$\mathcal{I}^2\mathcal{SDS}$

***The Institute for Integrating Statistics in Decision Sciences***

***Technical Report TR-2007-2***

**March 23, 2007**

# Multivariate Maximum Entropy Identification, Transformation and Dependence


Nader Ebrahimi
Division of Statistics
Northern Illinois University

Ehsan S. Soofi
Sheldon B. Lubar School of Business
University of Wisconsin-Milwaukee

Refik Soyer
Department of Decision Sciences
The George Washington University

# Multivariate Maximum Entropy Identification, Transformation, and Dependence

Nader Ebrahimi
Division of Statistics
Northern Illinois University
DeKalb, IL 60155
nader@math.niu.edu

Ehsan S. Soofi
School of Business Administration
University of Wisconsin-Milwaukee
P.O.Box 742, Milwaukee, WI 53201
esoofi@uwm.edu

Refik Soyer
Department of Decision Sciences
George Washington University
Washington D.C. 20052
soyer@gwu.edu

March 23, 2007

**Abstract**

This paper shows that multivariate distributions can be characterized as Maximum Entropy (ME) models based on the well-known general representation of density function of the ME distribution subject to moment constraints. In this approach, the problem of ME characterization simplifies to the problem of representing the multivariate density in the ME form, hence the need for case-by-case proofs by calculus of variations or other methods are eliminated. The main vehicle for the ME characterization is the the information distinguishability relationship, which easily extends to the multivariate case. Results are also formulated that encapsulate implications of the multiplication rule of probability and the entropy transformation formula for ME characterization. The dependence structure of multivariate ME distribution in terms of the moments and the support of distribution is studied. The relationships of ME families with the exponential families and with conditional exponential families are explored. Applications include the ME characterizations of more than twenty five bivariate families.

KEY WORDS: Bivariate distribution; Characterization; Dependence; Exponential family; Kullback-Leibler information; Mutual information.

## 1 Introduction

Maximum entropy (ME) methods produce probability models for random prospects that incorporate given information [26, 27, 40, 41]. Many well known univariate parametric families of distributions and several multivariate families have been characterized as ME models [6, 7, 29, 42, 44, 45]. In statistics, numerous diagnostics and tests of distributional hypotheses have been developed based on ME characterizations [11, 14, 18, 19, 20, 21, 22, 25, 32, 35, 37, 38, 39, 42, 43].

Shannon entropy of a distribution $F$ is defined as

$$H(F|\mathcal{S}) = -\int_{\mathcal{S}} f(\boldsymbol{x}) \log f(\boldsymbol{x}) d\nu(\boldsymbol{x}). \tag{1}$$

In most parts of this paper the density is with respect to the Lebesgue measure and subscript $\nu$ is omitted. In one case (Example 1a) we use the product measure, Lebesgue$\times$counting. In a few other cases (in Table 3) where the distributions do not have density with respect to the

Lebesgue measure, we include subscript $\nu$ to identify the densities. The support $\mathcal{S}$ will be included in information quantities as needed to emphasize its key role for some results.

For information characterization of multivariate distributions we consider a moment class of distributions

$$\Omega_F = \{F : E_F[T_j(\boldsymbol{X})] = \theta_j,\ j = 1, \cdots, J\}, \tag{2}$$

where $T_j(\boldsymbol{x})$ are real-valued integrable functions with respect to $dF(\boldsymbol{x})$, and $\theta_j,\ j = 1, \cdots, J$ are specified moments. The ME model in the moment class $\Omega_F$ is the distribution that maximizes (1).

The set of linearly independent moments

$$\mathcal{T} = \{T_j(\boldsymbol{X}),\ j = 1, \cdots, J\}, \tag{3}$$

that generates $\Omega_F$ will be referred to as the moment information set. If the expected value of elements of one moment set $\mathcal{T}_1$ can be obtained from the expected value of elements of another moment set $\mathcal{T}_2$, the two sets will be referred to as congruent, denoted as $\mathcal{T}_1 \cong \mathcal{T}_2$.

This paper provides some results for the ME characterizations of multivariate distributions. These results, which are formulated from some well-known relationships of information theory, eliminate the need for laborious proofs like Lagrangian and isoperimetric formulation of calculus of variations often used for ME characterizations of parametric families on case-by-case basis for each family.

The main vehicles for the ME characterization are the well-known general representation of density function of the ME distribution in $\Omega_F$, and the information distinguishability relationship, which easily extends to the multivariate case. In this approach, the problem of ME characterization simplifies to the problem of representing the multivariate density in the ME form, similar to the identification of distributions with densities in exponential families.

Results are also formulated that encapsulate implications of the multiplication rule of probability and the well-known entropy transformation formula for ME characterization.

The dependence structure of multivariate ME distribution in terms of the moment information set (3) and the support of distribution is studied. It is noted that, given the marginal distributions, the ME model in $\Omega_F$ is also the minimum dependence model. The concept of nested ME distributions is introduced and results are formulated for measures of information dependence for ME models.

The relationship and lack of equivalence between ME families of distributions and exponential families are explored. The relationship between families of ME distributions and distributions with conditional in exponential families is also explored. Numerous examples are presented. For simplicity, all examples are bivariate. Applications to ME characterizations of more than twenty five bivariate families are presented.

This paper is organized as follows. Section 2 presents the results for ME identification, and explores the relationships of ME families with the exponential families and with conditional exponential families. Section 3 presents ME characterizations of numerous bivariate families. Section 4 gives a result on ME transformation and its application to ME characterizations of several bivariate families. Section 5 discusses dependence structure of ME distributions.

## 2 Maximum Entropy Identification

The Kullback-Leibler information between two distributions $F$ and $G$ for a random vector $\boldsymbol{X}$ is

$$K(F:G|\mathcal{S}) = \int_{\mathcal{S}} \log \frac{f(\boldsymbol{x})}{g(\boldsymbol{x})} f(\boldsymbol{x}) d\nu(\boldsymbol{x}), \tag{4}$$

where $f$ and $g$ are densities with respect to measure $\nu$ over the support $\mathcal{S}$. It is assumed that $F$ is absolutely continuous with respect to $G$.

Kullback-Leibler information have many desirable properties for developing probability and statistical methodologies. Two properties utilized in this paper are: (a) $K(F:G) \geq 0$, where the equality holds if and only if $f(\boldsymbol{x}) = g(\boldsymbol{x})$ almost everywhere; and (b) for a given $g$, $K(F:G)$ is convex in $f$. $K(F:G)$ is also referred to as relative entropy, cross-entropy, and directed divergence, and $G$ is referred to as the reference distribution. However, unlike the Kullback-Leibler information, $-\infty \leq H[F(\boldsymbol{x}|\mathcal{S})] \leq \infty$.

The MDI model in $\Omega_F$ relative to $G$, when $F \in \Omega_F$ is absolutely continuous with respect to $G$, is defined by

$$F^* = \arg \min_{F \in \Omega_F} K(F:G).$$

The MDI Theorem and its extension to multivariate and multiparameter case ([30], pp. 38, 48) give the solution. If

$$C(\lambda, \mathcal{S}) = \left[\int_{\mathcal{S}} e^{-\lambda_1 T_1(\boldsymbol{x}) - \cdots - \lambda_J T_J(\boldsymbol{x})} g(\boldsymbol{x}) d\nu(\boldsymbol{x})\right]^{-1} > 0 \tag{5}$$

exists, then

$$K[F:G|\mathcal{S})] \geq K[F^*:G|\mathcal{S})] = \log C(\lambda, \mathcal{S}) - \lambda_1\theta_1 - \cdots - \lambda_J\theta_J, \quad \forall F \in \Omega_F, \tag{6}$$

and $F^*$ is unique and has density function in the form of

$$f^*(\boldsymbol{x}|\mathcal{S}) = C(\lambda, \mathcal{S}) g(\boldsymbol{x}) e^{-\lambda_1 T_1(\boldsymbol{x}) - \cdots - \lambda_J T_J(\boldsymbol{x})}, \tag{7}$$

where $\lambda = (\lambda_1, \cdots, \lambda_J)$ is the vector of Lagrange multipliers given by $\theta_j = \dfrac{\partial}{\partial \lambda_j} \log C(\lambda, \mathcal{S})$.

We will consider two reference distributions which give equivalent MDI solutions. In our first case the reference distribution $G$ is uniform (proper or improper) and the solution is the ME model.

The second case that we consider in Section 4 is when the reference distribution $G$ is the product of a set of marginal distributions.

When (5) exists for the uniform $G$, the MDI Theorem gives the ME model

$$f^*(\boldsymbol{x}|\mathcal{S}) = C(\lambda, \mathcal{S})e^{-\lambda_1 T_1(\boldsymbol{x})-\cdots-\lambda_J T_J(\boldsymbol{x})}. \tag{8}$$

The entropy of ME model is given by

$$H[F(\boldsymbol{x}|\mathcal{S})] \leq H[F^*(\boldsymbol{x}|\mathcal{S})] = -\log C(\lambda, \mathcal{S}) + \lambda_1\theta_1 + \cdots + \lambda_J\theta_J, \quad \forall F \in \Omega_F. \tag{9}$$

The density (8) can be obtained directly by calculus of variations (see, e.g., [27, 12]). It should be noted, however, that the ME (or MDI) model may not exist for a class of distributions, a reference distribution, or for a support. When it exists, the ME model is unique due to the fact that the entropy is concave in $f$. The calculus of variations solution suggest the form (8) as the solution, but does not show the uniqueness ([12], p. 267) without the additional burden of examining second-order variation.

The following multivariate information distinguishability relationship between Kullback-Leibler function and entropies is the key to the simple proof of ME characterization of a distribution.

**Lemma 1** *For any $F \in \Omega_F$,*

$$K(F : F^*|\mathcal{S}) = H(F^*|\mathcal{S}) - H(F|\mathcal{S}) \geq 0, \tag{10}$$

*where $F^*$ is the ME model in $\Omega_F$. The equality holds if and only if $f(\boldsymbol{x}|\mathcal{S}) = f^*(\boldsymbol{x}|\mathcal{S})$ almost everywhere.*

The proof is the same as the univariate case given in [42]. Note that

$$K(F : F^*|\mathcal{S}) = -H(F|\mathcal{S}) - E_f[\log f^*(\boldsymbol{X})].$$

Use (8) for $f^*$ and note that since $F^* \in \Omega_F$, $E_f[T_j(\boldsymbol{X}) = \theta_j]$. Then the result is given by (9). The equality $K(f : f^*|\mathcal{S}) = 0$ is attained if and only if $f(\boldsymbol{x}) = f^*(\boldsymbol{x})$ almost everywhere.

Proofs for ME characterization of particular families of distributions frequently appear in the literature. The information distinguishability relation is a simple but sufficiently general result that alleviates the burden of proof for particular families of distributions by calculus of variations and other mathematical techniques. Only identifying the moment conditions is needed. The following result encapsulates this ME characterization approach. Proof is given in [12], pp. 267-268.

**Lemma 2 (Maximum Entropy Identification)** *Any distribution with a density function in form of (8) is the unique ME model in the moment class of distributions (2) generated by the moment information set $\mathcal{T} = \{T_j(\boldsymbol{x}), j = 1, \cdots, J\}$ shown in exponent in (8).*

Some sufficient conditions for finiteness of multivariate entropy can be drawn from some well-known results.

(a) Boundedness of $f(\boldsymbol{x})$ implies $H(F) > -\infty$.

(b) If $F$ is absolutely continuous with respect to the product of marginals $F_1 \cdots F_d$, then finiteness of marginal entropies implies $H(F) < \infty$. This is seen by noting that $H(F) \leq \sum_{i=1}^{d} H(F_i)$, where $F_i$ denotes the marginal distribution.

(c) If $F$ is absolutely continuous with respect to the product of marginals $F_1 \cdots F_d$, then finiteness of marginal variances implies $H(F) < \infty$. Let $\sigma_i^2$ denote the variance of $F_i$. It is well-known that $H(F_i) \leq \frac{1}{2}\log(2\pi e) + \frac{1}{2}\log \sigma_i^2$, since the right-hand-side is entropy of the normal distribution which is ME for given variance.

A sharp upper bound for the multivariate entropy is given when $F$ has finite variances and covariances. If the covariance matrix $\Sigma$ is nonsingular, then

$$H(F) \leq \frac{d}{2}\log(2\pi e) + \frac{1}{2}\log|\Sigma|, \tag{11}$$

and the equality is attained if and only if $F$ is multivariate normal distribution.

Next result give the ME characterization of the joint distribution in terms of the ME characterization of a marginal and a set of conditional distributions.

