Maximum Entropy Characterizations of the Multivariate Liouville Distributions

Bhaskar Bhattacharya

Department of Mathematics, Southern Illinois University

Carbondale, IL 62901-4408, USA

Abstract

A random vector $\boldsymbol{X} = (X_1, X_2, \cdots, X_n)$ with positive components has a Liouville distribution with parameter $\boldsymbol{\theta} = (\theta_1, \theta_2, \cdots, \theta_n)$ if its joint probability density function is proportional to $h(\sum_{i=1}^n x_i) \prod_{i=1}^n x_i^{\theta_i - 1}$, $\theta_i > 0$ (Gupta and Richards [9]). Examples include correlated gamma variables, Dirichlet and inverted Dirichlet distributions. We derive appropriate constraints which establish the maximum entropy characterization of the Liouville distributions among all multivariate distributions. Matrix analogs of the Liouville distributions are considered. Some interesting results related to I-projection from a Liouville distribution are presented.

Key words and phrases: Dirichlet distribution, gamma variables, I-projections, Inverted Dirichlet distribution, maximum entropy principle, Shannon entropy.

1 Introduction

Let $\boldsymbol{X} = (X_1, X_2, \cdots, X_n)$ be a random vector with probability density function $p(\boldsymbol{x})$. The *entropy* (or Shannon entropy) of $\boldsymbol{X}$ which measures the uncertainty associated with $p(\boldsymbol{x})$ is

$$H(p) = -\int p(\boldsymbol{x}) \ln p(\boldsymbol{x}) d\boldsymbol{x}.$$

The role of entropy in statistical theory and information processing is well documented by Karlin and Rinott [14], Kullback [16], Marshall and Olkin [17], Rao [21], Rényi [22]. For applications of entropy in other areas such as thermodynamic systems, ecological structures, see Kapur [13] and Karlin and Rinott [15].

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 [11], and later treated axiomatically by Shore and Johnson [23].

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.* [12], Kapur [13], Guiasu [8], Preda [20], and the references therein. Although, literature is significantly less for the multivariate distributions, Kapur [13] considered several usual multivariate distributions, Zografos [25] considered the cases of Pearson's type II and VII multivariate distributions and Aulogiaris and Zografos [2] 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 [26], Ahmed and Gokhale [1] and Darbellay and Vajda [5].

In this paper, we obtain the Liouville distributions by using the maximum entropy principle. A comprehensive treatment of the Liouville distributions (including many of its characterization properties) has appeared in Gupta and Richards [9]. These distributions are also treated briefly by Marshall and Olkin [17] and Sivazlian [24]. Peddada and Richards [9] study the Schur-concavity and Schur-convexity of its entropy functional.

In Section 2, we derive the appropriate constraints which characterize the Liouville distributions among all multivariate distributions. Different Liouville distributions are generated with different choices of the function h (see (2.1) below). Examples are considered with special choices of the function h for well known cases and the corresponding characterizations are derived. It has been shown (Bhattacharya and Dykstra [3], among others) that the maximum entropy problem is essentially an I-projection problem. In Section 3, we show that the I-projection of a Liouville distribution onto a certain intersection of convex sets is a1so a Liouville distribution. In addition, we show that the I-projection of a Liouville distribution of the first kind onto the set of all probability measures (PM's) with support on the simplex $\mathcal{S}_n = \{(x_1, x_2, \cdots, x_n),\ x_i > 0,\ i = 1, \cdots, n; \sum_{i=1}^n x_i < 1\}$ is a Liouville distribution of the second kind.

2 Characterization Results

A continuous random vector $\boldsymbol{X} = (X_1, X_2, \cdots, X_n)$ has a (multivariate) Liouville distribution if its joint density function is given by

$$p(\boldsymbol{x};\boldsymbol{\theta}) = c_n(\boldsymbol{\theta}) h\left(\sum_{i=1}^n x_i\right) \prod_{i=1}^n x_i^{\theta_i - 1}, \tag{2.1}$$

where $\boldsymbol{x} = (x_1, x_2, \cdots, x_n),\ x_i > 0,\ i = 1, \cdots, n;\ h : \mathcal{R}_+ \to \mathcal{R}_+$ is continuous and the parameter $\boldsymbol{\theta} = (\theta_1, \theta_2, \cdots, \theta_n) \in \mathcal{R}_+^n$. The normalizing constant $c_n(\boldsymbol{\theta})$ is given by

$$[c_n(\boldsymbol{\theta})]^{-1} = \frac{\prod_{i=1}^n \Gamma(\theta_i)}{\Gamma(\sum_{i=1}^n \theta_i)} \int_0^\infty t^{\sum_{i=1}^n \theta_i - 1} h(t) dt, \tag{2.2}$$

see Edwards [6], Marshall and Olkin [17], Karlin and Rinott [15].

The first step to establish the characterization is to find $E_p\left[\ln\left[h\left(\sum_{i=1}^n X_i\right)\right)\right]$ and $E_p\left[\ln\ X_i\right]$ for a given function h for the $p(\boldsymbol{x};\boldsymbol{\theta})$ in (2.1). These are presented in Lemmas 2.1 and 2.2 below in a more general setting.

