Symmetric Kotz type and Burr multivariate distributions: A maximum entropy characterization

G. Aulogiaris and K. Zografos[1]
Department of Mathematics, University of Ioannina, 45110 Ioannina, GREECE

Abstract

In this paper a maximum entropy characterization is presented for Kotz type symmetric multivariate distributions as well as for multivariate Burr and Pareto type III distributions. Analytical formulae for the Shannon entropy of these multivariate distributions are also derived.

Keywords and phrases: Shannon Entropy, Maximum Entropy Principle, Kotz type multivariate distribution, Burr distribution, Pareto type III distribution.

1 Introduction

The maximum entropy method is a well-known approach to produce the unknown probability density function f, compatible to new information about f in the form of constraints on expected values. Although entropy maximization was first formulated in terms of thermodynamic entropy, the principle of maximum entropy was first introduced as a general method of inference by Jaynes (1957) and it was axiomatically characterized by Shore and Johnson (1980). It has been successfully applied in a remarkable variety of fields and has been also used for the characterization of several standard probability distributions (cf. Kapur (1989), Guiasu (1990), Gzyl (1995)).

Consider a $p-$variate random vector $\mathbf{X}^t = (X_1, ..., X_p)$, with unknown density f. Although f is unknown, suppose that we have access to some information about this density, formulated in terms of a set of information constraints on expected values. Consider the class of $p-$variate density functions $\mathcal{F} = \{f(\mathbf{x}) : E_f[T_i(\mathbf{X})] = \alpha_i,\ i = 0, 1, ..., m\}$, where T_i, $i = 0, 1, ..., m$, are absolutely integrable functions with respect to f and $T_0(\mathbf{x}) = \alpha_0 = 1$. We suppose further that the values of α_i and the form of $T_i, i = 0, 1, ..., m$, are known. The maximum entropy principle suggests to derive the unknown density function of the random vector $\mathbf{X}$, by the model that maximizes the Shannon entropy

$$H(\mathbf{X}) = -\int f(\mathbf{x}) \log f(\mathbf{x}) d\mathbf{x}, \tag{1}$$

subject to the information constraints that define the class $\mathcal{F}$. Jaynes states that the maximum entropy distribution, obtained by this constrained maximization problem, "is

[1] Corresponding author.
E-mail address: kzograf@cc.uoi.gr

the only unbiased assignement we can make; to use any other would amount to arbitrary assumption of information which by hypothesis we do not have." (Jaynes (1957), p. 623).

A considerable part of the literature related to the principle of maximum entropy is devoted to the maximum entropy characterization of the main univariate probability distributions. In the work of Kagan *et al.* (1973), Preda (1982), Bad Dumitrescu (1986), Kapur (1989), Guiasu (1977, 1990), Ebrahimi (2000), Kotz *et al.* (2000) and the references therein, the main univariate probability distributions have been reobtained by maximizing the Shannon entropy, subject to various types of constraints expressed by mean values of random variables. Comparatively little is the literature dealing with the characterization of multivariate distributions by means of the maximum entropy principle. The main reference, from this point of view, is the book of Kapur (1989) which devotes Chapters 4 and 5 for the characterization of some multivariate distributions, and the paper by Zografos (1999) where Pearson's Type II and VII multivariate distributions have reobtained by means of the maximum entropy principle.

In this paper, following Zografos (1999), we will concentrate on the characterization of Kotz type symmetric multivariate distributions as well as Burr and Pareto type III multivariate distributions. Analytical formulae for the Shannon entropy of these multivariate distributions are derived.

2 Symmetric Kotz type multivariate distribution

The $p-$variate random vector $\mathbf{X}^t = (X_1, ..., X_p)$ in R^p is said to have a symmetric Kotz type multivariate distribution, if the density function of $\mathbf{X}$ is defined by

$$f(\mathbf{x}) = C_p \left|\Sigma\right|^{-1/2} \left[(\mathbf{x}-\boldsymbol{\mu})^t\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})\right]^{m-1} \exp\left\{-r\left[(\mathbf{x}-\boldsymbol{\mu})^t\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})\right]^s\right\}, \quad (2)$$

for $r, s > 0$, $2m + p > 2$ and C_p a normalizing constant. The normalizing constant C_p is given by

$$C_p = \frac{s\Gamma(p/2)}{\pi^{p/2}\Gamma\left((2m+p-2)/2s\right)} r^{(2m+p-2)/2s}. \quad (3)$$

The parameter $\boldsymbol{\mu}$ is the mean vector $E(\mathbf{X})$ and the positive definite matrix Σ is related to the variance-covariance matrix of $\mathbf{X}$ (cf. Fang *et al.* (1990), p. 76, 77). When $m = 1$, $s = 1$ and $r = 1/2$, the distribution defined by (2) reduces to a multivariate normal distribution. The above, are particularly appealing family of distributions in constructing models in which the usual normality assumption is not satisfied.

In order to give a maximum entropy characterization of the density function defined by (2) we need the following lemmas. The proof of Lemma 1 is outlined in the Appendix.

