory from [11, Formulas (6.9) or (6.29), and (6.3)] or [12], and then we will explicitly describe (4.1) and (4.2), below. Thus,

Theorem 4.1. *Let $F(t,X,X')$ be given as in (2.2), $h>0$, $Nh=b-a$ and X_k, $k=0,1,\dots,N$, be a sequence of values which satisfy*

$$\begin{aligned} &F_{X'}\left(a_{k-1}^*, \frac{X_k+X_{k-1}}{2}, \frac{X_k-X_{k-1}}{h}\right) \\ &+\frac{h}{2}F_X\left(a_{k-1}^*, \frac{X_k+X_{k-1}}{2}, \frac{X_k-X_{k-1}}{h}\right) \\ &-F_X'\left(a_k^*, \frac{X_k+X_{k+1}}{2}, \frac{X_{k-1}-X_k}{h}\right) \\ &+\frac{h}{2}F_X\left(a_k^*, \frac{X_k+X_{k-1}}{2}, \frac{X_{k+1}-X_k}{h}\right)=0 \\ &\qquad\qquad \text{for } k=1,2,\dots,N-1 \end{aligned} \tag{4.1}$$

and

$$\begin{aligned} &F_{X'}\left(a_{N-1}^*, \frac{X_N+X_{N-1}}{2}, \frac{X_N-X_{N-1}}{h}\right) \\ &+\frac{h}{2}F_X\left(a_{N-1}^*, \frac{X_N+X_{N-1}}{2}, \frac{X_N-X_{N-1}}{h}\right)=0 \end{aligned} \tag{4.2}$$

where $a_k^*=(a_k+a_{k+1})/2$ with $a_k=a+k\,h$.

Let $X(t)$ be the solution to Problem (2.1) and $\{X_k\}$ be a computed, numerical solution of (4.1) and (4.2). Then there exists a constant $C>0$, independent of $h>0$, so that for h sufficintly small

$$\|X(a_k)-X_k\|\le Ch^2. \quad \square$$

We note that (4.1) "solves" the two point boundary value problem where $X(a)$ and $X(b)$ are given. The combination of (4.1) and (4.2) "solves" transversality problem at $t=b$.

Finally, to aid the reader in using these techniques, we return to $F_{X'}$ in (2.3) and look at the first term for a specific value of k. The remaining steps in applying our methods will be immediate and left to the reader. Hence,

$$F_{X'}\left(a_{k-1}^*, \frac{X_k+X_{k-1}}{2}, \frac{X_k-X_{k-1}}{h}\right) = \begin{pmatrix} \ln\left(\frac{x_{1k}-x_{1k-1}}{h}\right)+1-\left(a_k-\frac{h}{2}\right)\left(\frac{x_{4k}-x_{4k-1}}{h}\right)-\left(a_k-\frac{k}{2}\right)^2\left(\frac{x_{5k}-x_{5k-1}}{h}\right) \\ \frac{x_{4k}-x_{4k-1}}{h} \\ \frac{x_{5k}-x_{5k-1}}{h} \\ \frac{x_{2k}-x_{2k-1}}{h}-\left(a_k-\frac{h}{2}\right)\frac{x_{1k}-x_{1k-1}}{h} \\ \frac{x_{3k}-x_{3k-1}}{h}-\left(a_k-\frac{h}{2}\right)^2\left(\frac{x_{1k}-x_{1k-1}}{h}\right) \end{pmatrix}$$

We note that (4.1), (4.2) are a set of $5(N+1)$ equations in $5(N+1)$ unknowns with a unique solution by Lemma 2.1 and Theorem 2.2.

Acknowledgement

The authors thank the referee for useful comments which led to improvements in its presentation.