Lemma 2.1. *For fixed $\alpha > 0$ and $\beta_i > -1,\ \forall i$,*

$$\int_0^\infty \cdots \int_0^\infty \left[\ln\left(h\left(\sum_{i=1}^n x_i\right)\right)\right] h^\alpha\left(\sum_{i=1}^n x_i\right) \prod_{i=1}^n x_i^{\beta_i} dx_1 \cdots dx_n$$

$$= \frac{\prod_{i=1}^n \Gamma(\beta_i + 1)}{\Gamma(\sum_{i=1}^n \beta_i + n)} \int_0^\infty [\ln h(t)] t^{\sum_{i=1}^n \beta_i + n - 1} h^\alpha(t) dt. \tag{2.3}$$

Proof. Defining $I(\beta, g) = \int_0^\infty t^\beta g(t) dt$ as in Peddada and Richards [19], we get from (2.1) and (2.2) that

$$\int_0^\infty \cdots \int_0^\infty h\left(\sum_{i=1}^n x_i\right) \prod_{i=1}^n x_i^{\theta_i - 1} dx_1 \cdots dx_n = \frac{\prod_{i=1}^n \Gamma(\theta_i)}{\Gamma(\sum_{i=1}^n \theta_i)} I\left(\sum_{i=1}^n \theta_i - 1, h\right). \tag{2.4}$$

Applying (2.4) to the function h^α and using $\theta_i = \beta_i + 1$, we get

$$\int_0^\infty \cdots \int_0^\infty h^\alpha\left(\sum_{i=1}^n x_i\right) \prod_{i=1}^n x_i^{\beta_i} dx_1 \cdots dx_n = \frac{\prod_{i=1}^n \Gamma(\beta_i + 1)}{\Gamma(\sum_{i=1}^n \beta_i + n)} I\left(\sum_{i=1}^n \beta_i + n - 1, h^\alpha\right). \tag{2.5}$$

Differentiating both sides of (2.5) with respect to α we get the result. □

The next lemma is used to find $E[\ln X_j]$, for any $j = 1, \ldots, n$.

LEMMA 2.2. *For fixed $\alpha > 0$ and $\beta_i > -1$, $\forall i$,*

$$\int_0^\infty \cdots \int_0^\infty [\ln(x_j)] h^\alpha \left(\sum_{i=1}^n x_i\right) \prod_{i=1}^n x_i^{\beta_i} dx_1 \cdots dx_n$$

$$= A^{-1} \left[\Psi(\beta_j + 1) - \Psi\left(\sum_{i=1}^n \beta_i + n\right) + \frac{\int_0^\infty [\ln t] t^{\sum_{i=1}^n \beta_i + n - 1} h^\alpha(t) dt}{\int_0^\infty t^{\sum_{i=1}^n \beta_i + n - 1} h^\alpha(t) dt} \right] \tag{2.6}$$

$\forall j$, where A^{-1} is the right side of (2.5) and $\Psi(t) = (d/dt) \ln \Gamma(t)$ is the digamma function.

PROOF. Begin as in the proof of Lemma 2.1, and consider (2.5). Differentiating both sides of (2.5) with respect to β_j (for any j, $1 \le j \le n$) we get

$$\int_0^\infty \cdots \int_0^\infty [\ln(x_j)] h^\alpha \left(\sum_{i=1}^n x_i\right) \prod_{i=1}^n x_i^{\beta_i} dx_1 \cdots dx_n$$

$$\begin{aligned}
&= \frac{\partial}{\partial \beta_j} [A]^{-1} \\
&= -\frac{1}{A} \frac{\partial}{\partial \beta_j} (\ln[A]) \\
&= -\frac{1}{A} \frac{\partial}{\partial \beta_j} \left[\ln\left(\frac{\Gamma(\sum_{i=1}^n \beta_i + n)}{\prod_{i=1}^n \Gamma(\beta_i + 1)} \right) - \ln I\left(\sum_{i=1}^n \beta_i + n - 1, h^\alpha\right) \right] \\
&= -\frac{1}{A} \left[\frac{\partial}{\partial \beta_j} \left[\ln\left(\Gamma(\sum_{i=1}^n \beta_i + n)\right)\right] - \frac{\partial}{\partial \beta_j} \sum_{i=1}^n \ln \Gamma(\beta_i + 1) \right. \\
&\qquad \left. - \frac{\partial}{\partial \beta_j} \ln\left[I\left(\sum_{i=1}^n \beta_i + n - 1, h^\alpha\right)\right] \right] \\
&= -\frac{1}{A} \left[\Psi\left(\sum_{i=1}^n \beta_i + n\right) - \Psi(\beta_j + 1) - \frac{I\left(\sum_{i=1}^n \beta_i + n - 1, h_1\right)}{I\left(\sum_{i=1}^n \beta_i + n - 1, h^\alpha\right)} \right]
\end{aligned}$$

where $h_1(t) = (\ln t) h^\alpha(t)$, from which the desired result follows. □

Of course, to apply these two lemmas for the density function in (2.1), we need $\alpha = 1$ and $\beta_i = \theta_i - 1$. Using these values to the right sides of (2.3) and (2.6), we obtain after simplifications $E_p\left[\ln\left(h\left(\sum_{i=1}^n X_i\right)\right)\right] = w(\boldsymbol{\theta})$, $E_p[\ln X_i] = u_i(\boldsymbol{\theta})$, $\forall i$ where