Lemma 1. Let R^p is the $p-$dimensional Euclidean space. Then

a) $\int\limits_{R^p} (\mathbf{y}^t\mathbf{y})^{-\mu} \exp[-\kappa(\mathbf{y}^t\mathbf{y})^s]dy = \frac{\pi^{p/2}\Gamma((p-2\mu)/2s)}{s\Gamma(p/2)}\kappa^{(2\mu-p)/2s}, \quad \kappa > 0,\ \mu < \frac{p}{2}.$

b) $\int\limits_{R^p} (\mathbf{y}^t\mathbf{y})^{s-\mu} \exp[-\kappa(\mathbf{y}^t\mathbf{y})^s]dy = \frac{\pi^{p/2}\Gamma(((p-2\mu)/2s)+1)}{s\Gamma(p/2)}\kappa^{((2\mu-p)/2s)-1}, \quad \kappa > 0,\ \mu < s + \frac{p}{2}.$

c) $$\int\limits_{R^p} (\mathbf{y}^t\mathbf{y})^{-\mu} \exp[-\kappa(\mathbf{y}^t\mathbf{y})^s] \log(\mathbf{y}^t\mathbf{y})dy = \frac{\pi^{p/2}\kappa^{(2\mu-p)/2s}}{s^2\Gamma(p/2)} \times [\Gamma'((p-2\mu)/2s) - \log(\kappa)\Gamma((p-2\mu)/2s)],$$

for $\kappa > 0$, $\mu < \frac{p}{2}$. Γ denotes the gamma function and $\Gamma'(t) = (d/dt)\Gamma(t)$.

Lemma 2. For fixed $\alpha > 0$, consider the function $w(x;\alpha)$ of x, defined by

$$w(x;\alpha) = \Psi(x) - \log(\alpha x), \quad \text{for } x > 0,$$

with $\Psi(t) = (d/dt)\log\Gamma(t)$, being the digamma function. The equation $w(x;\alpha) = w(x_0;\alpha)$, has the unique solution $x = x_0$.

Proof. For $x > 0$ consider the function $\varphi(y) = x/(y+x)^2$, $y > 1$. It is obvious that φ is continuous, positive and decreasing in $y > 1$. Hence based on the Cauchy's integral test we have that

$$\sum_{k=1}^{\infty} \varphi(k) - \int_1^{\infty} \varphi(y)dy > 0 \qquad \text{or} \qquad \sum_{k=0}^{\infty} \frac{x}{(k+x)^2} > 1. \tag{4}$$

On the other hand it is well-known that $\frac{d}{dx}\Psi(x) = \sum\limits_{k=0}^{\infty} \frac{1}{(k+x)^2}$. Therefore, based on (4)

$$\frac{d}{dx}w(x;\alpha) = \sum_{k=0}^{\infty} \frac{1}{(k+x)^2} - \frac{1}{x} > 0, \quad x > 0,$$

which means that $w(x;\alpha)$ is strictly increasing in $x > 0$, which completes the proof of the lemma. ■

The following lemma proves that Shannon's entropy, given by (1), is not invariant under linear, non-singular transformations of the random vector $\mathbf{X}$. The proof is immediately obtained from the more general result that the Shannon entropy is not invariant under an invertible transformation of the variables (cf. Darbellay and Vajda (2000)).

Lemma 3. Suppose that $\mathbf{Y}$ is a $p-$variate random vector, $\mathbf{A}$ a non singular square matrix of order p and $\mathbf{u}$ a fixed $p-$dimensional vector. Then

$$H(\mathbf{AY}+\mathbf{u}) = \log|\det(\mathbf{A})| + H(\mathbf{Y}).$$

Theorem 1. Let $\mathbf{Y}^t = (Y_1, ..., Y_p)$ be a $p-$variate random vector in R^p with density g. Let also

$$E_g\left[(\mathbf{Y}^t\mathbf{Y})^s\right] = \frac{2m+p-2}{2sr}, \tag{C1}$$

and

$$E_g[\log(\mathbf{Y}^t\mathbf{Y})] = \frac{1}{s}w\left(\frac{2m+p-2}{2s}; \frac{2sr}{2m+p-2}\right), \tag{C2}$$

for $2m+p>2$, $r,s>0$, and w the function defined in Lemma 2. Then, the unique solution of the maximization problem

$$\max_g H(\mathbf{Y}) = \max_g \left\{-\int g(\mathbf{y})\log g(\mathbf{y})d\mathbf{y}\right\},$$

under the constraints (C1) and (C2), is given by the density

$$g(\mathbf{y}) = C_p(\mathbf{y}^t\mathbf{y})^{m-1}\exp[-r(\mathbf{y}^t\mathbf{y})^s], \quad r,s>0, \quad 2m+p>2,$$

and the normalizing constant C_p is given by (3).