**Corollary 1 (Maximum Entropy Chain Rule)** *Let* $\mathcal{T}_{X_1} = \{T_{j_1}(X_1),\ j_1 = 1, \cdots, J_1\}$ *and* $\mathcal{T}_{X_k|x_1:x_{k-1}} = \{T_{j_k}(x_1, \cdots, x_{k-1}, X_k),\ j_k = 1, \cdots, J_k\}$ $k = 2, \cdots, d$ *denote the moment information sets that characterize the marginal distribution* $F_{X_1}$ *and conditional distributions* $F_{X_k|x_1,\cdots,x_{k-1}}$ *as ME, respectively. Then the joint distribution* $F_{\boldsymbol{X}}$ *is ME in the class of distributions generated by the moment information set*

$$\mathcal{T}_{\boldsymbol{X}} \cong \mathcal{T} \subseteq \bigcup_{k=1}^{d} \mathcal{T}^*_{X_1:X_k},$$

*where* $\mathcal{T}^*_{X_1:X_0} = \mathcal{T}_{X_1}$ *and*

$$\mathcal{T}^*_{X_1:X_k} = \{\log C_k(X_1, \cdots, X_{k-1}), \lambda_{j_k}(X_1, \cdots, X_{k-1})T_{j_k}(X_1, \cdots, X_k),\ j_k = 1, \cdots, J_k\},\ k = 2, \cdots, d,$$

*in which* $\lambda_{j_k}(x_1, \cdots, x_{k-1}),\ j_k = 1, \cdots, J_k$ *are parameters and* $C_k(x_1, \cdots, x_{k-1})$ *is the normalizing factors for the conditional ME densities.*

**Proof.** Write the joint ME density using multiplication rule

$$f^*(\boldsymbol{x}) = f^*_{X_1}(x_1) f^*_{X_2|x_1}(x_2) \cdots f^*_{X_d|x_1,\cdots,x_{d-1}}(x_d).$$

The densities in the right-hand-side are ME with moment information sets $\mathcal{T}_{X_k|x_1:x_{k-1}}$, $k = 1, \cdots, d$, so they are in the form of (8). The parameters and normalizing factor of the marginal density $f^*_{X_1}(x_1)$ do not depend on $x_2, \cdots, x_d$. The moments, parameters, and normalizing factor of each conditional density $f^*_{X_k|x_1,\cdots,x_{k-1}}(x_k)$ are functions of $x_1, \cdots, x_{k-1}$, some possibly constant functions. Write each conditional ME density in the form of (8) with $\log C_k(x_1, \cdots, x_{k-1})$ in the exponent. Then the product in the right-hand-side is in the ME form (8). The result is obtained upon simplifications. When there is no cancelation of terms in the right-hand-side product, then $\mathcal{T}$ is the union. If there is a cancelation, $\mathcal{T}$ is a proper subset of the union.

The product decomposition of a joint density is order-dependent. By uniqueness of the ME distribution when it exists, one expects that the union for all $n!$ arrangements of the components of $\boldsymbol{X}$ to be congruent with $\mathcal{T}_{\boldsymbol{X}}$. When there is no cancelation of a factor in the product of conditional and marginal densities $\mathcal{T}_{\boldsymbol{X}}$ is congruent with the union. Next example illustrates the chain rule result for three bivariate distributions.

**Example 1**

(a) Consider the geometric distribution with $f_{y|p}(y) = p(1-p)^y$, $y = 0, 1, 2\cdots$ where $p$ has a beta distribution with $f_p(p) = \dfrac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1}(1-p)^{\beta-1}$. The density of beta distribution can be written in ME form (8) with $\mathcal{T}_p = \{T_1(p) = \log p,\ T_2(p) = \log(1-p)\}$, $\lambda_1 = 1-\alpha$ and $\lambda_2 = 1-\beta$. The geometric distribution can be written in ME form (8) with $\mathcal{T}_{Y|p} = \{T_{1_2}(p, Y) = Y\}$ and $\lambda_{1_2} = \lambda_{1_2}(p) = -\log(1-p)$, $C_2(p) = p$. Thus, $\mathcal{T}^*_{Y,p} = \{\log p, -Y\log(1-p)\}$. The joint (beta, geometric) distribution is ME in the class of bivariate distributions generated by the moment information set

$$\begin{aligned}\mathcal{T}_{(Y,p)} &= \{T_1(Y,p) = \log p,\ T_2(Y,p) = \log(1-p),\ T_3(Y,p) = Y\log(1-p)\} \\ &\cong \mathcal{T}_p \cup \mathcal{T}^*_{Y,p},\end{aligned}$$

and with $\lambda_1 = -\alpha$, $\lambda_2 = 1-\beta$ and $\lambda_3 = 1$.

(b) Mihram and Hultquist [34] motivated and derived the Beta-Stacy distribution as the product of the conditional beta distribution with density

$$f(x_2|x_1) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)x_1^{\beta-1}} x_2^{\alpha-1}(x_1-x_2)^{\beta-1}, \quad 0 \le x_2 \le x_1,$$

and the generalized gamma (Stacy) distribution with density

$$f(x_1) = \frac{\tau\lambda^{\tau\beta}}{\Gamma(\beta)} x_1^{\tau\beta-1} e^{-(\lambda x_1)^\tau}, \quad x_1 \ge 0.$$

We note that the marginal density $f(x_1)$ is in ME form (8) with $\mathcal{T}_{X_1} = \{T_1(X_1) = X_1^\tau,\ T_2(X_1) = \log X_1\}$ and with $\lambda_1 = \lambda^\tau$, $\lambda_2 = 1-\tau\beta$. The conditional density $f(x_2|x_1)$ is in ME form (8) with

$\mathcal{T}_{X_2|x_1} = \{T_{1_2}(x_1, X_2) = \log X_2,\ T_{2_2}(x_1, X_2) = \log(x_1 - X_2)\}$ and $\lambda_{1_2} = 1-\alpha,\ \lambda_{2_2} = 1-\beta$. In this case the model parameters are constant functions of $x_1$, but a moment and the normalizing factor includes $x_1$. Thus, $\mathcal{T}^*_{X_2,X_1} = \{\log X_1, \log X_2, \log(X_1 - X_2)\}$. Since there is no cancelation of terms in the product $f(x_1)f(x_2|x_1)$, the moment information set for $f(x_1, x_2)$ is

$$\begin{aligned}\mathcal{T}_{X_1,X_2} &= \mathcal{T}_{X_1} \cup \mathcal{T}^*_{X_1,X_2} \\ &= \{T_1(\boldsymbol{X}) = X_1^\tau,\ T_2(\boldsymbol{X}) = \log X_1,\ T_3(\boldsymbol{X}) = \log X_2,\ T_4(\boldsymbol{X}) = \log(X_1 - X_2)\}\end{aligned}$$

This moment information set characterizes the joint Beta-Stacy density

$$f(x_1, x_2) = \frac{|\tau|\lambda^{\alpha+\beta}\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)\Gamma(\delta)} x_1^{\tau\delta-\alpha-\beta} x_2^{\alpha-1}\,(x_1 - x_2)^{\beta-1} e^{-(\lambda x_1)^\tau},\quad 0 < x_2 < x_1,\quad \alpha,\ \beta,\ \delta,\ \lambda > 0,$$

which is in ME form (8) with $\lambda_1 = \lambda^\tau,\ \lambda_2 = \alpha+\beta-\tau\delta,\ \lambda_3 = 1-\alpha,\ \lambda_4 = 1-\beta$.

(c) Let the conditional distribution $f(x_2|x_1)$ be beta as above and the marginal distribution of $X_1$ be gamma with density

$$f(x_1) = \frac{\lambda^\beta}{\Gamma(\beta)} x_1^{\beta-1} e^{-\lambda x_1},\quad x_1 \geq 0.$$

This density is in ME form (8) with $\mathcal{T}_{X_1} = \{X_1,\ \log X_1\}$. Since the term $x_1^{\beta-1}$ cancels out in the product $f(x_1)f(x_2|x_1)$, the moment information set for $f(x_1, x_2)$ is a proper subset of the union

$$\begin{aligned}\mathcal{T}_{X_1,X_2} &= \{T_1(\boldsymbol{X}) = X_1,\ T_2(\boldsymbol{X}) = \log X_2,\ T_2(\boldsymbol{X}) = \log(X_1 - X_2)\} \\ &\subset \mathcal{T}_{X_1} \cup \mathcal{T}^*_{X_1,X_2}.\end{aligned}$$

This moment information characterizes McKay's bivariate gamma distribution [33] with density

$$f(x_1, x_2) = \frac{\lambda^{\alpha+\beta}}{\Gamma(\alpha)\Gamma(\beta)} x_2^{\alpha-1}(x_1 - x_2)^{\beta-1} e^{-\lambda x_1},\quad 0 < x_2 < x_1,\quad \alpha,\ \beta,\ \lambda > 0.$$

This density is in ME form (8) with $\lambda_1 = \lambda,\ \lambda_2 = 1-\alpha$, and $\lambda_3 = 1-\beta$.

## 2.1 Maximum Entropy Exponential Families

Lemma 2 shows that the problem of ME characterization by moments is only an identification problem, very much like identifying a distribution as a member of the exponential family. However, the sets of ME and exponential families are not isomorphic.

A multivariate family $\{F_\beta\}$ is said to be an $n$-parameter exponential family if the density of $F_\beta$ is in the form of

$$f_\beta(\boldsymbol{x}) = q(\boldsymbol{\beta})r(\boldsymbol{x})e^{\tau_1(\boldsymbol{\beta})W_1(\boldsymbol{x})+\cdots+\tau_n(\boldsymbol{\beta})W_n(\boldsymbol{x})}, \tag{12}$$

where $q(\boldsymbol{\beta})$ and $\tau_i(\boldsymbol{\beta})$ are real-valued functions free from $\boldsymbol{x}$, and $r(\boldsymbol{x})$ and $W_i(\boldsymbol{x})$ are real-valued functions free from $\boldsymbol{\beta}$.

(a) If $r(\boldsymbol{x})$ is a constant, then $f_\beta(\boldsymbol{x})$ is in the ME form (8), and the exponential family member is the ME model in the class of distributions generated by information set $\mathcal{T} = \mathcal{W} = \{W_i(\boldsymbol{X}),\ i = 1, \cdots, n\}$.

(b) If $\log r(\boldsymbol{x})$ is integrable, then $f_\beta(\boldsymbol{x})$ in (12) can be written in the ME form (8), hence $F = F^*$ is the ME model in the class of distributions generated by information set $\mathcal{T} = \mathcal{W} \cup \mathcal{W}_r(\boldsymbol{X})$, where $\mathcal{W}_r(\boldsymbol{X})$ is set of moments generated by $\log r(\boldsymbol{x})$.

(c) When $\log r(\boldsymbol{x})$ is not integrable, then $f_\beta(\boldsymbol{x})$ in (12) still can be written in the ME form (8), however, it is not a proper ME model with finite entropy.

(d) Some ME densities may not be written in the exponential family form (12). An example is $t$ distribution. Consider the bivariate $t$ density

$$f(x_1, x_2) = \frac{1}{2\pi}\left(1 + \frac{x_1^2 + x_2^2}{m}\right)^{-(m/2+1)}, \quad m > 0.$$

This density can be written in the ME form (8) with $T_1(\boldsymbol{X}) = \log(m + X_1^2 + X_2^2)$ and $\lambda_1 = m/2+1$, but it can not be written in the form of (12). The ME characterizations of multivariate $t$ as a transformation of Pearson Type VII is given by Zografos [45].

The following example illustrates that an exponential family of distribution may or may not be an ME distribution.

**Example 2**

(a) The most famous member of the distribution in exponential family and the most well known ME model is the normal distribution. The bivariate normal density is

$$\begin{aligned} f_\beta(x_1, x_2) = &\ \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \\ &\times exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(x_1-\mu_1)^2}{\sigma_1^2} - \frac{2\rho(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2} + \frac{(x_2-\mu_2)^2}{\sigma_2^2}\right]\right\}. \end{aligned} \tag{13}$$

This density can be written in the ME form (8) with

$$\mathcal{T}_X = \{T_1(\boldsymbol{X}) = X_1,\ T_2(\boldsymbol{X}) = X_2,\ T_3(\boldsymbol{X}) = X_1^2,\ T_4(\boldsymbol{X}) = X_2^2,\ T_5(\boldsymbol{X}) = X_1X_2\}, \tag{14}$$

and $\lambda_j = \lambda_j(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho),\ j = 1, 2, 3, 4, 5$; hence it is ME by Lemma 2. If $\boldsymbol{\beta} = (\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$, then $r(\boldsymbol{x})$ in (12) is constant and $W_j(\boldsymbol{X}) = T_j(\boldsymbol{X}),\ j = 1, 2, 3, 4, 5$ shown in the ME moment information set (14). Several other examples of parametric families with exponential family densities will be presented in Section 3.

(b) Consider the normal distribution when $\sigma_1^2, \sigma_2^2$, and $\rho$ are known and $\boldsymbol{\beta} = (\mu_1, \mu_2)$. Then, $W_1(\boldsymbol{x}) = x_1$, $W_2(\boldsymbol{x}) = x_2$, and $\log r(\boldsymbol{x}) = a_0(a_1x_1^2 + a_2x_2^2 + a_3x_1x_2)$, where $a_k = a_k(\sigma_1^2, \sigma_2^2, \rho),\ k = 0, 1, 2, 3$. Since $\log r(\boldsymbol{x})$ is integrable, the normal distribution $F_\beta = F^*$ is ME, however, subject to the same moment information set as (14) where the first two moments, $T_1(\boldsymbol{X}) = W_1(\boldsymbol{X}) = X_1$ and $T_2(\boldsymbol{X}) = W_2(\boldsymbol{X}) = X_2$, are from the exponential family exponents and the last three moments are from $\log r(\boldsymbol{x})$.