$$\begin{aligned}
w(\boldsymbol{\theta}) &= \frac{I\left(\sum_{i=1}^n \theta_i - 1, h_2\right)}{I\left(\sum_{i=1}^n \theta_i - 1, h\right)}, \\
u_i(\boldsymbol{\theta}) &= \Psi(\theta_i) - \Psi\left(\sum_{i=1}^n \theta_i\right) + \frac{I\left(\sum_{i=1}^n \theta_i - 1, h_3\right)}{I\left(\sum_{i=1}^n \theta_i - 1, h\right)},
\end{aligned} \tag{2.7}$$

$\forall i$, where $h_2(t) = h(t) \ln h(t)$, $h_3(t) = h(t) \ln t$.

The next theorem shows that the Liouville distribution has maximum entropy in the class of all probability distributions specified by (2.8).

THEOREM 2.3. *The random vector* $\boldsymbol{X}$ *with a Liouville distribution with parameter* $\boldsymbol{\theta}$ *has the maximum entropy among all positive, absolutely continuous random vectors* $\boldsymbol{Y} = (Y_1, \ldots, Y_n)$ *with pdf* $p(\boldsymbol{y})$ *subject to the restrictions*

$$\begin{aligned} E_p \left[\ln \left(h \left(\textstyle\sum_{i=1}^n Y_i\right)\right)\right] &= w(\boldsymbol{\theta}), \\ E_p \left[\ln\ Y_i\right] &= u_i(\boldsymbol{\theta}),\ \forall i \end{aligned} \tag{2.8}$$

where $w(\boldsymbol{\theta})$, $u_i(\boldsymbol{\theta})$ *are defined in (2.7).*

PROOF. It is well known using Jensen's inequality that for PM's P, Q, R,

$$\int \frac{dP}{dR} \ln \frac{dP/dR}{dQ/dR} dR \geq 0.$$

Hence,

$$\int \frac{dP}{dR} \ln \frac{dP}{dR} dR \geq \int \frac{dP}{dR} \ln \frac{dQ}{dR} dR.$$

Now consider a PM Q which satisfies the constraints (2.8) and has pdf of the form

$$\frac{dQ}{dR} = \exp \left[a_0 + a_1 \ln h \left(\sum_{i=1}^n X_i\right) + \sum_{i=1}^n b_i \ln\ X_i\right] \tag{2.9}$$

(R is the Lebesgue measure on $\mathcal{R}^n$) where a_0, a_1, b_i are chosen so that Q is a PM. Then we get

$$\begin{aligned} -\textstyle\int \frac{dP}{dR} \ln \frac{dP}{dR} dR &\leq -\textstyle\int \frac{dP}{dR} \ln \frac{dQ}{dR} dR \\ &= -\textstyle\int \frac{dP}{dR} \left(a_0 + a_1 \ln h \left(\sum_{i=1}^n X_i\right) + \sum_{i=1}^n b_i \ln\ X_i\right) dR \\ &= -\left(a_0 + a_1 w(\boldsymbol{\theta}) + \textstyle\sum_{i=1}^n b_i u_i(\boldsymbol{\theta})\right). \end{aligned} \tag{2.10}$$

The upper bound for entropy on the last line of (2.10) is achieved by the Liouville distribution, and its p.d.f. is of the form (2.9). Thus the Liouville distribution has the maximal entropy among all distributions satisfying the constraints (2.8). □

The last line of (2.10) also gives an expression for the entropy of the Liouville distribution, which after simplification yields the following expression

$$\sum_{i=1}^{n} \ln \Gamma(\theta_i) - \ln\left(\Gamma\left(\sum_{i=1}^{n}\theta_i\right)\right) + \ln\left(I\left(\sum_{i=1}^{n}\theta_i - 1, h\right)\right)$$
$$-\frac{I\left(\sum_{i=1}^{n}\theta_i - 1, h_2\right) + \left(\sum_{i=1}^{n}\theta_i - n\right) I\left(\sum_{i=1}^{n}\theta_i - 1, h_3\right)}{I\left(\sum_{i=1}^{n}\theta_i - 1, h\right)}$$
$$-\sum_{i=1}^{n}(\theta_i - 1)\Psi(\theta_i) + \left(\sum_{i=1}^{n}\theta_i - n\right)\Psi\left(\sum_{i=1}^{n}\theta_i\right).$$

An expression for the same entropy appears in Zografos and Nadarajah [26], but there appears to be some misprint present in their expression.

The density in (2.1) uses $h(\cdot)$ on a noncompact support, and is called the Liouville distribution of the *first kind* (Gupta and Richards, [9]). The Liouville distribution of the *second kind* arises when $h(\cdot)$ is supported on $(0,1)$ and the variables range on the simplex $\mathcal{S}_n = \{(x_1, x_2, \cdots, x_n),\ x_i > 0,\ i = 1, \cdots, n; \sum_{i=1}^{n} x_i < 1\}$. The expression in (2.2) and the results above can be duplicated in this case with the range of integrations replaced by 0 to 1.