Proof. Based on the Lagrange multipliers method we have

$$H(\mathbf{Y}) - \lambda - \mu\frac{2m+p-2}{2sr} - \kappa\frac{1}{s}w\left(\frac{2m+p-2}{2s}; \frac{2sr}{2m+p-2}\right)$$

$$\begin{aligned} &= \int_{R^p} g(\mathbf{y})\log[e^{-\lambda}(\mathbf{y}^t\mathbf{y})^{-\kappa}\exp[-\mu(\mathbf{y}^t\mathbf{y})^s]g^{-1}(\mathbf{y})]d\mathbf{y} \\ &\le \int_{R^p} g(\mathbf{y})[e^{-\lambda}(\mathbf{y}^t\mathbf{y})^{-\kappa}\exp[-\mu(\mathbf{y}^t\mathbf{y})^s]g^{-1}(\mathbf{y})]d\mathbf{y} - 1, \end{aligned}$$

with equality if and only if

$$g(\mathbf{y}) = e^{-\lambda}(\mathbf{y}^t\mathbf{y})^{-\kappa}\exp[-\mu(\mathbf{y}^t\mathbf{y})^s]. \tag{5}$$

In view of Lemma 1 (a) and the fact that $g(\mathbf{y})$, given by (5), is a density function we have that

$$e^{\lambda} = \frac{\pi^{p/2}\Gamma((p-2\kappa)/2s)}{s\Gamma(p/2)}\mu^{(2\kappa-p)/2s}, \ \mu>0, \ \kappa<\frac{p}{2}. \tag{6}$$

Constraint (C1) and Lemma 1 (b) lead to the relation

$$e^{\lambda}\frac{2m+p-2}{2sr} = \frac{\pi^{p/2}\Gamma(((p-2\kappa)/2s)+1)}{s\Gamma(p/2)}\mu^{((2\kappa-p)/2s)-1}, \mu>0, \ \kappa<\frac{p}{2}.$$

The last equation, taking into account relation (6), gives

$$\mu = \frac{2sr}{2m+p-2}\frac{p-2\kappa}{2s}. \tag{7}$$

On the other hand constraint (C2), Lemma 1 (c) and relation (6), lead to the relation

$$\Gamma\left(\frac{p-2\kappa}{2s}\right)w\left(\frac{2m+p-2}{2s}; \frac{2sr}{2m+p-2}\right) = \Gamma'\left(\frac{p-2\kappa}{2s}\right) - \log(\mu)\Gamma\left(\frac{p-2\kappa}{2s}\right). \tag{8}$$

Based on this last equation, and relations (7), (8) we have that

$$w\left(\frac{2m+p-2}{2s};\frac{2sr}{2m+p-2}\right)=w\left(\frac{p-2\kappa}{2s};\frac{2sr}{2m+p-2}\right).$$

In view of Lemma 2 and the last equation we obtain that $-\kappa = m-1$ and using (7) we obtain that $\mu = r$. Taking into account that $\kappa = 1-m$ and $\mu = r$, it is obtained that $e^{-\lambda} = C_p$, by using relation (6). This completes the proof of the theorem. ■

Based on Theorem 1 and Lemma 3 we can now state a similar characterization result for Kotz type symmetric distribution with density given by (2). In this context, for a $p-$dimensional vector $\boldsymbol{\mu}$ and a positive definite matrix $\boldsymbol{\Sigma}$ of order p, consider the linear transformation $\mathbf{X} = \boldsymbol{\Sigma}^{1/2}\mathbf{Y}+\boldsymbol{\mu}$, of the random vector $\mathbf{Y}$ of Theorem 1. The next corollary states that among all densities f in R^p that satisfy suitable constraints, the Kotz type symmetric distribution with density given by (2) is the unique density that maximizes Shannon's entropy.

Corollary 1. Let $\mathbf{X}^t = (X_1, ..., X_p)$ be a $p-$variate random vector in R^p with density f. Let also

$$E_f\left[(\mathbf{X}-\boldsymbol{\mu})^t\boldsymbol{\Sigma}^{-1}(\mathbf{X}-\boldsymbol{\mu})\right]^s=\frac{2m+p-2}{2sr}, \qquad \text{(C1*)}$$

and

$$E_f\left[\log\left((\mathbf{X}-\boldsymbol{\mu})^t\boldsymbol{\Sigma}^{-1}(\mathbf{X}-\boldsymbol{\mu})\right)\right]=\frac{1}{s}w\left(\frac{2m+p-2}{2s};\frac{2sr}{2m+p-2}\right), \qquad \text{(C2*)}$$

for $2m+p>2$, $r,s>0$. Then the unique solution of the maximization problem

$$\max_f H(\mathbf{X}) = \max_f\left\{-\int f(\mathbf{x})\log f(\mathbf{x})d\mathbf{x}\right\},$$

under the constraints (C1*) and (C2*), is the density of the Kotz type symmetric distribution given by (2).

In the next corollary the analytic formula for the entropy of the Kotz type symmetric distribution is presented. The proof can be immediately obtained in view of Corollary 1, relations (1) and (2) and taking into account constraints (C1*) and (C2*) of Corollary 1.

Corollary 2. The Shannon entropy of the Kotz type symmetric distribution with density given by (2) is

$$H(Kotz)=-\log C_p+\frac{1}{2}\log|\boldsymbol{\Sigma}|+r\frac{2m+p-2}{2sr}-(m-1)\frac{1}{s}w\left(\frac{2m+p-2}{2s};\frac{2sr}{2m+p-2}\right),$$

with $C_p = \frac{s\Gamma(p/2)}{\pi^{p/2}\Gamma((2m+p-2)/2s)}r^{(2m+p-2)/2s}$, $w(x;\alpha) = \Psi(x)-\log(\alpha x)$, for $x>0$, and $r,s>0$, $2m+p>2$.

An application of Corollary 2 for $m=1$, $s=1$ and $r=1/2$, leads to the well-known entropy of the multivariate normal distribution that is $\frac{1}{2}\left(p\log(2\pi)+\log|\boldsymbol{\Sigma}|+p\right)$.