(c) Consider the bivariate distribution with the following density

$$f_\beta(x_1, x_2) = \frac{\beta(\beta+1)}{x_1x_2(\log x_1 + \log x_2 - 1)^{\beta+2}}, \quad x_1, x_2 \geq e, \quad 0 < \beta \leq 1. \tag{15}$$

This is an exponential family density with $r(\boldsymbol{x}) = \dfrac{1}{x_1x_2}$ and $W_1(\boldsymbol{x}) = \log(\log x_1 + \log x_2 - 1)$. It can be shown that $\log r(\boldsymbol{x})$ is not integrable and hence $f_\beta(x_1, x_2)$ is not a proper ME model.

An $n$-parameter bivariate distribution is said to have conditionals in exponential family [3, 4] if its conditional densities $f_1(x_1|x_2)$ and $f_2(x_2|x_1)$ are in (12) form with $q_k,\ r_k,\ \tau_{k,i}$, and $W_{k,i},\ i = 1, \cdots, n_k,\ k = 1, 2$. The joint density is in the form of

$$f_B(x_1, x_2) = r_1(x_1)r_2(x_2)e^{\boldsymbol{W_1(x_1)'\Lambda W_2(x_2)}}, \tag{16}$$

where

$$\begin{aligned} \boldsymbol{W_1}(x_1) &= [1, W_{1,1}(x_1), \cdots, W_{1,n_1}(x_1)]' \\ \boldsymbol{W_2}(x_2) &= [1, W_{2,1}(x_2), \cdots, W_{2,n_2}(x_2)]', \end{aligned}$$

and $\Lambda$ is $(n_1+1) \times (n_2+1)$ matrix of parameters

$$\Lambda = \begin{bmatrix} \lambda_{00} & | & \lambda_{01} & \cdots & \lambda_{0n_2} \\ -- & + & -- & -- & -- \\ \lambda_{10} & | & & & \\ \vdots & | & & \Lambda_{12} & \\ \lambda_{n_10} & | & & & \end{bmatrix}$$

subject to the normalizing

$$\int_{\mathcal{S}_{X_1}} \int_{\mathcal{S}_{X_2}} f(x_1, x_2) d\nu_1(x_1)\ d\nu_2(x_2) = 1,$$

which imposes restrictions on the $\lambda_{ij}$.

The density (16) is in the exponential family density (12) form. Thus, the bivariate distributions with conditionals in exponential families are ME whenever $\log r(x_k),\ k = 1, 2$ are integrable. The ME information moment set is given by $\mathcal{T} = (\mathcal{W}_1 \times \mathcal{W}_2) \cup \mathcal{W}_{r_1}(x_1) \cup \mathcal{W}_{r_2}(x_2)$, where $(\mathcal{W}_1 \times \mathcal{W}_2) = \{W_{1i}(X_1)W_{2\ell}(X_2),\ i = 0, 1, \cdots, n_1,\ \ell = 0, 1, \cdots, n_2,\ i = \ell \neq 0,\ \beta_{ij} \neq 0\}$, and $\mathcal{W}_{r_k}(x_k)$ is set of moments generated by $\log r_k(x_k)$. Next example illustrates ME characterization of bivariate distribution with normal conditionals. Several other applications will be given in Section 3.

**Example 3**

Bivariate normal conditionals is given by (16) where $W_1(x_1) = (1, x_1, x_1^2),\ W_2(x_2) = (1, x_2, x_2^2)$, $r_1(x_1) = r_2(x_2) = 1$, and for integrability one of the two set of restrictions on $B$ must hold:

$$\lambda_{22} = \lambda_{12} = \lambda_{21} = 0,\quad \lambda_{20} < 0,\ \lambda_{02} < 0,\quad 4\lambda_{02}\lambda_{20} > \lambda_{11}^2 \tag{17}$$

$$\lambda_{22} < 0,\quad 4\lambda_{22}\lambda_{02} > \lambda_{12}^2,\quad 4\lambda_{22}\lambda_{20} > \lambda_{21}^2 \tag{18}$$

The parameter restrictions (17) give the bivariate normal distribution.

The moment information set for the ME characterization of bivariate distribution with normal conditionals is

$$\begin{aligned} \mathcal{T} &= (\mathcal{W}_1 \times \mathcal{W}_2) \\ &= \{T_1(\boldsymbol{X}) = X_1,\ T_2(\boldsymbol{X}) = X_2, \quad T_3(\boldsymbol{X}) = X_1^2,\ T_4(\boldsymbol{X}) = X_2^2,\ T_5(\boldsymbol{X}) = X_1X_2, \\ &\qquad T_6(\boldsymbol{X}) = X_1^2X_2,\ T_7(\boldsymbol{X}) = X_1X_2^2,\ T_8(\boldsymbol{X}) = X_1^2X_2^2\}. \end{aligned}$$

Thus the bivariate distribution with normal conditionals requires three additional moments than the bivariate normal distribution. Specific cases include the bivariate normal distribution obtained from the first five moments, bivariate distribution with normal conditionals of Castillo and Galambos [10] obtained from the subset $\{T_3(\boldsymbol{X}), T_4(\boldsymbol{X}), T_6(\boldsymbol{X}), T_7(\boldsymbol{X}), T_8(\boldsymbol{X})\}$ and the centered bivariate distribution with normal conditionals obtained from the subset $\{T_3(\boldsymbol{X}), T_4(\boldsymbol{X}), T_8(\boldsymbol{X})\}$.

# 3 Applications: ME Characterization of Bivariate Families

This section presents the ME characterizations of many bivariate distributions. Simplification of ME characterization is achieved through location-scale transformation $\boldsymbol{Y} = \Sigma^{1/2}(\boldsymbol{X} + \boldsymbol{\mu})$. We consider $F_X$. That is, without loss of generality, we ignore the location vector and scale matrix in most cases where no particular purpose being served. One can simply obtain the ME characterization of $F_Y$ by application of Lemma 3 of Section 4 to the affine transformation $\boldsymbol{X} = \Sigma^{-1/2}(\boldsymbol{Y} - \boldsymbol{\mu})$. If $\Sigma$ is nonsingular, Jacobian of transformation is $|\Sigma|^{-1/2}$.

The distributions are classified into three groups according to their supports. and one group shows applications of Lemma 3.

## 3.1 Distributions with Unrestricted Support

Examples of ME bivariate families with support the entire space $\mathcal{S} = \Re^2$, the first quadrant $\mathcal{S} = Q_1 = \{(x_1, x_2) : x_1, x_2 \geq 0\}$, and the first and second quadrants $\mathcal{S} = Q_1 \cup Q_2 = \{(x_1, x_2) : -\infty < x_1 < \infty, x_2 \geq 0\}$ are shown in Table 1. Each distribution is ME in the class of distributions generated by the corresponding moment information set with $T_j(X_1, X_2)$ shown in the second column of the table. It is easy to verify that the densities of the distributions listed in Table 1 can

Table 1. Examples of Maximum Entropy Bivariate Distributions with Rectangular Support

| ME Distribution and density | Information Set |
|---|---|
| Pareto, $x_1, x_2 \geq 0$ <br> $f(x_1,x_2) = \alpha(\alpha+1)(1+x_1+x_2)^{-\alpha-2}, \quad \alpha > 0$ | $T_1(\boldsymbol{X}) = \log(1+X_1+X_2)$ |
| Pearson Type VII, $-\infty < x_1, x_2 < \infty$ <br> $f(x_1,x_2) = \dfrac{\alpha-1}{\pi}(1+x_1^2+x_2^2)^{-\alpha}, \quad \alpha > 1.$ | $T_1(\boldsymbol{X}) = \log(1+X_1^2+X_2^2)$ |
| Kotz Type, $-\infty < x_1, x_2 < \infty$ <br> $f(x_1,x_2) = \dfrac{\tau\lambda^{\alpha/\tau}}{\pi\Gamma(\alpha/\tau)}(x_1^2+x_2^2)^{\alpha-1}e^{-\lambda(x_1^2+x_2^2)^\tau}, \quad \alpha,\lambda,\tau > 0$ | $\begin{cases} T_1(\boldsymbol{X}) = (X_1^2+X_2^2)^\tau, \\ T_2(\boldsymbol{X}) = \log(X_1^2+X_2^2) \end{cases}$ |
| Freund's bivariate exponential, $x_1, x_2 \geq 0$ <br> (Absolutely Continuous Bivariate Exponential) <br> $f(x_1,x_2) = \begin{cases} \dfrac{\lambda}{\lambda_1+\lambda_2}\lambda_1(\lambda_2+\lambda_{12})\ e^{-\lambda_1 x_1-(\lambda_2+\lambda_{12})x_2}, & x_1 < x_2 \\ \dfrac{\lambda}{\lambda_1+\lambda_2}\lambda_2(\lambda_1+\lambda_{12})\ e^{-(\lambda_1+\lambda_{12})x_1-\lambda_2 x_2}, & x_1 > x_2, \end{cases}$ <br> $\lambda = \lambda_1+\lambda_2+\lambda_{12}, \lambda_1, \lambda_2, \lambda_{12} > 0$ | $\begin{cases} T_1(\boldsymbol{X}) = X_1,\ T_2(\boldsymbol{X}) = X_2, \\ T_1(\boldsymbol{X}) = \max(X_1,X_2), \\ T_4(\boldsymbol{X}) = D(X_1 < X_2) \\ D(x_1 < x_2) = \begin{cases} 1 & \text{if } x_1 < x_2 \\ 0 & \text{otherwise} \end{cases} \end{cases}$ |
| Pareto Conditionals, $x_1, x_2 \geq 0$ <br> $f(x_1,x_2) = C(\alpha,a,b)(a+x_1+x_2+bx_1x_2)^{-\alpha-1}, \quad \alpha > 0$ | $T_1(\boldsymbol{X}) = \log(a+X_1+X_2+bX_1X_2)$ |
| Bivariate Exponential Conditional, $x_1, x_2 \geq 0$ <br> $f(x_1,x_2) = \lambda_1\lambda_2 c(\delta)e^{-\lambda_1x_1-\lambda_2x_2-\lambda_3x_1x_2},$ <br> $\lambda_1,\lambda_2 > 0, \lambda_3 \geq 0,\ \delta = \lambda_3/\lambda_1\lambda_2$ | $\begin{cases} T_1(\boldsymbol{X}) = X_1,\ T_2(\boldsymbol{X}) = X_2, \\ T_3(\boldsymbol{X}) = X_1X_2 \end{cases}$ |
| Bivariate gamma conditionals Type II, $x_1, x_2 \geq 0$ <br> $f(x_1,x_2) = \dfrac{c_{\alpha,\beta}(\lambda_3)\lambda_1^\alpha\lambda_2^\beta}{\Gamma(\alpha)\Gamma(\beta)}x_1^{\alpha-1}\ x_2^{\beta-1}e^{-\lambda_1x_1-\lambda_2x_2-\lambda_3x_1x_2},$ <br> $\alpha,\beta,\lambda_1,\lambda_2,\lambda_3 > 0$ | $\begin{cases} T_1(\boldsymbol{X}) = X_1,\ T_2(\boldsymbol{X}) = X_2, \\ T_3(\boldsymbol{X}) = \log X_1,\ T_4(\boldsymbol{X}) = \log X_2, \\ T_5(\boldsymbol{X}) = X_1X_2 \end{cases}$ |
| Gamma-Gamma Mix, $x_1, x_2 \geq 0$ <br> $f(x_1,x_2) = \dfrac{\lambda_1^\alpha\lambda_3^\beta}{\Gamma(\alpha)\Gamma(\beta)}x_1^{\alpha+\beta-1}\ x_2^{\beta-1}e^{-\lambda_1x_1-\lambda_3x_1x_2},$ <br> $\alpha,\beta,\lambda_1,\lambda_3 > 0$ | $\begin{cases} T_1(\boldsymbol{X}) = X_1, \\ T_2(\boldsymbol{X}) = \log X_1,\ T_3(\boldsymbol{X}) = \log X_2, \\ T_4(\boldsymbol{X}) = X_1X_2 \end{cases}$ |
| Normal and Gamma conditionals, $-\infty < x_1 < \infty,\ x_2 \geq 0$ <br> $f(x_1,x_2) = C(\alpha_1,\alpha_2,\alpha_3,\lambda_1,\lambda_2,\lambda_3,\lambda_4,\lambda_5)x_2^{\alpha_1+\alpha_2x_1+\alpha_3x_1^2}$ <br> $\times\ e^{-\lambda_1x_1-\lambda_2x_1^2-\lambda_3x_2-\lambda_4x_1x_2-\lambda_5x_1^2x_2},$ | $\begin{cases} T_1(\boldsymbol{X}) = X_1,\ T_2(\boldsymbol{X}) = X_1^2, \\ T_3(\boldsymbol{X}) = X_2,\ T_4(\boldsymbol{X}) = \log X_2 \\ T_5(\boldsymbol{X}) = X_1X_2,\ T_6(\boldsymbol{X}) = X_1^2X_2 \\ T_7(\boldsymbol{X}) = X_1\log X_2,\ T_8(\boldsymbol{X}) = X_1^2\log X_2, \end{cases}$ |

be written in the form of (8). We only make some brief comments about these distributions. The entropy expressions for these distributions are given in [36], so the existence is guaranteed.