Now we consider the matrix analogs of the Liouville distributions. The positive definite (symmetric) $m \times m$ matrices $\boldsymbol{A}_1, \ldots, \boldsymbol{A}_n$ are said to have a Liouville distribution of the *first kind* if their (continuous) joint density function exists and is given by

$$p(\boldsymbol{A};\boldsymbol{\theta}) = d_n(\boldsymbol{\theta}) g\left(\sum_{i=1}^{n}\boldsymbol{A}_i\right)\prod_{i=1}^{n}|\boldsymbol{A}_i|^{\theta_i - p}, \tag{2.11}$$

where $|\boldsymbol{A}|$ is the determinant of $\boldsymbol{A}$, $g : \mathcal{R}_+^{m\times m} \to \mathcal{R}_+$ is continuous and satisfies $\int_{\mathcal{R}_+^{m\times m}} |\boldsymbol{T}|^{\sum\theta_i - p} g(\boldsymbol{T}) d\boldsymbol{T} < \infty$, $p = (m+1)/2, \theta_i > p-1, i = 1, \cdots, n$, and the normalizing constant $d_n(\boldsymbol{\theta})$ is given by

$$[d_n(\boldsymbol{\theta})]^{-1} = \frac{\prod_{i=1}^{n}\Gamma_m(\theta_i)}{\Gamma_m(\sum_{i=1}^{n}\theta_i)}\int_{\mathcal{R}_+^{m\times m}} |\boldsymbol{T}|^{\sum\theta_i - p} g(\boldsymbol{T}) d\boldsymbol{T}, \tag{2.12}$$

where $\Gamma_m(\cdot)$ is the multidimensional gamma function (James, [10]).

We extend the results derived above to the matrix analogs of the Liouville distributions as follows. The proofs follow in a similar way, and hence they are omitted. First the relation (2.5) generalizes to

$$\int_{\mathcal{R}_+^{m\times m}} \cdots \int_{\mathcal{R}_+^{m\times m}} g^{\alpha}\left(\sum_{i=1}^{n} \boldsymbol{A}_i\right) \prod_{i=1}^{n} |\boldsymbol{A}|_i^{\beta_i} d\boldsymbol{A}_1 \cdots d\boldsymbol{A}_n$$

$$= \frac{\prod_{i=1}^{n} \Gamma_m(\beta_i + p)}{\Gamma(\sum_{i=1}^{n} \beta_i + np)} \int_{\mathcal{R}_+^{m\times m}} |\boldsymbol{T}|^{\sum_{i=1}^{n} \beta_i + (n-1)p} g^{\alpha}(\boldsymbol{T}) d\boldsymbol{T}. \qquad (2.13)$$

Next two lemmas generalize Lemma 2.1 and Lemma 2.2.

LEMMA 2.4. *For fixed $\alpha > 0$ and $\beta_i > -1$, $\forall i$,*

$$\int_{\mathcal{R}_+^{m\times m}} \cdots \int_{\mathcal{R}_+^{m\times m}} \left[\ln\left(g\left(\sum_{i=1}^{n} \boldsymbol{A}_i\right)\right)\right] g^{\alpha}\left(\sum_{i=1}^{n} \boldsymbol{A}_i\right) \prod_{i=1}^{n} |\boldsymbol{A}_i|^{\beta_i} d\boldsymbol{A}_1 \cdots d\boldsymbol{A}_n \qquad (2.14)$$

$$= \frac{\prod_{i=1}^{n} \Gamma_m(\beta_i + p)}{\Gamma_m(\sum_{i=1}^{n} \beta_i + np)} \int_{\mathcal{R}_+^{m\times m}} [\ln g(\boldsymbol{T})] |\boldsymbol{T}|^{\sum_{i=1}^{n} \beta_i + (n-1)p} g^{\alpha}(\boldsymbol{T}) d\boldsymbol{T}.$$

LEMMA 2.5. *For fixed $\alpha > 0$ and $\beta_i > -1$, $\forall i$,*

$$\int_{\mathcal{R}_+^{m\times m}} \cdots \int_{\mathcal{R}_+^{m\times m}} [\ln(|\boldsymbol{A}_j|)] g^{\alpha}\left(\sum_{i=1}^{n} \boldsymbol{A}_i\right) \prod_{i=1}^{n} |\boldsymbol{A}_i|^{\beta_i} d\boldsymbol{A}_1 \cdots d\boldsymbol{A}_n$$

$$= B^{-1}\left[\Psi_m(\beta_j + p) - \Psi_m\left(\sum_{i=1}^{n} \beta_i + np\right) + \frac{\int_{\mathcal{R}_+^{m\times m}} [\ln(|\boldsymbol{T}|)] |\boldsymbol{T}|^{\sum_{i=1}^{n} \beta_i + (n-1)p} g^{\alpha}(\boldsymbol{T}) d\boldsymbol{T}}{\int_{\mathcal{R}_+^{m\times m}} |\boldsymbol{T}|^{\sum_{i=1}^{n} \beta_i + (n-1)p} g^{\alpha}(\boldsymbol{T}) d\boldsymbol{T}}\right] \qquad (2.15)$$

$\forall j$, where B^{-1} is the right side of (2.13) and $\Psi_m(t) = (d/dt) \ln \Gamma_m(t)$.