3 Burr and Pareto type III multivariate distributions

The aim of this section is to obtain Burr and Pareto type III multivariate distributions by means of the maximum entropy principle. A random vector $\mathbf{X}^t = (X_1, ..., X_p)$ follows a Burr multivariate distribution if the density function of $\mathbf{X}$ is defined by (cf. Johnson and Kotz (1972)),

$$f(\mathbf{x}) = \prod_{i=1}^{p} (\alpha + i - 1) d_i c_i x_i^{c_i - 1} (1 + \sum_{i=1}^{p} d_i x_i^{c_i})^{-(\alpha+p)}, \tag{9}$$

for $x_i > 0$, $c_i > 0$, $d_i > 0$, $i = 1, ..., p$, and $\alpha > 0$. From here and in the sequel we shall be concerned with the case $\alpha = 1$.

The multivariate Pareto type III distribution has density (cf. Johnson and Kotz (1972))

$$f^*(\mathbf{z}) = \prod_{i=1}^{p} \frac{i}{\gamma_i \theta_i} \left(\frac{z_i - \lambda_i}{\theta_i} \right)^{\frac{1}{\gamma_i} - 1} \left(1 + \sum_{i=1}^{p} \left(\frac{z_i - \lambda_i}{\theta_i} \right)^{\frac{1}{\gamma_i}} \right)^{-(1+p)} \tag{10}$$

for $z_i > \lambda_i$, $\gamma_i > 0$, $\theta_i > 0$, $i = 1, ..., p$. If $\alpha = 1$, $c_i = 1/\gamma_i$, and $d_i = 1$, then a Pareto random vector of type III, to be denoted as $\mathbf{Z}$, can be obtained from a Burr random vector $\mathbf{X}$ by the following component-wise transformation

$$z_i = \theta_i x_i + \lambda_i, \quad i = 1, ..., p.$$

Hence the maximum entropy characterization of Pareto type III multivariate distribution can be achieved by the respective one of the multivariate Burr distribution in view of Lemma 3 and the above transformation. In order to present the maximum entropy characterization of Burr distribution we will follow ideas of the previous section. In this context the following lemmas are necessary. The proof of Lemma 4 is outlined in the Appendix.

Lemma 4. Let $S = \{\mathbf{y} \in R^p : \ y_i > 0, \ i = 1, ..., p\}$. Then for $c_i > 0$, $i = 1, ..., p$,

$$a) \quad \int_S (1 + \sum_{i=1}^{p} y_i^{c_i})^{-\mu} \prod_{i=1}^{p} y_i^{-\kappa_i} d\mathbf{y} = \Gamma(\mu - \sum_{i=1}^{p} \beta_i) \prod_{i=1}^{p} \Gamma(\beta_i) / \Gamma(\mu) \prod_{i=1}^{p} c_i,$$

$$b) \quad \int_S (1 + \sum_{i=1}^{p} y_i^{c_i})^{-\mu} \prod_{i=1}^{p} y_i^{-\kappa_i} \log(1 + \sum_{i=1}^{p} y_i^{c_i}) d\mathbf{y} = (\prod_{i=1}^{p} \Gamma(\beta_i) / \Gamma^2(\mu) \prod_{i=1}^{p} c_i) \times$$
$$\times (\Gamma(\mu - \sum_{i=1}^{p} \beta_i) \Gamma'(\mu) - \Gamma(\mu) \Gamma'(\mu - \sum_{i=1}^{p} \beta_i)),$$

$$c) \quad \int_S \log(y_j) (1 + \sum_{i=1}^{p} y_i^{c_i})^{-\mu} \prod_{i=1}^{p} y_i^{-\kappa_i} d\mathbf{y} = (\prod_{\substack{i=1 \\ i \neq j}}^{p} \Gamma(\beta_i) / c_j \Gamma(\mu) \prod_{i=1}^{p} c_i) \times$$
$$\times (\Gamma(\mu - \sum_{i=1}^{p} \beta_i) \Gamma'(\beta_j) - \Gamma(\beta_j) \Gamma'(\mu - \sum_{i=1}^{p} \beta_i)),$$

for $\beta_i = (1-\kappa_i)/c_i$, with $\beta_i > 0$, κ_i real constants and $\mu > \sum_{i=1}^{p} \beta_i$, $i, j = 1, ..., p$.

Lemma 5. For $p \in N$, consider the system of equations

$$\begin{array}{l} \Psi(y) - \Psi(y-px) = \Psi(1+p) - \Psi(1) \\ \Psi(x) - \Psi(y-px) = 0 \end{array}, \tag{11}$$

for $(x,y) \in V = \{(x,y) \in R^2 : \ x > 0, \ \ y > px\}$, with Ψ the digamma function. The equations (11) have the unique solution $x = 1$ and $y = 1+p$.