The Pearson Type VII, Kotz type, densities of Pareto, and the Freund's bivariate exponential/Absolutely Continuous Bivariate Exponential (ACBE) distributions shown in Table 1 are members of exponential family with a constant $r(\boldsymbol{x})$, hence they are ME. The ME characterizations of Pearson Type VII is given by Zografos [45] and the ME characterizations of Kotz Type is given by Auglogiaris and Zografos [7] with laborious proofs.

The ACBE distribution is the absolutely continuous part of Marshall-Olkin exponential distribution (shown in Table 3) and is a reparamterization of Freund's bivariate exponential distribution [9, 31]. The four constraints for ME characterization of ACBE distribution corresponds to four independent sufficient statistics that appear in the maximum likelihood equations for a sample of $n$ observations given in [9], where the last constraint corresponds to $n_1 = \sum_{i=1}^{n} D(x_{1i} < x_{2i})$. Also note that the density of ACBE distribution can be represented in the ME form (8) as:

$$f^*(x_1, x_2) = \frac{\lambda}{\lambda_1 + \lambda_2}\lambda_2(\lambda_1 + \lambda_{12})\ e^{-\lambda_1 x_1 - \lambda_2 x_2 - \lambda_{12}\max(x_1,x_2) - \lambda_3 D(x_1 < x_2)}, \tag{19}$$

where $D(x_1 < x_2)$ is the indicator function shown in Table 1. Since, $E[D(X_1 < X_2)] = \dfrac{\lambda_1}{\lambda_1 + \lambda_2}$, the parameters in the representation (19) are restricted as

$$\lambda_3 = \log\left(\frac{\lambda_1(\lambda_2 + \lambda_{12})}{\lambda_2(\lambda_1 + \lambda_{12})}\right).$$

That is, the role of the last constraint for ME characterization of ACBE distribution is the parameter restriction required for the identification of its two branches.

The bivariate Pareto conditionals, bivariate exponential conditionals, bivariate gamma conditionals, the gamma-gamma mixture and gamma normal conditionals are examples of (16) with $\lambda_{i\ell} = 0$ for some $i$ and $\ell$ [2, 3, 4]. The normalizing constant for bivariate Pareto conditionals depends on $a$ and $b$ of the information moment. For $a = 1$ and $b = 0$, the moment reduces to the bivariate Pareto case. For $a = 0$ and $b > 0$, the normalizing constant is $C(\alpha, b) = \dfrac{\alpha\sin(\alpha\pi)}{\pi b^{\alpha-1}}$. For $a > 0$ and $b > 0$, the normalizing constant is awkward (see [3] pp. 106, 107). The normalizing constant for the bivariate exponential conditionals includes $c(\delta) = \left[\int_0^\infty e^{-z}(1+\delta z)^{-1}dz\right]^{-1}$.

The Type II bivariate distribution with gamma conditionals shown in Table 1 is Model II of Arnold et al [3]. It is obtained from (16) with $W_1(x_1) = (1, x_1, \log x_1)$, $W_2(x_2) = (1, x_2, \log x_2)$, $r_k(x_k) = x_k,\ k = 1, 2,\ \lambda_{12} = \lambda_{21} = \lambda_{22} = 0,\ \lambda_{10} = \alpha,\ \lambda_{01} = \beta,\ \lambda_{20} = \lambda_1,\ \lambda_{02} = \lambda_2 > 0,\quad \lambda_{11} = \lambda_3 > 0$; and the normalizing constant includes confluent hypergeometric function given by

$$c_{\alpha,\beta}(\lambda_3) = \frac{1}{\Gamma(\lambda_3)}\int_0^\infty e^{-\lambda_3 z} z^{\alpha-1}(1+z)^{\beta-\alpha-1}dz.$$

Arnold et al [3] list five other bivariate distributions with gamma conditionals (including independent bivariate gamma) obtained from (16) with restrictions on $\lambda_{i\ell}$. The ME characterizations of these models can be obtained similarly.

The Gamma-Gamma Mix distribution is a generalization of the gamma-exponential distribution studied by Darbellay and Vajda [15]. Both conditional distributions are gamma, thus it is in the form of (16) whose moment information set has one less element than Type II bivariate distribution with gamma conditionals. It may also be viewed as the limiting case of Type II bivariate gamma conditionals when $\lambda_2 \to 0$.

The bivariate distribution with normal and gamma conditionals shown in Table 1 is discussed by Arnold et al [3, 4]. It is obtained from (16) with $W_1(x_1) = (1, x_1, x_1^2),\ W_2(x_2) = (1, x_2, \log x_2)$, $r_1(x_1)$ a constant, $r_2 = x_2$, and

$$\lambda_{12},\ \lambda_{22} > 0,\ 4\lambda_{10}\lambda_{12} > \lambda_{12}^2,\ 4\lambda_{20}\lambda_{22} > \lambda_{21}^2,\ \lambda_{02} < \lambda_{22}\left(1 - \log\frac{\lambda_{22}}{\lambda_{12}}\right).$$

There is no closed form expression for the normalizing constant. For $\lambda_{10} = \lambda_{20} = \lambda_{12} = \lambda_{21} = \lambda_{22} = 0$, this distribution reduces to the conjugate normal-gamma prior distribution for normal likelihood [3, 4, 16]. Thus, the family of conjugate normal-gamma prior distributions for the univariate normal likelihood $f(x|\mu,\tau) = N(\mu, \tau^{-1})$ is the ME prior distribution subject to moment constraints

$$\mathcal{T} =\ T_3(\mu,\tau) = \tau,\ T_4(\mu,\tau) = \log\tau\{T_1(\mu,\tau) = \mu\tau,\ T_2(\mu,\tau) = \mu_2^2\tau\}.$$

### 3.2 ME Distributions with Restricted Supports

Table 2 shows a few examples where the supports are restricted. Each distribution shown in Table 2 is ME in the class of distributions generated from the corresponding moment information set.

The support of bivariate distribution with normal marginals shown in Table 2 is the first and third quadrant in $\Re^2$, [3]. This density is in the ME form (8). The bivariate exponential with triangular support gives the normalized first (second) branch of ACBE distribution shown in Table 1 for $\alpha = 0$ ($\alpha = 1$) and $\beta = 1$ ($\beta = 0$). This density is also in the ME form (8). The densities of bilateral Pareto [16], bivariate Pareto conditionals, bivariate beta conditionals [3], Dirichlet, and Pearson Type II distributions can be tranformed in the ME form (8) using the logarithmic moments shown in the table. The ME characterization of Dirichlet distribution is given by Kapur [29], and the ME characterization of Pearson Type II is given by Zografos [45], with laborious proofs. Application of Lemma 2 simplifies the proofs.

### 3.3 Distributions with Support in $\Re^2 \cup \Re$

The distributions shown in Table 3 give positive probability to a two-dimensional region as well as to a line (two lines in the case of natural exponential distribution). Formally, their supports are in the form $\mathcal{S} \subset \Re^2 \cup \Re$. Hence these distributions include absolutely continuous and singular parts, with respect to two-dimensional Lebesgue measure.

Nadarajah and Zografos [36] computed entropies of Marshall-Olkin, Cuadras-Augé and natural bivariate exponential distributions using densities shown in Table 3. Ebrahimi et al [23], following

Table 2. Examples of Maximum Entropy Bivariate Distributions with Restricted Supports

| ME Distribution and density | Information Set |
|---|---|
| Bivariate normal marginals, $x_1x_2 > 0$ <br> $f(x_1,x_2) = \dfrac{1}{\pi\sqrt{\sigma_1\sigma_2}}\, e^{-\frac{1}{2}(x_1^2/\sigma_1^2+x_2^2/\sigma_2^2)}, \quad \sigma_1,\ \sigma_2 > 0.$ | $T_1(\boldsymbol{X}) = X_1^2, \quad T_2(\boldsymbol{X}) = X_2^2$ |
| Bivariate Exponential on Triangle, $0 \le \alpha x_2 < x_1 < \beta x_2$ <br> $f(x_1,x_2) = \dfrac{(\alpha\lambda_1+\lambda_2)(\beta\lambda_1+\lambda_2)}{\beta-\alpha}\, e^{-\lambda_1x_1-\lambda_2x_2}, \quad \lambda_1,\lambda_2 > 0.$ | $T_1(\boldsymbol{X}) = X_1, \quad T_2(\boldsymbol{X}) = X_2$ |
| Bilateral Pareto, $0 < x_1 < \eta < \xi < x_2$ <br> $f(x_1,x_2) = \alpha(\alpha+1)(\xi-\eta)^{\alpha}(x_2-x_1)^{-\alpha+2}, \quad \alpha > 1$ | $T_1(\boldsymbol{X}) = \log(X_2 - X_1)$ |
| Pareto Conditionals $0 < x_1, x_2 < 1,\ x_1 + x_2 - Ax_1x_2 < 1$ <br> $f(x_1,x_2) = C(\alpha,\eta)(1-x_1-x_2+Ax_1x_2)^{1/\alpha-1}, \quad \alpha > 1$ | $T_1(\boldsymbol{X}) = \log(1 - X_1 - X_2 + AX_1X_2)$ |
| Beta conditionals, $0 < x_1, x_2 < 1,\ x_1 + x_2 < 1$ <br> $f(x_1,x_2) = C(\alpha_1,\alpha_2,\alpha_3,\beta)x_1^{\alpha_1-1}\, x_2^{\alpha_2-1}(1-x_1-x_2)^{\alpha_3-1}e^{\beta(\log x_1)(\log x_2)},$ <br> $\alpha_1,\ \alpha_2,\ \alpha_3 > 0$ | $\begin{cases} T_1(\boldsymbol{X}) = \log X_1,\ T_2(\boldsymbol{X}) = \log X_2 \\ T_3(\boldsymbol{X}) = \log(1 - X_1 - X_2) \\ T_4(\boldsymbol{X}) = (\log X_1)(\log X_2) \end{cases}$ |
| Dirichlet, $0 < x_1, x_2 < 1,\ x_1 + x_2 < 1$ <br> $f(x_1,x_2) = \dfrac{\Gamma(\alpha_1+\alpha_2+\alpha_3)}{\Gamma(\alpha_1)\Gamma(\alpha_2)\Gamma(\alpha_3)}x_1^{\alpha_1-1}\, x_2^{\alpha_2-1}(1-x_1-x_2)^{\alpha_3-1},$ <br> $\alpha_1,\ \alpha_2,\ \alpha_3 > 0$ | $\begin{cases} T_1(\boldsymbol{X}) = \log X_1,\ T_2(\boldsymbol{X}) = \log X_2 \\ T_3(\boldsymbol{X}) = \log(1 - X_1 - X_2) \end{cases}$ |
| Pearson Type II, $x_1^2 + x_2^2 \le 1$ <br> $f(x_1,x_2) = \dfrac{\alpha-1}{\pi}(1-x_1^2-x_2^2)^{\alpha}, \quad \alpha > 0.$ | $T_1(\boldsymbol{X}) = \log(1 - X_1^2 - X_2^2)$ |

the probabilistic argument of Marshall and Olkin ([31], p. 35), obtained the entropy of the Marshall-Olkin Bivariate Exponential (MOBE) using one and two dimensional Lebesgue measures via the partitioning property of (1). The entropies of other distributions shown in Table 3 can be obtained similarly.

For the ME characterizations of these distribution we use the densities relative to the suitable measures $\nu$ for which (1) is well-defined. The joint survival function $\bar{F}(x_1,x_2) = P(X_1 > x_1, X_2 > x_2)$ of MOBE can be represented as

$$\bar{F}(x_1,x_2) = \frac{\lambda_1+\lambda_2}{\lambda}\bar{F}_c(x_1,x_2) + \frac{\lambda_{12}}{\lambda}\bar{F}_s(x_1,x_2), \quad x_1,x_2 \ge 0, \quad \lambda_1,\ \lambda_2 > 0,\ \lambda_{12} \ge 0,$$

where $\lambda = \lambda_1 + \lambda_2 + \lambda_{12}$, $\bar{F}_c(x_1,x_2)$ is the absolutely continuous part with density of ACBE shown in Table 1, and

$$\bar{F}_s(x_1,x_2) = e^{-\lambda x_1},\ x_1 = x_2, \quad \lambda = \lambda_1 + \lambda_2 + \lambda_{12}$$

is a singular part. The singular part reflects the fact that $X_1 = X_2$ has positive probability, whereas the line $x_1 = x_2$ has two-dimensional Lebesgue measure zero.