Let $I_m(\beta, f) = \int_{\mathcal{R}_+^{m\times m}} |\boldsymbol{T}|^{\beta} f(\boldsymbol{T}) d\boldsymbol{T}$. Now apply Lemma 2.4 and Lemma 2.5 with $\alpha = 1$ and $\beta_i = \theta_i - p$ for the density function in (2.11). Using these values to the right sides of (2.14) and (2.15), we obtain $E_p[\ln(g(\sum_{i=1}^{n} \boldsymbol{A}_i))] = w^m(\boldsymbol{\theta})$, $E_p[\ln\ \boldsymbol{A}_i] = u_i^m(\boldsymbol{\theta})$, $\forall i$ where

$$\begin{aligned} w^m(\boldsymbol{\theta}) &= \frac{I_m\left(\sum_{i=1}^{n} \theta_i - p, g_2\right)}{I_m\left(\sum_{i=1}^{n} \theta_i - p, g\right)}, \\ u_i^m(\boldsymbol{\theta}) &= \Psi_m(\theta_i) - \Psi_m\left(\sum_{i=1}^{n} \theta_i\right) + \frac{I_m\left(\sum_{i=1}^{n} \theta_i - 1, g_3\right)}{I_m\left(\sum_{i=1}^{n} \theta_i - 1, g\right)}, \end{aligned} \qquad (2.16)$$

$\forall i$, where $g_2(\boldsymbol{T}) = g(\boldsymbol{T}) \ln g(\boldsymbol{T})$, $g_3(\boldsymbol{T}) = g(\boldsymbol{T}) \ln |\boldsymbol{T}|$.

The next theorem generalizes the maximum entropy characterization for the Liouville distribution.

THEOREM 2.6. *The random matrices $\boldsymbol{A}_1, \ldots, \boldsymbol{A}_n$ with a Liouville distribution with parameter $\boldsymbol{\theta}$ has the maximum entropy among all positive, absolutely continuous random matrices $\boldsymbol{Y} = (\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_n)$ with pdf $p(\boldsymbol{y})$ subject to the restrictions*

$$\begin{aligned} E_p \left[\ln \left(g \left(\textstyle\sum_{i=1}^n \boldsymbol{Y}_i\right)\right)\right] &= w^m(\boldsymbol{\theta}), \\ E_P \left[\ln \ \boldsymbol{Y}_i\right] &= u_i^m(\boldsymbol{\theta}), \ \forall i \end{aligned} \tag{2.17}$$

where $w^m(\boldsymbol{\theta})$, $u_i^m(\boldsymbol{\theta})$ are defined in (2.16).

An expression for the entropy for the matrix analog of the Liouville distribution is obtained from Theorem 2.6, which after simplification yields

$$\begin{aligned} &\sum_{i=1}^n \ln \Gamma_m(\theta_i) - \ln \left(\Gamma_m \left(\sum_{i=1}^n \theta_i\right)\right) + \ln \left(I_m \left(\sum_{i=1}^n \theta_i - p, g\right)\right) \\ &- \frac{I_m \left(\sum_{i=1}^n \theta_i - p, g_2\right) + \left(\sum_{i=1}^n \theta_i - p\right) I \left(\sum_{i=1}^n \theta_i - p, g_3\right)}{I \left(\sum_{i=1}^n \theta_i - p, g\right)} \\ &- \sum_{i=1}^n (\theta_i - p) \left[\Psi_m(\theta_i) - \Psi_m \left(\sum_{i=1}^n \theta_i\right)\right]. \end{aligned}$$

To the best of our knowledge, the above expression has not previously been derived in the literature.

The matrix Liouville distribution of the *second kind* are those for which $\boldsymbol{I} - \sum_{i=1}^n \boldsymbol{A}_i$ is also positive definite, where $\boldsymbol{I}$ is the $m \times m$ identity matrix. Similar results hold for this case also.

Example 1. (Correlated gamma variables) Let $(X_1, \ldots, X_n)$ has a Liouville distribution of the first kind with $h(t) = t^{a-1} e^{-bt}, t > 0, a > 0, b > 0$. Then the joint density function is proportional to

$$\left(\sum_{i=1}^n x_i\right)^{a-1} \prod_{i=1}^n x_i^{a-1} e^{-bx_i}.$$

The constraints that characterize this distribution are obtained from (2.8). By setting $\theta_i = a$, $\forall i$ and using the formula for the gamma integrals, the expressions in (2.7) are simplified as follows

$$\begin{aligned} w(a,b) &= (a-1)\Psi(k) - (a-1)\ln b - k, \\ u_i(a,b) &= \Psi(a) - \Psi(na) + \Psi(k) - \ln b, \end{aligned}$$

$i = 1, \ldots, n$, where $k = na + a - 1$.