Proof. Taking into account that the digamma function is strictly increasing, the second equation of (11) leads to

$$y = (p+1)x. \tag{12}$$

Hence the first equation of (11) becomes

$$\Psi((p+1)x) - \Psi(x) = \Psi(1+p) - \Psi(1). \tag{13}$$

For $x > 0$ define the function $\omega(x) = \Psi((p+1)x) - \Psi(x)$. It is well known that $\Psi(mx) = \log m + \frac{1}{m} \sum_{\kappa=0}^{m-1} \Psi\left(x + \frac{\kappa}{m}\right)$. The derivative of Ψ with respect to x at the point $m = p+1$ gives that $(p+1)\Psi'((p+1)x) = \frac{1}{p+1} \sum_{\kappa=0}^{p} \Psi'(x + \frac{\kappa}{p+1})$. Taking into account that the function Ψ' is strictly decreasing the last identity leads to $(p+1)\Psi'((p+1)x) < \Psi'(x)$. Hence the function $\omega(x)$, defined above, is strictly decreasing and the equation (13) has the unique solution $x = 1$. From (12) the unique solution with respect to y is $y = 1+p$ which completes the proof of the lemma. ■

Theorem 2. Let $\mathbf{Y}^t = (Y_1, ..., Y_p)$ be a $p-$variate random vector in $S = \{\mathbf{y} \in R^p : y_i > 0, i = 1, ..., p\}$ with density g. Let also for $c_i > 0$, $i = 1, ..., p$, the following constraints are satisfied,

$$E_g(\log(1+\sum_{i=1}^{p} Y_i^{c_i})) = \Psi(1+p) - \Psi(1), \tag{C3}$$

$$E_g(\log(Y_j)) = 0, \quad j = 1, ..., p, \tag{C4}$$

for Ψ the digamma function. In this context, the unique solution of the maximization problem

$$\max_g H(\mathbf{Y}) = \max_g \left\{ -\int_S g(\mathbf{y}) \log g(\mathbf{y}) d\mathbf{y} \right\},$$

under the constraints (C3) and (C4), is given by the density

$$g(\mathbf{y}) = \prod_{i=1}^{p} i c_i y_i^{c_i - 1} (1+\sum_{i=1}^{p} y_i^{c_i})^{-(1+p)}, \quad c_i > 0, \ \ i = 1, ..., p, \ \text{ and } \ \mathbf{y} \in S.$$

Proof. Following the steps of the proof of Theorem 1, for constants λ, μ, and κ_i, $i = 1, ..., p$, we have that

$$H(\mathbf{Y}) - \lambda - \mu(\Psi(1+p) - \Psi(1)) \leq \int_S g(\mathbf{y}) \left(e^{-\lambda}(1+\sum_{i=1}^p y_i^{c_i})^{-\mu} \prod_{i=1}^p y_i^{-\kappa_i} \right) g^{-1}(\mathbf{y}) d\mathbf{y} - 1,$$

with equality if and only if

$$g(\mathbf{y}) = e^{-\lambda}(1+\sum_{i=1}^p y_i^{c_i})^{-\mu} \prod_{i=1}^p y_i^{-\kappa_i}. \tag{14}$$

From Lemma 4 (a) and the fact that the density $g(\mathbf{y})$ above integrates to 1, we have that

$$e^{\lambda} = \Gamma(\mu - \sum_{i=1}^p \beta_i) \prod_{i=1}^p \Gamma(\beta_i)/\Gamma(\mu) \prod_{i=1}^p c_i, \tag{15}$$

for $\beta_i = (1-\kappa_i)/c_i$, with $\beta_i > 0$, $i = 1, ..., p$, and $\mu > \sum_{i=1}^p \beta_i$. Constraint (C3) and Lemma 4 (b) lead, after a little algebra, to the equality

$$\Psi(1+p) - \Psi(1) = \Psi(\mu) - \Psi(\mu - \sum_{i=1}^p \beta_i). \tag{16}$$

In a similar manner constraint (C4), Lemma 4 (c) and relation (15) give that

$$\Psi(\beta_j) - \Psi(\mu - \sum_{i=1}^p \beta_i) = 0, \quad j = 1, ..., p. \tag{17}$$

Equations (16) and (17) lead,

$$\Psi(\beta_j) = \Psi(\mu) - \Psi(1+p) + \Psi(1), \quad \text{for every } j = 1, ..., p,$$

which means that

$$\Psi(\beta_1) = \Psi(\beta_2) = .. = \Psi(\beta_p).$$

This last equation and the strict monotonicity of digamma function Ψ ensure that

$$\beta_1 = \beta_2 = .. = \beta_p = \beta. \tag{18}$$

Based on them, equations (16) and (17) are equivalent to the following

$$\begin{aligned} &\Psi(1+p) - \Psi(1) = \Psi(\mu) - \Psi(\mu - p\beta), \\ &\Psi(\beta) - \Psi(\mu - p\beta) = 0, \end{aligned} \tag{19}$$

for $\beta > 0$ and $\mu > p\beta$. From Lemma 5 we have that the unique solution of the equations (19) is

$$\beta = 1 \quad \text{and} \quad \mu = 1 + p. \tag{20}$$

It is $\beta_i = (1 - \kappa_i)/c_i$, $i = 1, ..., p$, then from (19) and (20) we have that

$$-\kappa_i = c_i - 1, \quad i = 1, ..., p. \tag{21}$$

Relations (15), (18), and (20) lead to $e^{-\lambda} = \prod_{i=1}^{p} ic_i$, which completes the proof of the theorem in view of (14). ■