Table 3. Examples of Maximum Entropy Bivariate Distributions with Probability Mass on a Line

| ME Distribution and density | Information Set |
|---|---|
| Marshall-Olkin Bivariate Exponential (MOBE), $x_1, x_2 > 0$ <br> $f_\nu(x_1,x_2) = \begin{cases} \lambda_1(\lambda_2+\lambda_{12})e^{-\lambda_1 x_1-(\lambda_2+\lambda_{12})x_2} & \text{for } x_2 > x_1 > 0 \\ \lambda_2(\lambda_1+\lambda_{12})e^{-(\lambda_1+\lambda_{12})x_1-\lambda_2 x_2} & \text{for } x_1 > x_2 > 0 \\ \lambda_{12}e^{-(\lambda_1+\lambda_2+\lambda_{12})x_1} & \text{for } x_1 = x_2 > 0 \\ & \lambda_1, \lambda_2 > 0,\ \lambda_{12} \geq 0 \end{cases}$ | $\begin{cases} T_1(\boldsymbol{X}) = X_1,\ T_2(\boldsymbol{X}) = X_2, \\ T_3(\boldsymbol{X}) = \max(X_1, X_2), \\ T_4(\boldsymbol{X}) = D(X_1 < X_2), \\ T_5(\boldsymbol{X}) = D(X_1 = X_2) \end{cases}$ |
| Cuadras-Augé bivariate distribution, $0 \leq x_1, x_2 \leq 1$ <br> $f_\nu(x_1,x_2) = \begin{cases} (1-\alpha)x_1^{-\alpha} & \text{for } 0 \leq x_2 < x_1 \leq 1 \\ (1-\alpha)x_2^{-\alpha} & \text{for } 0 \leq x_1 < x_2 \leq 1 \\ \alpha x_1^{-\alpha} & \text{for } 0 \leq x_1 = x_2 \leq 1 \\ & 0 \leq \alpha \leq 1 \end{cases}$ | $T_1(\boldsymbol{X}) = \log X_1,\ T_2(\boldsymbol{Y}) = \log X_2,$ <br> $T_3(\boldsymbol{X}) = \log[\min(X_1, X_2)],$ <br> $T_4(\boldsymbol{X}) = D(X_1 = X_2)$ |
| Natural Bivariate Exponential, $x_1, x_2 > 0$ <br> $f_\nu(x_1,x_2) = \begin{cases} \lambda e^{-x_1/\alpha} & \text{for } x_1 = \alpha x_2 \\ \lambda(1-\lambda)\beta e^{-\lambda x_1-(1-\lambda)\beta x_2} & \text{for } \alpha x_2 < x_1 < \beta x_2 \\ (1-\lambda)e^{-x_1} & \text{for } x_1 = \beta x_2 \\ \lambda = \dfrac{\beta-1}{\beta-\alpha}, & \alpha, \beta > 0 \end{cases}$ | $\begin{cases} T_1(\boldsymbol{X}) = X_1,\ T_2(\boldsymbol{X}) = X_2, \\ T_3(\boldsymbol{X}) = \max(X_1, \beta X_2), \\ T_4(\boldsymbol{X}) = \max(X_1, \alpha X_2), \\ T_5(\boldsymbol{X}) = D(X_1 = \beta X_2), \\ T_6(\boldsymbol{X}) = D(X_1 = \alpha X_2) \end{cases}$ |
| Exponential Autoregressive Bivariate distribution, $x_n, x_{n+1} > 0$ <br> $f_\nu(x_n,x_{n+1}) = \begin{cases} (1-\rho)\lambda^2 e^{-\lambda(1-\rho)x_n-\lambda x_{n+1}} & \text{for } x_{n+1} > \rho x_n > 0 \\ \rho\lambda e^{-\lambda x_n} & \text{for } x_{n+1} = \rho x_n > 0 \\ & \lambda > 0,\ 0 \leq \rho < 1 \end{cases}$ | $\begin{cases} T_1(\boldsymbol{X}) = X_n,\ T_2(\boldsymbol{X}) = X_{n+1}, \\ T_3(\boldsymbol{X}) = D(X_{n+1} = \rho X_n) \end{cases}$ |

Note:

* $D(z) = \begin{cases} 1 & \text{if } z \text{ holds} \\ 0 & \text{otherwise} \end{cases}$

Let $\mathcal{S}_c = \{(x_1, x_2) : x_1 \neq x_2\}$ and $\mathcal{S}_s = \{(x_1, x_2) : x_1 = x_2\}$. Then $\mathcal{S} = \mathcal{S}_c \cup \mathcal{S}_s$ and the suitable measure for MOBE distribution is

$$\nu(\mathcal{A}) = \nu_1([\mathcal{A} \cap \mathcal{S}_s]_p) + \nu_2(\mathcal{A}), \tag{20}$$

where $\nu_1$ and $\nu_2$ are one and two dimensional Lebesgue measures with $\nu_1(\mathcal{S}_s) = \lambda_{12}/\lambda$ and $\nu_2(\mathcal{S}_c) = (\lambda_1 + \lambda_2)/\lambda$; and the subscript $p$ denotes the projection of the set onto the $\boldsymbol{x}_1$-axis. This measure was used by Bemis et al [8] for maximum likelihood estimation of the MOBE parameters. The density $f_\nu(x_1, x_2)$ shown in Table 3 for the MOBE distribution is relative to this measure.

The MOBE density can be represented in the ME form (8) as:

$$f^*_\nu(x_1,x_2) = \lambda_{12}\ e^{-\lambda_1 x_1 - \lambda_2 x_2 - \lambda_{12}\max(x_1,x_2) - \lambda_3 D(x_1<x_2) - \lambda_4 D(x_1=x_2)}, \tag{21}$$

where

$$D(z) = \begin{cases} 1 & \text{if } z \text{ holds} \\ 0 & \text{otherwise.} \end{cases}$$

Thus, MOBE distribution is ME in the class of distributions generated by the five moments shown in Table 4. Note that $E_\nu[D_k(X_1 < X_2)] = \pi_1$ and $E_\nu[D_k(X_1 > X_2)] = \pi_2$ where

$$\pi_k = \nu(\mathcal{S}_k) = \int_{\mathcal{S}_k} f(x_1,x_2)d\nu(x_1,x_2) = \frac{\lambda_k}{\lambda},\ k = 1,2$$

where $\mathcal{S}_k = \{(x_1,,x_2) : x_k < x_\ell,\quad \ell \neq k = 1,2\}$. Thus, for the representation (21) the MOBE model parameters are subject the following constraints

$$\lambda_3 = \log\left(\frac{\lambda_{12}}{\lambda_1(\lambda_2+\lambda_{12})}\right),\quad \lambda_4 = \log\left(\frac{\lambda_{12}}{\lambda_2(\lambda_1+\lambda_{12})}\right).$$

That is, the roles of the last two constraints are the parameter restrictions required for the identification of the three branches of the MOBE distribution. Note that the five constraints for ME characterization of MOBE distribution corresponds to five independent sufficient statistics that appear in the maximum likelihood equations given in [8]. The last two constraints correspond to $n_k = \sum_{i=1}^{n} D_k(x_{1i} > x_{2i}),\ k = 1,2.$

The ME characterization of the Cuadras-Augé [13] bivariate distribution shown in Table 3 is obtained using (20) where $\nu_1(\mathcal{S}_s) = \alpha/(2-\alpha)$ and $\nu_2(\mathcal{S}_c) = 2(1-\alpha)/(2-\alpha)$ are the probabilities for the singular and continuous parts of $\mathcal{S}$, respectively. The joint density shown in Table 3 is with respect to $\nu$. It may be written in the ME form (8) as

$$f_\nu(x_1,x_2) = (1-\alpha)e^{-\lambda_1\log x_1 - \lambda_2\log x_2 - \lambda_3\log[\min(x_1,x_2)] - \lambda_4 D(x_1=x_2)},\quad 0 \le x_1,x_2 \le 1,\qquad 0<\alpha<1.$$

where $\lambda_1 = \lambda_2 = \alpha,\ \lambda_3 = -\alpha$ and $\lambda_4 = -\log(1/\alpha - 1)$.

The natural bivariate exponential (NBE) distribution shown in Table 3 has two singular parts on lines $x = \alpha y$ and $x = \beta y$. Thus, $\mathcal{S}_s = \{(x_1,x_2) : x_1 = \alpha x_2 \ge 0\} \cup \{(x_1,x_2) : x_1 = \beta x_2 \ge 0\}$ and a continuous part $\mathcal{S}_c = \{(x_1,x_2) : \alpha x_2 < x_1 < \beta x_2\}$. The ME characterization is obtained using a measure similar to (20), however with two one-dimensional Lebesgue components $\nu(\mathcal{A}) = \nu_1([\mathcal{A}\cap\mathcal{S}_s]_{p_\alpha}) + \nu_1([\mathcal{A}\cap\mathcal{S}_s]_{p_\beta}) + \nu_2(\mathcal{A})$, where as before $\nu_1$ and $\nu_2$ are one and two dimensional Lebesgue measures, and the subscripts $p_\alpha$ and $p_\beta$ denote the projections of the set onto the lines $x_1 = \alpha x_2$ and $x_1 = \beta x_2$, respectively. The joint density of NBE distribution with respect to this measure is shown in Table 3. It may be written in the ME form (8) as

$$f_\nu(x_1,x_2) = \alpha(1-\alpha)\beta e^{-\lambda_1 x_1 - \lambda_2 x_2 - \lambda_3\max(x_1,\alpha x_2) - \lambda_4\max(x_1,\beta x_2) - \lambda_5 D(x_1=\alpha x_2) - \lambda_6 D(x_1=\beta x_2)},$$
$$x_1,x_2 \ge 0,\qquad 0<\alpha<\beta<1,$$

where $\lambda_1 + \lambda_3 = \lambda,\ \lambda_2 + \lambda_4/\beta = (1-\lambda)\beta,\ \lambda_5 = -\log(\alpha\beta),\ \lambda_6 = -\log[(1-\alpha)\beta]$, and $\lambda = \dfrac{\beta - 1}{\beta - \alpha}$.

The exponential autoregressive (EAR) bivariate distribution shown in Table 3 is the distribution of adjacent terms in the first order autoregressive process $X_n = \rho X_{n-1} + \epsilon_n$, where $\{\epsilon_n\}$ is a sequence of independent and identical exponential variate with survival function $\bar{F}(\epsilon) = P(\epsilon_n > \epsilon) = \rho + (1-\rho)e^{-\lambda\epsilon},\ \epsilon \geq 0,\ \lambda > 0$, and $0 \leq \rho < 1$, [24]. Any pair $(X_n, X_{n+k})$ is stationary with a bivariate exponential distribution. Gaver and Lewis [24] give the Laplace transform for the joint distribution of $(X_n, X_{n+1})$. The bivariate distribution has a singular part (one dimensional Lebesgue ) on the support $\mathcal{S}_s = \{(x_n, x_{n+1}) : x_{n+1} = \rho x_n > 0\}$ with probability $\rho$ and a continuous part (two dimensional Lebesgue) on the support $\mathcal{S}_c = \{(x_n, x_{n+1}) : x_{n+1} > \rho x_n > 0\}$ with probability $1-\rho$. The measure $\nu(A)$ therefore is the same as (20) with $\nu_1 = \rho$ and $\nu_2 = 1-\rho$. The bivariate density shown in Table 3 is relative to this measure. It can be written in ME form (8) as

$$f_\nu(x_n, x_{n+1}) = (1-\rho)\lambda^2 e^{-\lambda_1 x_n - \lambda_2 x_{n+1} - \lambda_3 D(x_{n+1} = \rho x_n)},\quad x_{n+1} \geq \rho x_n,\ \lambda > 0,\ 0 \leq \rho < 1,$$

where $\lambda_1 = (1-\rho)\lambda,\ \lambda_2 = \lambda$, and $\lambda_3 = \log[(1/\rho - 1)\lambda]$. Note that the EAR bivariate distribution with $\lambda = 1$ can be obtained as a limiting case of NBE as $\alpha \to 0$ and $\beta = 1/\rho$.

# 4 Maximum Entropy Transformation

This section shows that ME characterizations of numerous distributions can be obtained by transformation. Unlike the discrimination function (4), the entropy is not invariant under one-to-one transformations of $\boldsymbol{X}$.

Let $\phi : \Re^p \to \Re^p$ be one-to-one transformation and let $\boldsymbol{Y} = \phi(\boldsymbol{X})$. Then the well-known entropy transformation formula gives

$$H(F_Y) = H(F_X) + E[T_\phi(\boldsymbol{Y})], \tag{22}$$

where

$$T_\phi(\boldsymbol{y}) = \log \left| \left[ \frac{\partial \phi^{-1}(y_i)}{\partial y_k} \right] \right|,\quad i, k = 1, \cdots, d$$

is the logarithm of the Jacobian of transformation.

Kapur [29] presents a few examples of distributions that are obtained from ME models by transformations, without relating the information moments for ME characterizations of $F_X$ and $F_Y$. From the relationship $H(F_Y) = H(F_X) + \log|A|$, for the affine transformation $\boldsymbol{Y} = A\boldsymbol{X} + \boldsymbol{b}$ with a nonsingular $A$, Auglogiaris and Zografos [7] and Zografos [45] deduced the relationship between characterizations of $F_Y^*$ and $F_X^*$ for some particular distributions. The next result identifies the class of distributions in which the distribution of an arbitrary one-to-one transformation of $\boldsymbol{X}$ is the ME model.