Example 2. (Dirichlet distribution) Let $(X_1, \ldots, X_n)$ has a Liouville distribution of the second kind with $h(t) = (1-t)^{a_{n+1}-1}, 0 < t < 1, a_{n+1} > 0$. Here the joint density function is proportional to

$$\left(1 - \sum_{i=1}^{n} x_i\right)^{a_{n+1}-1} \prod_{i=1}^{n} x_i^{a_i - 1}.$$

The constraints that characterize this distribution are in (2.8), where the expressions for $w(\boldsymbol{\theta})$, $u_i(\boldsymbol{\theta})$ can be simplified by setting $\theta_i = a_i$, $i = 1, \ldots, n$ and considering derivatives of the beta integrals as follows

$$\begin{aligned} w(a_1, \ldots, a_{n+1}) &= (a_{n+1} - 1)[\Psi(a_{n+1}) - \Psi(\textstyle\sum_{i=1}^{n+1} a_i)], \\ u_i(a_1, \ldots, a_{n+1}) &= \Psi(a_i) - \Psi(\textstyle\sum_{i=1}^{n+1} a_i), \end{aligned}$$

$i = 1, \ldots, n$.

Example 3. (Inverted Dirichlet distribution) Let $(X_1, \ldots, X_n)$ has a Liouville distribution of the first kind with $h(t) = (1+t)^{-\sum_{i=1}^{n+1} a_i}, t > 0, a_{n+1} > 0$. Here the joint density function is proportional to

$$\left(1 + \sum_{i=1}^{n} x_i\right)^{-\sum_{i=1}^{n+1} a_i} \prod_{i=1}^{n} x_i^{a_i - 1}.$$

The constraints that characterize this distribution are in (2.8), where the expressions for $w(\boldsymbol{\theta})$, $u_i(\boldsymbol{\theta})$ can be simplified by setting $\theta_i = a_i$, $i = 1, \ldots, n$, with change of variables and techniques similar to Example 2. It follows that

$$\begin{aligned} w(a_1, \ldots, a_{n+1}) &= (\textstyle\sum_{i=1}^{n+1} a_i) \left[\Psi(a_{n+1}) - \Psi(\textstyle\sum_{i=1}^{n+1} a_i)\right], \\ u_i(a_1, \ldots, a_{n+1}) &= \Psi(a_i) - \Psi(a_{n+1}), \end{aligned}$$

$i = 1, \ldots, n$.

3 I-projections from Liouville distribution

It is well known that finding the distribution with maximum entropy is the same as finding the I-projection of the Lebesgue measure onto the set of constraints of interest (Bhattacharya and Dykstra [3], among others). In this section we show that the I-projection of a Liouville distribution onto certain intersection of sets of PM's is also a Liouville distribution. We also show that many Liouville distributions of the second kind can be expressed as an I-projection of corresponding Liouville distributions of the first kind.

For two PM's P and Q defined on an arbitrary measurable space $(\mathcal{X}, \mathcal{B})$, the *I-divergence* or the *Kullback-Leibler distance* between P and Q is defined as

$$I(P|Q) = \begin{cases} \int \ln\,(dP/dQ)\,dP, & \text{if } P \ll Q, \\ +\infty, & \text{elsewhere.} \end{cases}$$

If R is any PM with $P \ll R$, $Q \ll R$, then $I(P|Q)$ can equivalently be expressed as $I(P|Q) = \int (dP/dR) \ln((dP/dR)/(dQ/dR))\,dR$. Here and in the sequel we observe the conventions that $\ln 0 = -\infty,\ \ln(a/0) = +\infty,\ 0 \cdot (\pm\infty) = 0$.

Although $I(P|Q)$ is not a metric, it is always nonnegative and equals 0 if and only if $P = Q$ (a.e.). Hence it is often interpreted as a measure of 'divergence' or 'distance' between P and Q. Other popular names of $I(P|Q)$ are information for discrimination, cross-entropy, information gain, etc. For a given Q and a specified set of PM's $\mathcal{C}$, it is often of interest to find the $R \in \mathcal{C}$ which satisfies

$$I(R|Q) = \inf_{P \in \mathcal{C}} \int \ln(dP/dQ)\,dP \quad (< \infty). \tag{3.1}$$

This R is called the *I-projection* of Q onto $\mathcal{C}$. Csiszár (1975) has shown that R exists uniquely if $\mathcal{C}$ is variation-closed and there exists $P \in \mathcal{C}$ such that $I(P|Q) < \infty$.

Let the underlying probability space be denoted by $(\Omega, \mathcal{F}, Q)$ where Ω is the sample space, $\mathcal{F}$ is the σ-field of subsets of Ω, and Q is a given PM defined on elements of $\mathcal{F}$.

If we define the convex function f on $L_1(Q)$ as $f(x) = \int x \ln x \, dQ$ if $x \geq 0, \int x \, dQ = 1$, and $= +\infty$, otherwise, and $\mathcal{C}_0 = \{x \in L_1(Q) : x = dP/dQ, \text{for some} P \in \mathcal{C}\}$, then (3.1) and

$$\inf_{x \in \mathcal{C}_0} f(x) \tag{3.2}$$

are equivalent problems. A subset K of Ω is said to be a *cone* if $x \in K \Rightarrow \alpha x \in K, \ \forall \alpha \geq 0$. For an arbitrary subset S of $L_1(Q)$, the *dual* cone of S is given by $S^{\oplus} = \left\{ y \in \overline{M}(\Omega, \mathcal{F}) : \int xy \, dQ \geq 0, \text{ for all } x \in S \right\}$ where $\overline{M}(\Omega, \mathcal{F})$ is the set of extended valued, $\mathcal{F}$-measurable functions on Ω. We first state the following theorem from Bhattacharya and Dykstra [3] which provides a way to find the solution to an I-projection problem. In short, the theorem states that if we know the solution to the dual problem $\inf_{y \in \mathcal{C}^{\oplus}} \int e^y \, dQ$, say y_0, then the solution to the I-projection problem (3.2) can be expressed in terms of y_0. This theorem is useful when y_0 is easy to find.