The density

$$g(\mathbf{y}) = \prod_{i=1}^{p} ic_i y_i^{c_i - 1}(1 + \sum_{i=1}^{p} y_i^{c_i})^{-(1+p)}, \quad c_i > 0, \ i = 1, ..., p, \tag{22}$$

and $\mathbf{y} \in S = \{\mathbf{y} \in R^p : y_i > 0, \ i = 1, ..., p\}$, obtained in Theorem 2, can be also used in order to generate the multivariate Burr distribution with density given by (9) and the parameter $\alpha = 1$. Indeed, if the $p-$variate random vector $\mathbf{Y}^t = (Y_1, ..., Y_p)$ has density $g(\mathbf{y})$, given by (22), then the random vector $\mathbf{X}^t = (X_1, ..., X_p)$ defined by the following component-wise transformation

$$x_i = d_i^{-1/c_i} y_i, \quad i = 1, ..., p,$$

has a multivariate Burr distribution with density given by (9) and the parameter $\alpha = 1$. This remark associated with Theorem 2 and Lemma 3 lead to the following corollary which states the maximum entropy characterization of Burr multivariate distribution for the parameter $\alpha = 1$.

Corollary 3. Let $\mathbf{X}^t = (X_1, ..., X_p)$ be a $p-$variate random vector in $S = \{\mathbf{x} \in R^p : x_i > 0, \ i = 1, ..., p\}$ with density f. Let also for $c_i > 0$, $i = 1, ..., p$, the following constraints are satisfied,

$$E_f(\log(1 + \sum_{i=1}^{p} d_i X_i^{c_i})) = \Psi(1 + p) - \Psi(1), \tag{C3*}$$

$$E_f(\log(d_j^{1/c_j} X_j)) = 0, \quad j = 1, ..., p, \tag{C4*}$$

for Ψ the digamma function. Then the unique solution of the maximization problem

$$\max_f H(\mathbf{X}) = \max_f \left\{ - \int_S f(\mathbf{x}) \log f(\mathbf{x}) d\mathbf{x} \right\},$$

under the constraints (C3*) and (C4*), is given by the density

$$f(\mathbf{x}) = \prod_{i=1}^{p} id_i c_i x_i^{c_i - 1}(1 + \sum_{i=1}^{p} d_i x_i^{c_i})^{-(1+p)}, \quad c_i > 0, \ d_i > 0, \ i = 1, ..., p, \ \mathbf{x} \in S. \tag{23}$$

Based on the discussion at the beginning of this section, the multivariate Pareto type III distribution with density $f^*(\mathbf{z})$, given by (10), can be generated from (23) for $c_i = 1/\gamma_i$, and $d_i = 1$, and by using the component-wise transformation

$$z_i = \theta_i x_i + \lambda_i, \quad i = 1, ..., p,$$

for $z_i > \lambda_i$, $\gamma_i > 0$, $\theta_i > 0$, $i = 1, ..., p$. This transformation, in association with Corollary 3 and Lemma 3, leads to a similar maximum entropy characterization of the Pareto type III density, given by (10), under the following constraints,

$$E_{f^*}\left(\log\left(1+\sum_{i=1}^{p}\left(\frac{Z_i-\lambda_i}{\theta_i}\right)^{\frac{1}{\gamma_i}}\right)\right) = \Psi(1+p) - \Psi(1), \tag{C3**}$$

$$E_{f^*}\left(\log\left(\frac{Z_j-\lambda_j}{\theta_j}\right)\right) = 0, \quad j = 1, ..., p. \tag{C4**}$$

The above results can be also used in order to evaluate the Shannon entropy of the multivariate Burr, for $\alpha = 1$, and the multivariate Pareto type III distributions. These entropies are presented in the corollary that follows the proof of which can be immediately obtained in view of Corollary 3, relations (9) and (10) and taking into account constraints (C3*), (C4*) and (C3**), (C4**).

Corollary 4. a) The Shannon entropy of the Burr distribution with density given by (9) and $\alpha = 1$, $c_i > 0$, $d_i > 0$, $i = 1, ..., p$, is

$$H(Burr, \alpha = 1) = -\sum_{i=1}^{p}\log i + (1+p)[\Psi(1+p) - \Psi(1)] - \sum_{i=1}^{p}\log\left(c_i d_i^{1/c_i}\right).$$

b) The Shannon entropy of the Pareto type III distribution, with density given by (10), is

$$H(Pareto\ III) = -\sum_{i=1}^{p}\log\frac{i}{\theta_i} + (1+p)[\Psi(1+p) - \Psi(1)] + \sum_{i=1}^{p}\log(\gamma_i),$$

for $\gamma_i > 0$, $\theta_i > 0$, $i = 1, ..., p$.

The above expressions for the Shannon entropy of Burr and Pareto III multivariate distributions have been also obtained by Darbellay and Vajda (2000) in a different framework.

A Appendix

Proof of Lemma 1. The proof of parts (a) and (c) are given in the sequel. Part (b) can be proved in a similar manner.