**Lemma 3 (Maximum Entropy Transformation)** *Let $F_X^*$ be the ME model in the moment class of distributions (2) generated by $\mathcal{T}_X = \{T_j(\boldsymbol{X}), j = 1, \cdots, J\}$. Let $\phi : \Re^p \to \Re^p$ be one-to-one transformation. Then the distribution of $\boldsymbol{Y} = \phi(\boldsymbol{X})$ is the ME model $F_Y^*$ in the moment class of distributions generated by $\mathcal{T}_Y = \{T_j[\phi^{-1}(\boldsymbol{Y})], j = 1, \cdots, J\} \cup \mathcal{T}_\phi(\boldsymbol{Y})$. where $\mathcal{T}_\phi(\boldsymbol{Y})$ is the set of moments generated by the Jacobian of transformation, provided all the moments in $\mathcal{T}_Y$.*

This result is seen by noting that

$$\begin{aligned} f_Y^*(\boldsymbol{y}) &= f_X^*[\phi^{-1}(\boldsymbol{y})] \left| \left[ \frac{\partial \phi^{-1}(y_i)}{\partial y_k} \right] \right| \\ &= C(\lambda, \mathcal{S}_y) \left| \left[ \frac{\partial \phi^{-1}(y_i)}{\partial y_k} \right] \right| e^{-\lambda_1 T_1[\phi^{-1}(\boldsymbol{x})] - \cdots - \lambda_J T_J[\phi^{-1}(\boldsymbol{x})]} \\ &= C(\lambda, \mathcal{S}_y) e^{\log T_\phi^{-1}(\boldsymbol{y}) - \lambda_1 T_1[\phi^{-1}(\boldsymbol{y})] - \cdots - \lambda_J T_J[\phi(\boldsymbol{y})]}, \end{aligned}$$

which is in the form of ME model (8).

The following example illustrates some points regarding implementation of Lemma 3.

## Example 4

(a) Consider $\boldsymbol{X} = (X_1, X_2)$ having a bivariate normal distribution with density (13). Let $y_i = \phi(x_i) = e^{x_i},\ i = 1, 2$. The distribution of $\boldsymbol{Y} = (Y_1, Y_2)$ is bivariate lognormal with density

$$\begin{aligned} f_Y(y_1, y_2) &= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}\ y_1 y_2} \\ &\quad \times exp\left\{ -\frac{1}{2(1-\rho^2)} \left[ \frac{(\log y_1 - \mu_1)^2}{\sigma_1^2} - \frac{2\rho(\log y_1 - \mu_1)(\log y_2 - \mu_2)}{\sigma_1\sigma_2} + \frac{(\log y_2 - \mu_2)^2}{\sigma_2^2} \right] \right\}. \end{aligned}$$

The ME characterization of multivariate lognormal, which is given in Kapur (1988) where the proof uses the usual calculus of variations technique for ME. The lognormal density $f_Y$ can be written in the ME form (8). However, application of Lemma 3 is illuminating. Noting that $\phi^{-1}(y_i) = \log y_i$ and the Jacobian of transformation is $\frac{1}{y_1 y_2}$, in this case $\{T_\phi(\boldsymbol{Y})\} \subset \{T_j[\phi^{-1}(\boldsymbol{Y})], j = 1, \cdots, 5\}$, so no additional moment is needed, i.e., $\mathcal{T}_Y = \{T_j[\phi^{-1}(\boldsymbol{Y})], j = 1, \cdots, 5\}$

(b) Since $\mu_i = E(X_i),\ i = 1, 2$ are location parameters for the normal distribution, they are not needed for the ME characterization. The information moment set

$$\mathcal{T}_X = \{T_1(\boldsymbol{X}) = X_1^2,\ T_2(\boldsymbol{X}) = X_2^2,\ T_3(\boldsymbol{X}) = X_1 X_2\}$$

characterizes the centered bivariate normal distribution with $\mu_i = 0,\ i = 1, 2$. Again let $y_i = \phi(x_i) = e^{x_i},\ i = 1, 2$. The distribution of $\boldsymbol{Y}$ is "centered" bivariate lognormal with

$E[\log(X_i)] = \mu_i = 0,\ i = 1, 2$. In this case,

$$\begin{aligned} \mathcal{T}_Y &= \{T_\phi(\boldsymbol{Y})\} \cup \{T_j[\phi^{-1}(\boldsymbol{Y})], j = 1, \cdots, 3\} \\ &= \{\log Y_1, \log Y_2, (\log Y_1)^2, (\log Y_2)^2, (\log Y_1)(\log Y_2)\}. \end{aligned}$$

That is, two more moments in addition to the moments for the centered normal are needed.

(c) Consider $\boldsymbol{X} = (\boldsymbol{X}_1, X_2)$ having a bivariate Pareto distribution with density shown in Table 1. Let $y_i = \phi(x_i) = e^{x_i+1},\ i = 1, 2$. The Jacobian of transformation is $\frac{1}{y_1 y_2}$, however, in this case $\log Y_i,\ i = 1, 2$ is not integrable and the distribution of $\boldsymbol{Y} = (Y_1, Y_2)$ is not ME; it has density (15).

Table 4 shows more examples of bivariate distributions $f_Y$ where the components of $\boldsymbol{Y}$ is obtained by one-to-one transformations of random variables $\boldsymbol{X}$ whose distributions are presented in previous. Other than the inverted Dirichlet distribution (also referred to as beta prime), all examples are component-wise transformations illustrating various issues of interest. The inverted Dirichlet distribution is an example of vector transformation. Densities of these distributions are given in an Appendix.

The bivariate Weibull conditionals is obtained by the power transformation of the bivariate exponential conditional shown in Table 1, which is ME subject to three moment constraints. The Jacobian is $\frac{1}{y_1 y_2}$, and thus $T_\phi(\boldsymbol{Y}) = \{\log Y_1, \log Y_2\}$ which adds two constraints to $\mathcal{T}_X$.

The generalized gamma-gamma mix is related to gamma-gamma mix distribution shown in Table 1, which is ME subject to four moment constraints. The transformation is the same as the Weibull conditionals case. However, unlike the Weibull case no additional constraints are needed. Here, $\{T_\phi(\boldsymbol{Y})\} \subset \{T_j[\phi^{-1}(\boldsymbol{Y})],\ j = 1, \cdots, 4\}$.

The bivariate logistic distribution is obtained by the log transformation of the bivariate Pareto shown in Table 1, which is ME subject to a single moment constraint. The Jacobian is $e^{y_1} e^{y_2}$, and thus $T_\phi(\boldsymbol{Y}) = \{Y_1, Y_2\}$ which adds two mean constraints to $\mathcal{T}_X$. Logistic distribution is presented in Kapur (1988) as an example of transformation of Dirichlet, without exploring the relationship between the sets of moment constraints for $f_X$ and $f_Y$ (the support is identified erroneously).

The bivariate Gumbel conditionals distribution is obtained by the log transformation of the bivariate exponential conditionals shown in Table 1, which is ME subject to three moment constraints. The Jacobian is $e^{y_1} e^{y_2}$, and thus $T_\phi(\boldsymbol{Y}) = \{Y_1, Y_2\}$ which adds two mean constraints to $\mathcal{T}_X$.

The Muliere and Scarsini's Pareto distribution is obtained by the exponential transformation of the Marshall-Olkin distribution shown in Table 3. The Jacobian is $\frac{1}{y_1 y_2}$, and thus $T_\phi(\boldsymbol{Y}) = \{\log Y_1, \log Y_2\}$ which does not add any constraint to $\mathcal{T}_X$, because $\{T_\phi(\boldsymbol{Y})\} \subset \{T_j[\phi^{-1}(\boldsymbol{Y})],\ j = 1, \cdots, 5\}$.

Table 4. Examples of Maximum Entropy Bivariate Distributions Obtained by Transformation

| ME Distributions | Transformation $\phi(\boldsymbol{x})$ | Information Set $\mathcal{T}_Y$ |
|---|---|---|
| Bivariate Weibull Conditional, $y_1, y_2 \geq 0$<br>Bivariate Exponential Conditional, $x_1, x_2 \geq 0$ | $y_i = x_i^{1/\tau_i}$ | $\begin{cases} T_1(\boldsymbol{Y}) = Y_1^{\tau_1},\ T_2(\boldsymbol{Y}) = Y_2^{\tau_2}, \\ T_3(\boldsymbol{Y}) = Y_1^{\tau_1} Y_2^{\tau_2} \\ T_4(\boldsymbol{Y}) = \log Y_1,\ T_5(\boldsymbol{Y}) = \log Y_2 \end{cases}$ |
| Generalized Gamma-Gamma Mix, $y_1, y_2 \geq 0$<br>Gamma-Gamma Mix, $x_1, x_2 \geq 0$ | $y_i = x_i^{1/\tau_i}$ | $\begin{cases} T_1(\boldsymbol{Y}) = Y_1^{\tau_1},\ T_2(\boldsymbol{Y}) = \log Y_1, \\ T_3(\boldsymbol{Y}) = \log Y_2,\ T_4(\boldsymbol{Y}) = Y_1^{\tau_1} Y_2^{\tau_2} \end{cases}$ |
| Logistic, $-\infty < y_1, y_2 < \infty$<br>Pareto, $x_1, x_2 > 0$ | $y_i = \log x_i$ | $\begin{cases} T_1(\boldsymbol{Y}) = \log\left(1 + e^{-Y_1} + e^{-Y_2}\right), \\ T_2(\boldsymbol{Y}) = Y_1,\ T_3(\boldsymbol{Y}) = Y_2 \end{cases}$ |
| Bivariate Gumbel conditionals, $-\infty < x_1, x_2 < \infty$<br>Bivariate exponential conditionals, $x_1, x_2 > 0$ | $y_i = -\log x_i$ | $\begin{cases} T_1(\boldsymbol{Y}) = e^{-Y_1},\ T_2(\boldsymbol{Y}) = e^{-Y_2}, \\ T_3(\boldsymbol{Y}) = e^{-Y_1 - Y_2}, \\ T_4(\boldsymbol{Y}) = Y_1,\ T_5(\boldsymbol{Y}) = Y_2, \end{cases}$ |
| Muliere and Scarsini's Pareto , $y_1, y_2 > 1$<br>Marshall-Olkin | $y_i = e^{x_i}$ | $T_1(\boldsymbol{Y}) = \log Y_1,\ T_2(\boldsymbol{Y}) = \log Y_2$<br>$T_3(\boldsymbol{Y}) = \max(Y_1, Y_2)$,<br>$T_4(\boldsymbol{Y}) = D(Y_1 < Y_2)$,<br>$T_5(\boldsymbol{Y}) = D(Y_1 = Y_2)$ |
| Inverted Dirichlet, $y_1, y_2 \geq 0$<br>Dirichlet, $0 \leq x_1, x_2 \leq 1,\ x_1 + x_2 \leq 1$ | $y_i = \dfrac{x_i}{1 - x_1 - x_2}$ | $\begin{cases} T_1(\boldsymbol{Y}) = \log Y_1,\ T_2(\boldsymbol{Y}) = \log Y_2 \\ T_3(\boldsymbol{Y}) = \log(1 + Y_1 + Y_2) \end{cases}$ |

Note:

* $D(z) = \begin{cases} 1 & \text{if } z \text{ holds} \\ 0 & \text{otherwise} \end{cases}$

The bivariate inverted Dirichlet distribution is related to the Dirichlet distribution shown in Table 2, which is ME subject to three moment constraints. In this case, $T_1[\phi^{-1}(\boldsymbol{Y})] = \log y_1, T_2[\phi^{-1}(\boldsymbol{Y})] = \log y_2$, and $T_3[\phi^{-1}(\boldsymbol{Y})] = \log(1 + Y_1 + Y_2)$. The Jacobian is $\dfrac{1}{(1 + y_1 + y_2)^3}$, and thus $T_\phi(\boldsymbol{Y}) = \{\log(1 + Y_1 + Y_2)\}$ which adds no additional constraint to $\mathcal{T}_X$.

Asadi et al [5] lists more than twenty families of distributions that can be obtained from the univariate version of Pareto distribution shown in Table 1. These distributions and their transformation functions are given in [5]. The ME characterizations of bivariate (multivariate) versions of these distributions can easily be obtained from the Pareto distribution by Lemma 3.

# 5 Dependence Information

An important information function is defined when the reference distribution in (4) is the product(s) of some marginals. The mutual information between two subvectors $\boldsymbol{X}_a$ and $\boldsymbol{X}_b$ has Kullback-Leibler and entropy representations

$$\begin{aligned} M(\boldsymbol{X}_a, \boldsymbol{X}_b | \mathcal{S}) &= K(F : F_{\boldsymbol{X}_a} F_{\boldsymbol{X}_b} | \mathcal{S}) \\ &= H(F_{\boldsymbol{X}_a} | \mathcal{S}_a) + H(F_{\boldsymbol{X}_b} | \mathcal{S}_b) - H(F | \mathcal{S}), \end{aligned} \tag{23}$$

where $F_{\boldsymbol{X}_a}$ and $F_{\boldsymbol{X}_b}$ are the marginal distributions of $\boldsymbol{X}_a$ and $\boldsymbol{X}_b$, and $a + b \geq d$. For $a + b > d$, (23) gives a measure of conditional dependence between $\boldsymbol{X}_a$ and $\boldsymbol{X}_b$, given the common components $\boldsymbol{X}_{ab} = \boldsymbol{x}_{ab}$. For $a + b = d$, (23) gives a measure of dependence between $\boldsymbol{X}_a$ and $\boldsymbol{X}_b$. The mutual information for more than two subvectors are defined similarly. For simplicity, we consider the bivariate case. The results given in this section extends to higher dimensions when $a + b = d$.