THEOREM 3.1 *Assume $\mathcal{L}$ is a subset of $L_1(Q)$ and y_0 is a solution to $\inf_{y \in \mathcal{L}^{\oplus}} \int e^y \, dQ < \infty$. Then a PM R solves (3.1) where $dR/dQ = x_0 = e^{y_0} / \int e^{y_0} \, dQ$ (equivalently, x_0 solves (3.2)) if $x_0 \in \mathcal{L}$. If $\mathcal{L}$ is either (1) convex, variation-closed and contained in dom(f) or (2) a variation-closed, convex cone, then $x_0 \in \mathcal{L}$, and hence x_0 must solve (3.2).* □

Let Q be the Liouville distribution with pdf given by (2.1). Let $\mathcal{P}$ be the class of all PM's P where $(X_1, \ldots, X_n) \sim P$ with $X_i > 0, \forall i$. One might want to approximate Q subject to the constraints in $\mathcal{S} = \mathcal{C} \cap \mathcal{C}_1 \cap \ldots \cap \mathcal{C}_n$ where $\mathcal{C} = \{P \in \mathcal{P} : E_P[\ln(h(\sum_{i=1}^n X_i))] = k\}$, $\mathcal{C}_i = \{P \in \mathcal{P} : E_P[\ln X_i] = k_i\}$, $\forall i = 1, \ldots, n$, where k, k_i are given constants. To solve this problem we consider finding the I-projection of Q onto $\mathcal{S}$. The next result shows that this I-projection is also a Liouville distribution but possibly with a different h function.

PROPOSITION 3.2. *The I-projection of a given Liouville distribution Q onto $\mathcal{S}$ is also* *a Liouville distribution.*

PROOF. Expressing $\mathcal{C} = \{P : \int[\ln(h(\sum_{i=1}^n X_i)) - k](dP/dQ)dQ = 0\}$ and $\mathcal{C}_i = \{P : \int[\ln X_i - k_i](dP/dQ)dQ = 0\}$, $\forall i = 1, \ldots, n$, it follows that the dual cone of $\mathcal{S}$ is given by

$$\mathcal{S}^{\oplus} = \left\{\alpha\left(\ln\left(h\left(\sum_{i=1}^n X_i\right)\right) - k\right) + \sum_{i=1}^n \beta_i\left((\ln X_i) - k_i\right) : \alpha \in \mathcal{R}, \beta_i \in \mathcal{R}, \forall i\right\}.$$

The dual problem can be expressed as

$$\min_{\alpha\in\mathcal{R},\beta_i\in\mathcal{R},\forall i} c_n(\boldsymbol{\theta}) \int_0^\infty \cdots \int_0^\infty h^{\alpha+1}\left(\sum_{i=1}^n x_i\right) \prod_{i=1}^n x_i^{\beta_i+\theta_i-1} dx_1 \cdots dx_n$$

$$= \min_{\alpha\in\mathcal{R},\beta_i\in\mathcal{R},\forall i} c_n(\boldsymbol{\theta}) \frac{\prod_{i=1}^n \Gamma(\beta_i+\theta_i)}{\Gamma(\sum_{i=1}^n \beta_i + \sum_{i=1}^n \theta_i + n)} \int_0^\infty t^{\sum_{i=1}^n \beta_i + \sum_{i=1}^n \theta_i + n - 1} h^{\alpha+1}(t) dt.$$

If $(\alpha^*, \beta_1^*, \ldots, \beta_n^*)$ solves the above dual problem, then the solution to the I-projection problem by Theorem 3.1 is given by

$$p^*(\boldsymbol{x}; \boldsymbol{\theta}) = c_n^*(\boldsymbol{\theta}) h^{\alpha^*+1}\left(\sum_{i=1}^n x_i\right) \prod_{i=1}^n x_i^{\theta_i+\beta_i^*-1}$$

where $c_n^*(\boldsymbol{\theta})$ is the appropriate normalizing constant. Since $p^*(\boldsymbol{x}; \boldsymbol{\theta})$ is also a pdf of the form (2.1), the result follows. $\square$

Next we show that there exists a Liouville distribution of the second kind which may be expressed as an I-projection of a Liouville distribution of the first kind onto the set of all PM's with support on the simplex $\mathcal{S}_n$.