(a) Consider the generalized spherical coordinate transformation

$$\begin{aligned} x_1 &= r\prod_{k=1}^{p-1}\sin\theta_k \\ x_j &= r(\prod_{k=1}^{p-j}\sin\theta_k)\cos\theta_{p-j+1}, \quad 2\le j\le p-1, \\ x_p &= r\cos\theta_1 \end{aligned}$$

for $0 < r \leq 1$, $0 < \theta_i \leq \pi$, $i = 1, ..., p-2$ and $0 < \theta_{p-1} \leq 2\pi$. Clearly, we have $\mathbf{x}^t\mathbf{x} = x_1^2 + \cdots + x_p^2 = r^2$ and the Jacobian of the transformation from $x_1, ..., x_p$ to $r, \theta_1, ..., \theta_{p-1}$ is $r^{p-1}\sin^{p-2}\theta_1 \sin^{p-3}\theta_2 \cdots \sin\theta_{p-2}$ (cf. Muirhead (1982), p. 37). Taking into account the equality

$$\int_0^{\pi}\cdots\int_0^{\pi}\int_0^{2\pi} \sin^{p-2}\theta_1 \sin^{p-3}\theta_2 \cdots \sin\theta_{p-2} d\theta_1 \cdots d\theta_{p-2} d\theta_{p-1} = \frac{2\pi^{\frac{p}{2}}}{\Gamma(\frac{p}{2})},$$

(cf. Muirhead (1982), p. 37), we have

$$\begin{aligned} \int_{R^p} (\mathbf{y}^t\mathbf{y})^{-\mu}\exp[-\kappa(\mathbf{y}^t\mathbf{y})^s]d\mathbf{y} &= \frac{2\pi^{p/2}}{\Gamma(p/2)}\int_0^{\infty}(r^2)^{-\mu}\exp[-\kappa(r^2)^s]\,r^{p-1}dr \\ &= \frac{\pi^{p/2}}{s\Gamma(p/2)}\int_0^{\infty} z^{\frac{p-2\mu}{2s}-1}\exp(-\kappa z)dz. \end{aligned}$$

The last integral is the gamma function $\Gamma((p-2\mu)/2s)\kappa^{(2\mu-p)/2s}$, $\kappa > 0$, $\mu < \frac{p}{2}$, and the proof of part (a) of the lemma is completed.
(c) If we consider again the generalized spherical coordinate transformation, we have

$$\begin{aligned} \int_{R^p} (\mathbf{y}^t\mathbf{y})^{-\mu}\exp[-\kappa(\mathbf{y}^t\mathbf{y})^s]\log(\mathbf{y}^t\mathbf{y})d\mathbf{y} &= \frac{2\pi^{p/2}}{\Gamma(p/2)}\int_0^{\infty}(r^2)^{-\mu}\exp[-\kappa(r^2)^s]\log(r^2)\,r^{p-1}dr \\ &= \frac{\pi^{p/2}\kappa^{\frac{p-2\mu}{2s}+1}}{s^2\Gamma(p/2)}\int_0^{\infty}(\kappa z)^{\frac{p-2\mu}{2s}-1}\exp(-\kappa z)\log(z)dz. \end{aligned} \tag{A1}$$

The derivative with respect to t of the gamma function $\Gamma(t) = \int_0^{\infty} x^{t-1}\exp(-x)dx$, $t > 0$, is given by $\Gamma'(t) = \int_0^{\infty} x^{t-1}\exp(-x)\log(x)dx$. For $x = \alpha z$, $\alpha > 0$, we have

$$\Gamma'(t) = \alpha\int_0^{\infty}(\alpha z)^{t-1}\exp(-\alpha z)\log(\alpha)dz + \alpha\int_0^{\infty}(\alpha z)^{t-1}\exp(-\alpha z)\log(z)dz.$$

Hence

$$\alpha\int_0^{\infty}(\alpha z)^{t-1}\exp(-\alpha z)\log(z)dz = \Gamma'(t) - \log(\alpha)\Gamma(t).$$

An application of this last equality, for $\alpha = \kappa$ and $t = (p - 2\mu)/2s$, to the relation (A1) leads to the desired result. ■

Proof of Lemma 4. a) Consider the transformation $u_i = y_i^{c_i}$, $i = 1, ..., p$. Then if we denote by $I = \int\limits_S (1 + \sum\limits_{i=1}^{p} y_i^{c_i})^{-\mu} \prod\limits_{i=1}^{p} y_i^{-\kappa_i} d\mathbf{y}$, we have

$$I = (\prod_{i=1}^{p} c_i)^{-1} \int_0^\infty \int_0^\infty ... \int_0^\infty \left[\prod_{i=1}^{p-1} u_i^{\beta_i - 1} \left(\int_0^\infty u_p^{\beta_p - 1} (1 + \sum_{i=1}^{p} u_i)^{-\mu} du_p \right) \right] du_1 ... du_{p-1} \quad \text{(A2)}$$

Consider now the integral $I_p = \int\limits_0^\infty u_p^{\beta_p - 1} (1 + \sum\limits_{i=1}^{p} u_i)^{-\mu} du_p$. If we will use the transformation $\omega = u_p/(1 + \sum\limits_{i=1}^{p-1} u_i)$, then

$$I_p = (1 + \sum_{i=1}^{p-1} u_i)^{-\mu + \beta_p} \int_0^\infty \omega^{\beta_p - 1} (1 + \omega)^{-\mu} d\omega = (1 + \sum_{i=1}^{p-1} u_i)^{-\mu + \beta_p} B(\beta_p, \mu - \beta_p),$$

for $\beta_p > 0$, $\mu > \beta_p$ and B the beta function. Taking into account relation (A2)

$$I = (\prod_{i=1}^{p} c_i)^{-1} B(\beta_p, \mu - \beta_p) \int_0^\infty \int_0^\infty ... \int_0^\infty \left[\prod_{i=1}^{p-1} u_i^{\beta_i - 1} \left(1 + \sum_{i=1}^{p-1} u_i \right)^{-\mu + \beta_p} \right] du_1 ... du_{p-1}.$$