References

[1] N.A. Ahmed and D.V. Gokhale, Entropy expressions and their estimators for multivariate distributions, IEEE. Trans. Inform. Theory **35** (1989), 688–692.
[2] Arfken, G., *Mathematical Methods for Physicists*, Academic Press, Inc., Orlando, FL, 1985.
[3] G. Aulogiaris and K. Zografos, *A maximum entropy characterization of symmetric Kotz type and multivariate Burr distributions*, Test **13** (2004), 65–83.
[4] Bhattacharya, B., *Maximum entropy characterization for the multivariate Louisville distribution*, Journal of Multivariate Analysis, **97** (2006), 1272-1283.
[5] Bliss, G.A., *Lectures on the Calculus of Variations*, Phoenix Science Series, University of Chicago Press, Chicago, IL, 1963.
[6] Coles, P., *From Cosmos to Chaos: The Science of Unpredictability*, Oxford University Press, New York, 2006.
[7] Cover, T.M. and Thomas, J.A., *Elements of Information Theory*, Chapter 11, John Wiley & Sons, 1991.
[8] G.A. Darbellay and T. Vajda, *Entropy expressions for multivariate continuous distributions*, IEEE. Trans. Inform. Theory **46** (2000), 709–712.
[9] S. Guiasu, *A classification of the main probability distributions by minimizing the weighted logarithmic measure of deviation*, Ann. Inst. Statist. Math. **42** (1990), 269–279.
[10] Gregory, J., A New, Systematic Method for Efficiently Solving Holonomic (and Nonholonomic) Constraint Problems, in preparation.
[11] Gregory, J. and Lin, C., *Constrained Optimization in the Calculus of Variations and Optimal Control Theory*, Van Nostrand Reinhold, New York, 1992.
[12] Gregory, J. and Lin, C., *Discrete variable metohds for the m-dependent variable, non-linear extremal problem in the calculus of variations, II*, SIAM J. Numer. Anal. **30** (1993), 871–883.
[13] Gregory, J. and Lin, C., *An unconstrained calculus of variations formulation for generalized optimal control problems and for the constrained problem of Bolza*, J. Math. Anal. Appl. **187** (1994), 826–841.
[14] Hestenes, M.R., *Calculus of Variations and Optimal Control Theory*, John Wiley and Sons: New York, 1966.
[15] E.T. Jaynes, *Information theory and statistical mechanics*, Phys. Rev., **106** (1957), 620–630.
[16] A.M. Kagan, Y.V. Linnik, C.R. Rao, *Characterization Problems in Mathematical Statistics*, Translated from Russian text by B. Ramachandran, Wiley, New York, 1973.
[17] J.N. Kapur, *Maximum Entropy Models in Engineering*, Wiley, New York, 1989.
[18] S. Karlin and Y. Rinott, *Entropy inequalities for classes of probability distributions-II, The multivariate case*, Adv. Appl. Probab. **13** (1981), 325–351.
[19] S. Karlin and Y. Rinott, *Classes of orderings of measures and related correlation inequalities-I, Multivariate totally positive distributions*, J. Multivariate Anal. **10** (1980), 467–498.
[20] S. Kullback, *Information Theory and Statistics*, Wiley, New York, 1959.
[21] A.W. Marshall and I. Olkin, *Theory of Majorization and its Applications*, Wiley, New York, 1979.
[22] V. Preda, *The student distribution and the maximum entropy*, Ann. Inst. Statist. Math. **34** (1982), 335–338.
[23] C.R. Rao, *Linear Statistical Inference and Its Applications*, Wiley, New York, 1973.

[24] A. Rényi, *On measures of entropy and information*, Proc. 4th Berkeley Sympos. Math. Stat. Probab. Vol. 1 (1961), 541–561.

[25] J.E. Shore and R.W. Johnson, *Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy*, IEEE Trans. Inform. Theory **27** (1980), 26–37.

[26] K. Zografos, On maximum entropy characterization of Pearson's type II and VII multivariate distributions, J. Multivariate Anal. 71 (1999), 67-75.

[27] K. Zografos and S. Nadarajah, *Expressions for Rényi and Shannon entropies for multivariate distributions*, Statist. & Probab. Lett. **71** (2005), 71–84.

[28] O. Vasicek, *A test for normality based on sample entropy*, J. Roy. Statist. Soc. B **38** (1976), 54–59.