The mutual information between to random variables is given by

$$\begin{aligned} M(X_1, X_2|\mathcal{S}) &= K(F : F_1F_2)|\mathcal{S}] \\ &= H(F_1|\mathcal{S}_1) + H(F_2|\mathcal{S}_2) - H(F|\mathcal{S}), \end{aligned} \tag{24}$$

where $F = F_{X_1,X_2}$, $F_1 = F_{X_1}$, and $F_2 = F_{X_2}$ are the joint and marginal distributions, respectively. $M(X_1, X_2|\mathcal{S}) \geq 0$, where the equality holds if and only if the variables are independent. This is a widely used measure of dependence in statistics and many other fields (see, e.g, [28, 41]).

Note that by the Kullback-Leibler representation in (24), the mutual information is well-defined when the joint distribution is absolutely continuous relative to the product $F_1F_2$. For example, for Marshall-Olkin distribution all three entropies in (24) can be computed, but the mutual information is not well-defined due to the singularity. The product $F_1F_2$ has bivariate density relative to two-dimensional Lebesgue, but the bivariate distribution is not absolutely continuous. In fact, the entropy representation gives negative values.

It is well-known that $M(X_1, X_2)$ is invariant under one-to-one transformations of each component. For example, for the bivariate normal and lognormal, and for the first four pairs of distributions listed in Table 4, $M(Y_1, Y_2) = M(X_1, X_2)$. The transformation for last pair of distributions shown in Table 4, inverted Dirichlet and Dirichlet, is not component-wise and it can be shown that $M(Y_1, Y_2) \neq M(X_1, X_2)$.

The MDI model relative to given marginal distributions, minimizes the mutual information and has density in the form of

$$f^*(x_1, x_2|\mathcal{S}) = C(\lambda, \mathcal{S}) f_1(x_1) f_2(x_2) e^{-\lambda_1 T_1(x_1,x_2) - \cdots - \lambda_J T_J(x_1,x_2)}. \tag{25}$$

For specified marginal distributions, the marginal entropies $H(f_1)$ and $H(f_2)$ are determined. Then from the entropy representation in (24) it is clear that minimization of $M(X_1, X_2)$ is equivalent to the maximization of the joint entropy $H(F)$.

Dependence in an ME model is induced by the joint moments and/or the support $\mathcal{S}$. Examples of some distributions where the dependence is only due to the joint moments are bivariate normal and lognormal and other distributions with rectangular supports shown in Table 1 and Table 3. The bivariate normal marginals and bivariate exponential on triangle shown in Table 2 are examples of distributions where the dependence is only due to the support. Other distributions shown in Table 2 are examples where dependence is due to joint moments and support.

The ME distributions with joint moment constraints are two types in terms of the ME properties of their marginal distributions. For example, the marginal distributions of bivariate normal, lognormal, Gamma-Gamma, and Marshall-Olkin's bivariate exponential are ME models in the univariate classes of distributions defined by the marginal moments in their moment information sets. The marginal distributions of bivariate $t$, Pareto, bivariate conditional exponential, and absolutely continuous bivariate exponential are not ME models in the univariate classes of distributions defined by the marginal moments in their moment information sets.

Let $\mathcal{T}^*_{X_k} = \{T^*_{j_k}(X_k),\ j_k = 1, \cdots, J_k\}, \quad k = 1, 2,$ be the moment information sets that characterize the marginal distributions of $F$ as the ME model $F^*_k$. Then the marginal distributions of $F^*$ are ME subject to the marginal moments in $\mathcal{T}$ if and only if each $\mathcal{T}^*_{X_k},\ k = 1, 2$ is congruent to a subset $\mathcal{T}_{X_k} \subseteq \mathcal{T},\ k = 1, 2$. The next result gives an information characterization of ME nested models.

**Lemma 4** *The components of an ME model $F^*$ are independent if and only if $\mathcal{T} \cong \mathcal{T}^*_{X_1} \cup \mathcal{T}^*_{X_2}$ and $\mathcal{S} = \mathcal{S}_1 \times \mathcal{S}_2$, where $\mathcal{S}_k,\ k = 1, 2$ does not depend on $x_\ell,\ \ell \neq k = 1, 2$.*

**Proof.** For the case of $J_1 + J_2 = J$, without loss of generality, let $\mathcal{T}_{X_1} = \{T_j(\boldsymbol{X}),\ j = 1, \cdots J_1\}$ and $\mathcal{T}_{X_2} = \{T_j(\boldsymbol{X}),\ j = J_1 + 1, \cdots J\}$. Then it can be seen from (8) that the ME model factors as

$$f^*(x_1, x_2|\mathcal{S}) = C(\lambda, \mathcal{S}) w_1(x_1) w_2(x_2),$$

where

$$\begin{aligned} w_1(x_1) &= e^{-\lambda_1 T_1(x_1) - \cdots - \lambda_{J_1} T_{J_1}(x_1)} \\ w_2(x_2) &= e^{-\lambda_{J_1+1} T_{J_1+1}(x_2) - \cdots - \lambda_J T_J(x_2)}. \end{aligned}$$

Observe that $w_k(x_k)$ is the univariate version of the kernel of the ME density (8). That is, we must have $C(\lambda, \mathcal{S}) = C_1(\lambda, \mathcal{S}_1) C_2(\lambda, \mathcal{S}_2)$, and

$$f^*_k(x_k) = C_k(\lambda, \mathcal{S}) w_k(x_k), \quad k = 1, 2$$

is the density of the univariate ME distribution subject to the respective moment constraints.

This result confirms the well-known belief that since the ME is maximally noncommittal to missing information, in the absence of information on the dependence, the ME solution gives independent components.

When the marginal distributions are not ME subject to the given marginal moments in $\mathcal{T}$ and the dependence is only due to a set of joint moments in $\mathcal{T}$, the marginal distributions are given by

$$f_k(x_k) = C(\lambda, \mathcal{S}) W_k(x_k) w_k(x_k), \quad k = 1, 2. \tag{26}$$

where

$$W_k(x_k) = C(\lambda, \mathcal{S}) \int_{\mathcal{S}_\ell} \boldsymbol{w}_\ell(x_\ell) dx_\ell, \quad k \neq \ell = 1, 2. \tag{27}$$

That is, the marginal distributions are ME subject to a set of additional marginal moments which are not in $\mathcal{T}$ of the bivariate ME distributions. The marginal entropies therefore are given by (9) that includes these additional moments. Because of these additional constraints, the marginal entropies are lower than the above case.

The support induces dependence when its boundary is a function of $x_1$ and $x_2$. It is well-known that for a rectangular support (e.g., $\mathcal{S} = \Re^2$, the first quadrant $\mathcal{S} = Q_1$, the unit square $\mathcal{S} = Sq$), with sides parallel to the coordinate axes when the boundary does not depend on $x_1$ and $x_2$, the product decomposition implies independence (see, [17]). Note that for such a rectangular support the dependence is only due to the joint moments and in the absence of a joint moment $M(X_1, X_2) = 0$.

The incremental contribution of a subset of moment information set $\{T_j(\boldsymbol{X}),\ j = q{+}1, \cdots, J\} \subseteq \mathcal{T}$ to the information content of another subset $\{T_j(\boldsymbol{X}),\ j = 1, \cdots, q\} \subseteq \mathcal{T}$ is measured by

$$K(F^*_{1,\cdots,J} : F^*_{q+1,\cdots,J}|\mathcal{S}) = H(F^*_{q+1,\cdots,J}|\mathcal{S}) - H(F^*_{1,\cdots,J}|\mathcal{S}); \tag{28}$$

$K(F^*_{1,\cdots,J} : F^*_{q+1,\cdots,J}|\mathcal{S}) \geq 0$, where the equality holds when the additional constraints are redundant.

Next result gives mutual information for ME nested and non-nested models.

**Lemma 5** *Let $\mathcal{T}$ and $\mathcal{T}^*_{X_k},\ k = 1, 2$ be the information moment sets for the joint and marginal distributions.*

*(a) If each $\mathcal{T}^*_{X_k},\ k = 1, 2$ is congruent to a subset $\mathcal{T}_{X_k} \subseteq \mathcal{T},\ k = 1, 2$, then, the mutual information (24) is given by the ME difference (28):*

$$\begin{aligned} M(X_1, X_2) &= H(F^*_{1,\cdots,m}|\mathcal{S}) - H(F^*_{1,\cdots,J}|\mathcal{S}) \\ &= \log C(\lambda, \mathcal{S}) - \log C_1(\lambda, \mathcal{S}) - \log C_2(\lambda, \mathcal{S}) - \lambda_{m+1}\theta_{m+1} - \cdots - \lambda_J \theta_J, \end{aligned}$$

*where, without loss of generality, $\mathcal{T}_{X_1} = \{T_j(X_1, X_2) = T_j(X_1),\ j = 1, \cdots, q\}$ and $\mathcal{T}_{X_2} = \{T_j(X_1, X_2) = T_j(X_2),\ j = q + 1, \cdots, m\}$, with $q < m \leq J$.*

*(b) If each $\mathcal{T}^*_{X_k},\ k = 1, 2$ is not congruent to a subset $\mathcal{T}_{X_k} \subseteq \mathcal{T},\ k = 1, 2$, then,*

$$M(X_1, X_2) = \log C(\lambda, \mathcal{S}) - E \log[W_1(X_1)] - E \log[W_2(X_2)] + \Delta'\theta, \tag{29}$$

*where $W_k(X_k)$ is defined in (27), $\Delta = (\Delta_1, \cdots, \Delta_m, -1, \cdots, -1)$, $\Delta_j = \lambda_{j1} - \lambda_{j2}$ is the vector of the differences between the ME marginal and joint model parameters for the moments $\theta = (\theta_1, \cdots, \theta_J)$.*

**Example 5**

(a) The ME model subject to the four marginal constraints shown in (14) for the bivariate normal distribution of Example 2 is the independent normal model with $H(F^*_{1,2,3,4}|\Re^2) = 1 + \log(2\pi\sigma_1\sigma_2)$. With the addition of $E(X_1X_2)$ as the fifth constraint, the marginals are still normal and $H(F^*_{1,2,3,4,5}|\Re^2) = 1+\log(2\pi\sigma_1\sigma_2)+.5\log(1-\rho^2)$, where $\rho$ is the correlation coefficient. We have

$$M(X_1, X_2) = K(F^*_{1,2,3,4,5} : F^*_{1,2,3,4}) = H(F^*_{1,2,3,4}|\Re^2) - H(F^*_{1,2,3,4,5}|\Re^2) = -.5\log(1-\rho^2).$$

(b) The ME model subject to the two marginal constraints shown in shown in Table 2 for the bivariate distribution with normal marginals is not the independent normal model because the support is a function of $x_1x_2$. The mutual information for this distribution is simply $M(X_1, X_2) = \log 2$.

**Example 6**

Consider the moment information set $\mathcal{T} = \{T_1(X_1, X_2) = X_1, T_2(X_1, X_2) = X_2\}$.

(a) The ME model with $\mathcal{S} = \Re^2$ does not exist.

(b) Let $\mathcal{S} = Q_1 = \{(x_1, x_2) : x_1, x_2 \geq 0\}$. Then the ME model is independent bivariate exponential. The ME mode entropy is

$$H(X_1, X_2) = H(F^*_{1,2}) = 2 - \log\lambda^*_1 - \log\lambda^*_2,$$

where the parameters are determined by the moments $E(X_k) = \dfrac{1}{\lambda^*_k} = \mu^*_k, k = 1, 2$.

(c) Consider the moment information set $\mathcal{T} = \{T_1(X_1, X_2) = X_1, T_2(X_1, X_2) = X_2, T_3(X_1, X_2) = X_1X_2\}$, and the rectangular support $\mathcal{S} = Q_1$. Then the ME model is bivariate exponential conditional (BEC) model shown in Table 1. The marginal distributions of BEC are in the form of (26) with

$$W_k(x_k) = \frac{1}{1 + \delta\lambda_k x_k}, \quad k = 1, 2.$$

The marginal information moment sets $\mathcal{T}^*_{X_k} = \{T_{1_k}(X_k) = X_k, T_{2_k}(X_k) = \log(1 + \delta\lambda_k X_k),\ k = 1, 2\}$ are not congruent to any subset of $\mathcal{T}$, so the dependence is not only through $E(X_1X_2)$. In this case the ME model entropy is

$$H(X_1, X_2) = H(F^*_{1,2,3}) = 1 - \log[\lambda_1\lambda_2 C(\delta)] + \frac{C(\delta) - 1}{\delta}$$

where the BEC parameters are determined by the marginal and joint moments given by

$$E(X_k) = \frac{c(\delta) - 1}{\delta\lambda_k} = \mu_k, \quad k = 1, 2 \quad \text{and} \quad E(X_1X_2) = \frac{1 + \delta - c(\delta)}{\delta\lambda_3} = \mu_{12}.$$

The ME entropy difference measure (28) is given by

$$K(F^*_{1,2,3}:F^*_{1,2}) = H(F^*_{1,2}) - H(F^*_{1,2,3}) = 1 + \log\left(\frac{\lambda_1\lambda_2}{\lambda_1^*\lambda_2^*}\right) + \log[c(\delta)] - \frac{c(\delta)-1}{\delta}, \qquad (30)$$

where $\lambda_1^*$ and $\lambda_2^*$ are the parameters of the independent exponential model of (b). Thus, (30) measures the incremental contribution of the joint moment $E(X_1X_2) = \mu_{12}$ to the independent exponential. The mutual information measures departure of the BEC model from the product of its marginals. It is given by

$$M(X_1, X_2) = -\log[c(\delta)] + \zeta_1 + \zeta_2 + \frac{c(\delta)-1}{\delta} - 1,$$

where $\zeta_k = E[\log(W_k(X_k)] = E[\log(1 + \delta\lambda_k X_k)]$.