If the same procedure is repeated $(p-1)$–times then the integral I becomes

$$I = (\prod_{i=1}^{p} c_i)^{-1} \prod_{i=1}^{p} B(\beta_i, \mu - \sum_{j=i}^{p} \beta_j),$$

and leads to the desired result.
b) The proof of this part follows immediately from part a) if we observe that

$$\int_S (1 + \sum_{i=1}^{p} y_i^{c_i})^{-\mu} \prod_{i=1}^{p} y_i^{-\kappa_i} \log(1 + \sum_{i=1}^{p} y_i^{c_i}) d\mathbf{y} = -\frac{d}{d\mu} \int_S (1 + \sum_{i=1}^{p} y_i^{c_i})^{-\mu} \prod_{i=1}^{p} y_i^{-\kappa_i} d\mathbf{y}.$$

c) Let $I^* = \int\limits_S \log(y_j)(1 + \sum\limits_{i=1}^{p} y_i^{c_i})^{-\mu} \prod\limits_{i=1}^{p} y_i^{-\kappa_i} d\mathbf{y}$. Consider the transformation $u_i = y_i^{c_i}$, $i = 1, ..., p$. Then

$$I^* = (c_j \prod_{i=1}^{p} c_i)^{-1} \int_S \log(u_j)(1 + \sum_{i=1}^{p} u_i)^{-\mu} \prod_{i=1}^{p} u_i^{\beta_i - 1} d\mathbf{u}, \quad \text{(A3)}$$

with $\beta_i = (1-\kappa_i)/c_i$, $i = 1, ..., p$. If we use the transformation $\omega = u_j/(1+\sum_{i=1,i\neq j}^{p-1} u_i)$, and take the derivatives, with respect to β_j, of both sides of the identity $\int_0^\infty \omega^{\beta_j-1}(1+\omega)^{-\mu}d\omega = \Gamma(\beta_j)\Gamma(\mu-\beta_j)/\Gamma(\mu)$, $\beta_j > 0$, $\mu > \beta_j$, after a little algebra, we obtain that

$$\int_0^\infty \log(u_j)u_j^{\beta_j-1}(1+\sum_{i=1}^{p} u_i)^{-\mu}du_j = (1+\sum_{i=1,i\neq j}^{p} u_i)^{-\mu+\beta_j}\left(W_1+W_2\log(1+\sum_{i=1,i\neq j}^{p} u_i)\right), \tag{A4}$$

with $W_1 = [\Gamma'(\beta_j)\Gamma(\mu-\beta_j) - \Gamma(\beta_j)\Gamma'(\mu-\beta_j)]/\Gamma(\mu)$ and $W_2 = \Gamma(\beta_j)\Gamma(\mu-\beta_j)/\Gamma(\mu)$. Using relation (A4) and parts (a) and (b) of Lemma 4, relation (A3) completes the proof of the lemma. ■

References

Bad Dumitrescu, M. (1986). The application of the principle of minimum cross-entropy to the characterization of the exponential-type probability distributions. *Ann. Inst. Statist. Math.* **38**, 451-457.

Darbellay, G. and Vajda, I. (2000). Entropy expressions for multivariate continuous distributions. *IEEE Trans. Inf. Theory.* **IT-46**, 709-712.

Ebrahimi, N. (2000). The maximum entropy method for lifetime distributions. *Sankhya*, Series A, **62**, 236-243.

Fang, K. T., Kotz, S. and Ng, K. W. (1990). *Symmetric multivariate and related distributions.* Chapman and Hall.

Guiasu, S. (1977). *Information Theory with Applications.* McGraw-Hill. New York.

Guiasu, S. (1990). A classification of the main probability distributions by minimizing the weighted logarithmic measure of deviation. *Ann. Inst. Statist. Math.* **42**, 269-279.

Gzyl, H. (1995). *The Method of Maximum Entropy.* World Scientific.

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

Johnson, N. L. and Kotz, S. (1972). *Distributions in statistics: Continuous multivariate distributions.* Wiley. New York.

Kagan, A. M., Linnik, Yu. V. and Rao, C. R. (1973). *Characterization Problems in Mathematical Statistics.* Wiley, New York.

Kapur, J. N. (1989). *Maximum Entropy Models in Science and Engineering.* Wiley.

Kotz, S., Kozubowski, T. J. and Podgórski, K. (2000). Maximum entropy characterization of asymmetric Laplace distribution. To appear in: *Int. Math. Journal*, **1** (2002), 31-35.

Muirhead, R. J. (1982). *Aspects of Multivariate Statistical Theory.* Wiley, New York.

Preda, V. (1982). The Student distribution and the principle of maximum entropy. *Ann. Inst. Statist. Math.* **34**, 335-338.

Shore, J. E. and Johnson, R. W. (1980). Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy. *IEEE Trans. Inf. Theory.* **IT-26**, 26-37.

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