PROPOSITION 3.3. *The I-projection of a Liouville distribution of the first kind onto the set of all PM's with support on $\mathcal{S}_n$ is a Liouville distribution of the second kind.*

PROOF. If we like to restrict Q, a Liouville distribution of the first kind with pdf (2.1), to have support $\mathcal{S}_n$, it would be natural to consider the conditional distribution of Q restricted to the support $\mathcal{S}_n$. To find this restricted distribution, we find the I-projection of Q onto $\mathcal{S} = \{P : P(\mathcal{S}_n) = 1\}$. Since $\mathcal{S} = \{P : \int(I_{\mathcal{S}_n} - 1)dP = 0\}$

where $I_{\mathcal{S}_n}$ is the indicator function of $\mathcal{S}_n$, the dual cone is $\mathcal{S}^{\oplus} = \{\alpha(I_{\mathcal{S}_n} - 1) : \alpha \in \mathcal{R}\}$. Thus the dual problem is

$$\inf_{\alpha \in \mathcal{R}} \int e^{\alpha(I_{\mathcal{S}_n} - 1)} dQ = Q(\mathcal{S}_n) + \inf_{\alpha \in \mathcal{R}} \int_{\mathcal{S}_n^c} e^{-\alpha} dQ,$$

which is attained at $\alpha = \infty$. Using Theorem 3.1 above, it follows that the I-projection of Q onto $\mathcal{S}$ is given by

$$\frac{dP^*}{dQ} = \frac{e^{y_0}}{\int e^{y_0}\, dQ} = \frac{1}{Q(\mathcal{S}_n)}.$$

Thus $dP^* = dQ/Q(\mathcal{S}_n)$, or P^* is a Liouville distribution of the second kind. $\square$

Thus one may find the I-projections of the Liouville distributions of the first kind in Examples 1 and 3 in Section 2 onto the set of all PM's with support on $\mathcal{S}_n$ and obtain the corresponding Liouville distributions of the second kind. These resulting I-projections would have the same expressions as the Liouville distributions of the first kind we start with but with a different normalizing constant and support restricted to $\mathcal{S}_n$. However not all Liouville distributions of the second kind may be expressed as an I-projection of a Liouville distributions of the first kind in this way. For example, in the distribution in Example 2 in Section 2, the function h would be undefined when $t > 1$, $a_{n+1} < 1$.

Characterizations of distributions are very useful in probability theory and statistics, and play an important role, among others, for the construction of goodness-of-fit tests, e.g. Vasicek [27], Gokhale [7], Mudholkar and Tian [18]. Applicability of characterizations depends on the simplicity of the criteria used. In this context, the constraints that characterize the multivariate distributions are inherently difficult to handle both from a theoretical point of view and from a practical point of view. For the Liouville distributions this will also depend on the nature of the function h. Further research is needed to investigate how the characterizations developed in this paper can be useful in practice.

Acknowledgment

The author thanks two referees for their comments which have improved this paper. Thanks are also due to Professor Philip Feinsilver for making the paper by Peddada and Richards [19] available to the author which has led to this work.

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] G. Aulogiaris and K. Zografos, A maximum entropy characterization of symmetric Kotz type and multivariate Burr distributions, Test 13 (2004), 65-83.

[3] B. Bhattacharya and R. Dykstra, A general duality approach to I-projections, J. Statist. Plan. Inf. 47 (1995), 203-216.

[4] I. Csiszár, I-divergence geometry of probability distributions and minimization problems, Ann. Probab. 3 (1975), 146-158.

[5] G.A. Darbellay and T. Vajda, Entropy expressions for multivariate continuous distributions, IEEE. Trans. Inform. Theory 46 (2000), 709-712.

[6] J. Edwards, A treatise on the integral calculus, vol. 2, McMillan, New York, 1922.

[7] D.V. Gokhale, On entropy based goodness-of-fit tests, Comput. Statist. Data Anal. 96 (1983), 157-165.

[8] 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.

[9] R.D. Gupta and D.S.P. Richards, Multivariate Liouville distributions, J. Multivariate Anal. 23 (1987), 233-256.

[10] A.T. James, Distributions of matrix variates and latents roots derived from normal samples, Ann. Math. Stat. 35 (1964), 475-501.

[11] E.T. Jaynes, Information theory and statistical mechanics, Phys. Rev., 106 (1957), 620-630.

[12] 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.

[13] J.N. Kapur, Maximum entropy models in engineering, Wiley, New York, 1989.

[14] S. Karlin and Y. Rinott, Entropy inequalities for classes of probability distributions-II, The multivariate case, Adv. Appl. Probab. 13 (1981), 325-351.

[15] 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.

[16] S. Kullback, Information Theory and Statistics, Wiley, New York, 1959.

[17] A.W. Marshall and I. Olkin, Theory of Majorization and its Applications, Wiley, New York, 1979.

[18] G.S. Mudholkar and L. Tian, An entropy based characterization of the inverse Gaussian distribution and related goodness-of-fit test, J. Statist. Plan. Inf. 102 (2002), 211-221.

[19] S.D. Peddada and D.S.P. Richards, Entropy inequalities for some multivariate distributions, J. Multivariate Anal. 39 (1991), 202-208.

[20] V. Preda, The student distribution and the maximum entropy, Ann. Inst. Statist. Math. 34 (1982), 335-338.

[21] C.R. Rao, Linear Statistical Inference and Its Applications, Wiley, New York, 1973.

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

[23] 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.

[24] B.D. Sivazlian, On a multivariate extension of the gamma and beta distributions, SIAM J. Appl. Math. 41 (1981), 205-209.

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

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

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