(d) Now let $\mathcal{S} = \mathcal{S}_{TR} = \{(x_1, x_2) : 0 < x_1 < x_2\}$. From Table 2 we find that the bivariate exponential on the triangle with $\alpha = 0$ and $\beta = 1$ gives Half ACBE on the triangle as the ME model. The marginal distributions have densities:

$$f_1(x_1) = (\lambda_1 + \lambda_2)e^{-(\lambda_1+\lambda_2)x_1}, \quad x_1 > 0$$

$$f_2(x_2) = \frac{\lambda_2}{\lambda_1}(\lambda_1 + \lambda_2)e^{-\lambda_2 x_2}\left(1 - e^{-\lambda_1 x_2}\right), \quad x_2 > 0.$$

The ME model parameters are determined by

$$E(X_1) = \frac{1}{\lambda_1 + \lambda_2} = \mu_1, \quad \text{and} \quad E(X_2) = \frac{1}{\lambda_2} + \frac{1}{\lambda_1 + \lambda_2} = \mu_2.$$

For computing the mutual information we let $Y_2 = 1 - e^{-\lambda_1 X_2}$. Then $Y_2$ has a beta distribution with density

$$f_{Y_2}(y_2) = \beta(\beta + 1)y_2(1 - y_2)^{\beta-1}, \quad 0 \le y_2 \le 1,$$

where $\beta = \dfrac{\lambda_2}{\lambda_1}$. The joint distribution of $(X_1, Y_2)$ has density

$$f_{X_1,Y_2}(x_1, y_2) = \beta(\lambda_1 + \lambda_2)e^{-\lambda_1 x_1}(1 - y_2)^{\beta-1}.$$

Using the Kullback-Leibler representation in (24) gives

$$\begin{aligned} M(X_1, X_2) = M(X_1, Y_2) &= -\log(\beta + 1) + \lambda_2 E(X_1) - E(\log Y_2) \\ &= \psi(\beta + 1) - \log(\beta + 1) + \gamma, \end{aligned}$$

where $\psi(\cdot)$ is the digamma function and $\gamma = -\psi(1) = .5772\cdots$ is Euler's constant. The last expression is obtained by noting that $E(\log Y_2) = \psi(2) - \psi(\beta+2)$ and using $\psi(z+1) = \psi(z) + 1/z$. Since there is no joint moment in the information set, $M(X_1, X_2)$ measures the dependency due to the support. Note that $M(X_1, X_2)$ is strictly positive, is an increasing function of $\beta > 0$, and $0 < M(X_1, X_2) < \gamma$.

## Appendix

Densities of Maximum Entropy Bivariate Distributions of Table 4 are as follows.

Bivariate Weibull Conditionals density

$$f_Y(y_1, y_2) = C(\alpha_1, \alpha_2, \tau_1\tau_2, \lambda_1, \lambda_2, \lambda_3) y_1^{\alpha_1 - 1} y_2^{\alpha_2 - 1} exp\left\{-(\lambda_1 y_1)^{\tau_1} - (\lambda_2 y_2)^{\tau_2} - \lambda_3 y_1^{\tau_1} y_2^{\tau_2}\right\}, \quad y_1, y_2 \geq 0$$

$$\alpha_1, \alpha_2, \lambda_1, \lambda_2, \lambda_3, \tau_1, \tau_2 > 0.$$

Bivariate Generalized Gamma-Gamma Mix density

$$f_Y(y_1, y_2) = \frac{\tau_1 \tau_2 \lambda_1^{\tau_1 \alpha_1} \lambda_3^{\tau_2 \alpha_2}}{\Gamma(\alpha_1)\Gamma(\alpha_2)} y_1^{\tau_1 \alpha_1 + \tau_2 \alpha_2 - 1} y_2^{\tau_2 \alpha_2 - 1} exp\left\{-\lambda_1 y_1^{\tau_1} - \lambda_3^{\tau_2} y_1^{\tau_1} y_2^{\tau_2}\right\}, \quad y_1, y_2 \geq 0$$

$$\alpha_1, \alpha_2, \lambda_1, \lambda_2, \lambda_3, \tau_1, \tau_2 > 0.$$

Bivariate Gumbel conditionals, $-\infty < x_1, x_2 < \infty$

$$f(x_1, x_2) = C(\alpha) e^{-\left(x_1 + x_2 + e^{-x_1} + e^{-x_2} + \alpha e^{-x_1 - x_2}\right)}, \quad \alpha \geq 0.$$

Bivariate logistic density

$$f_Y(y_1, y_2) = \frac{2e^{-(y_1 + y_2)}}{1 + e^{-y_1} + e^{-y_2}}, \quad -\infty < y_1, y_2 < \infty.$$

Muliere and Scarsini's Pareto survival function

$$\bar{F}(y_1, y_2) = y_1^{-\lambda_1} y_1^{-\lambda_2} \max(y_1, y_2)^{-\lambda_{12}}, \quad y_1, y_2 \geq 1, \quad \lambda_1,\ \lambda_2 > 0,\ \lambda_{12} \geq 0.$$

Bivariate inverted Dirichlet density

$$f_Y(y_1, y_2) = \frac{\Gamma(\alpha_1 + \alpha_2 + \alpha_3)}{\Gamma(\alpha_1)\Gamma(\alpha_2)\Gamma(\alpha_3)} \frac{y_1^{\alpha_1 - 1} y_2^{\alpha_2 - 1}}{(1 + y_1 + y_2)^{\alpha_1 + \alpha_2 + \alpha_3}}, \quad y_1, y_2 \geq 0$$

$$\alpha_1, \alpha_2, \alpha_3 > 0.$$

## References


[1] I. Arizono, H. Ohta, A Test for normality based on Kullback-Leibler information, The American Statistician 34 (1989) 20-23.

[2] B.C. Arnold, D. Strauss, Bivariate distribution with exponential Conditionals, Journal of the American Statistical Association 83 (1988) 522-527.

[3] B.C. Arnold, E. Castillo, J.M. Sarabia, Conditional Specification of Statistical Models, Springer, New York, 1999.

[4] B.C. Arnold, E. Castillo, J.M. Sarabia, Conditionally specified distributions: An Introduction, Statistical Science 16 (2001) 249-274.

[5] M. Asadi, N. Ebrahimi, G.G. Hamedani, E.S. Soofi, Information measures for Pareto distributions and order statistics", in Advances on Distribution Theory and Order Statistics, N. Balakrishnan, E. Castillo, and J.M. Sarabia, Birkhauser, Boston, 2006, 207-223.

[6] M. Asadi, N. Ebrahimi, G.G. Hamedani, E.S. Soofi, Maximum dynamic entropy models, Journal of Applied Probability 41 (2004) 379-390.

[7] G. Aulogiaris and K. Zografos, A Maximum Entropy Characterization of Symmetric Kotz Type and Burr Multivariate Distributions, Test 13 (2004) 65-83.

[8] B.M. Bemis, L.J. Bain, and J.J. Higgins, Estimation and hypothesis testing for the parameters of bivariate exponential distribution, Journal of the American Statistical Association 67 (1972) 927-929.

[9] H.B. Block, A.P. Basu, A continuous bivariate exponential extension, Journal of the American Statistical Association 69 (1974) 1031-1037.

[10] E. Castillo, J. Galambos, Conditional distributions and the bivariate normal distribution, Metrika 36 (1989) 209-214.

[11] B. Choi, K. Kim, Testing goodness-of-fit for Laplace distribution based on maximum entropy, Statistics 40 (2006) 517-531.

[12] T.M. Cover, J.A. Thomas, Elements of Information Theory, Wiley, New York, 1991.

[13] C.M. Cuadras, J. Augé, A continuous general multivariate distribution and its properties, Communication in Statistics – Theory and Methods 10 (1981) 339-353.

[14] A. Dadpay, E.S. Soofi, R. Soyer, Information Measures for Generalized Gamma Family, Journal of Econometrics, in press.

[15] G.A. Darbellay, I. Vajda, Entropy expressions for multivariate continuous distributions, IEEE Transactions on Information Theory, IT 46 (2000) 709-712.

[16] M.H. DeGroot, Optimal Statistical Decisions, McGraw-Hill, New York, 1970.

[17] M.H. DeGroot, Probability and Statistics 2nd ed., Addison-Wesley, Reading MA, 1986.

[18] D.J. DeWaal, Goodness of fit of the generalized extreme value distribution based on the Kullback-Leibler information, South African Statistical Journal 30 (1996) 139-153.

[19] E. J. Dudewicz, E.C. Van Der Meulen, Entropy-based Tests of uniformity, Journal of the American Statistical Association 76 (1981) 967-974.

[20] N. Ebrahimi, Testing for exponentiality of the residual lifetime based on dynamic Kullback-Leibler information, IEEE Transactions on Reliability R-47 (1998) 197-201.

[21] N. Ebrahimi, Testing for uniformity of the residual lifetime based on dynamic Kullback-Leibler information, Annals of Institute of Statistical Mathematics 53 (2001) 325-337.

[22] N. Ebrahimi, M. Habibullah, E.S. Soofi, Testing exponentiality based on Kullback-Leibler information, Journal of Royal Statistical Society B 54 (1992) 739-748.

[23] N. Ebrahimi, S.N.U.A. Kirmani, and E.S. Soofi, Dynamic multivariate information, Journal of Multivariate Analysis 98 (2007) 328-349.

[24] D.P. Gaver, P.A.W. Lewis, First-order gamma sequences and point processes, Advances in Applied Probability 12 (1980) 727-745.

[25] D. V. Gokhale, On entropy-based goodness-of-fit tests, Computational Statistics and Data Analysis 1 (1983) 157-165.

[26] E.T. Jaynes, Information theory and statistical mechanics, Physics Review 106 (1957) 620-630.

[27] E.T. Jaynes, On the rationale of maximum-entropy methods, Proceedings of IEEE 70 (1982) 939-952.

[28] H. Joe, Relative entropy measures of multivariate dependence Journal of the American Statistical Association 84 (1989) 157-164.

[29] J.N. Kapur, Maximum Entropy Models in Science and Engineering, Wiley, New York, 1989.

[30] S. Kullback, Information Theory and Statistics, Wiley, New York 1959 (reprinted in 1968 by Dover).

[31] A. W. Marshall, I. Olkin, A multivariate exponential distribution, Journal of the American Statistical Association 62 (1967) 30-44.

[32] T. A. Mazzuchi, E.S. Soofi, R. Soyer, Computations of maximum entropy Dirichlet for modeling lifetime data, Computational Statistics and Data Analysis 32 (2000) 361-378.

[33] A.T. McKay, Sampling from batches, Journal of Royal Statistical Society, Supplement 1 (1934) 207-216.

[34] G.A. Mihram and R.A. Hultquist, A bivariate warning-time/failure-time distribution Journal of the American Statistical Association 62 (1967) 589-599.

[35] G. S. Mudholkar, L. Tian, An entropy characterization of the inverse Gaussian distribution and related goodness-of-fit test, Journal of Statistical Planning and Inference 102 (2002) 211-221.

[36] S. Nadarajah, K. Zografos, Expressions for Rényi and Shannon entropies for bivariate distributions, Information Sciences 170 (2005) 173-189.

[37] S.G. Park, Testing exponentiality based on the Kullback-Leibler information with the Type II censored data, IEEE Transactions on Reliability 54 (2005) 22-26.

[38] S.G. Park, A goodness-of-fit test for normality based on the sample entropy of order statistics, Statistics and Probability Letters 44 (1999) 359-363.

[39] S.G. Park, D. Park, Correcting moments for goodness-of-fit tests based on two entropy estimates, Journal of Statistical Computation and Simulation 73 (2003) 685-694.

[40] J.E. Shore, R.W. Johnson, Axiomatic derivation of the principle of maximum entropy and principle of minimum cross-entropy, IEEE Transactions on Information Theory, IT 26 (1980) 26-37.

[41] E.S. Soofi, Capturing the intangible concept of Information, Journal of the American Statistical Association 89 (1994) 1243-1254.

[42] E.S. Soofi, N. Ebrahimi, M. Habibullah, Information distinguishability with application to analysis of failure data, Journal of the American Statistical Association 90 (1995) 657-668.

[43] O. Vasicek, A test for normality based on sample entropy, Journal of Royal Statistical Society B 38 (1976) 54-59.

[44] A. Zellner, R.A. Highfield, Calculation of maximum entropy distributions and approximation of marginal posterior distributions, Journal of Econometrics 37 (1988) 195-210.

[45] K. Zografos, On maximum entropy characterization of Pearson's Type II and VII multivariate distributions, Journal of Multivariate Analysis 71 (1999) 67-75.