Probability in the Engineering and Informational Sciences, **24**, 2010, 561–584.
doi:10.1017/S0269964810000185

MODELS BASED ON PARTIAL INFORMATION ABOUT SURVIVAL AND HAZARD GRADIENT

MAJID ASADI AND SOMAYEH ASHRAFI
Department of Statistics
University of Isfahan
Isfahan, 81744, Iran
E-mail: m.asadi@sci.ui.ac.ir

NADER EBRAHIMI
Division of Statistics
Northern Illinois University
DeKalb, IL 60155 E-mail: nader@math.niu.edu

EHSAN S. SOOFI
Sheldon B. Lubar School of Business
University of Wisconsin-Milwaukee,
Milwaukee, WI 53201
E-mail: esoofi@uwm.edu

This article develops information optimal models for the joint distribution based on partial information about the survival function or hazard gradient in terms of inequalities. In the class of all distributions that satisfy the partial information, the optimal model is characterized by well-known information criteria. General results relate these information criteria with the upper orthant and the hazard gradient orderings. Applications include information characterizations of the bivariate Farlie–Gumbel–Morgenstern, bivariate Gumbel, and bivariate generalized Gumbel, for which no other information characterization are available. The generalized bivariate Gumbel model is obtained from partial information about the survival function and hazard gradient in terms of marginal hazard rates. Other examples include dynamic information characterizations of the bivariate Lomax and generalized bivariate Gumbel models having marginals that are transformations of exponential such as Pareto, Weibull, and extreme value. Mixtures of bivariate Gumbel and generalized Gumbel are obtained from partial information given in terms of mixtures of the marginal hazard rates.

© Cambridge University Press 2010 0269-9648/10 $25.00

https://doi.org/10.1017/S0269964810000185 Published online by Cambridge University Press

1. INTRODUCTION

Consider a system consisting of n components. Let $X_i,\ i = 1, \ldots, n$, be nonnegative random variables representing the lifetimes of the components of the system. Assessing the reliability of a system requires a model for the joint distribution of the random vector $\boldsymbol{X} = (X_1, \ldots, X_n)$. For example, if we have a system in which components are arranged in series, the reliability of the system is given by $S(x) = P(\min\{X_1, \ldots, X_n\} > x) = P(X_1 > x, \ldots, X_n > x)$. Specification of a joint distribution $F(\boldsymbol{x}) = P(X_1 \leq x_1, \ldots, X_n \leq x_n)$ that gives such a probability requires *full information* about the lifetime distribution of each component and dependence among all the components.

Frequently, information is available about some characteristics of the system which can be formulated in terms of partial information that do not determine a unique model for the joint distribution of the components. When the partial information can be formulated in terms of a set of moment constraints, a unique model for the joint distribution can be found by applying the maximum entropy information criteria (MEIC) (Jaynes [12]) and minimum discrimination information criteria (MDIC) (Kullback [18]). In reliability problems, it is natural to think of constraints on the hazard rate, residual moments, hazard gradients, and so on. In the univariate case, the mean residual and hazard rate equality constraints provide full information about the distribution and uniquely determine the distribution. However, the partial information can be formulated in terms of bounds on the hazard rates and mean residual functions of the components of a system. Asadi, Ebrahimi, Hamedani, and Soofi [1] and Asadi, Ebrahimi, Hamedani, and Soofi [3] defined dynamic entropy information criteria and dynamic discrimination information criteria in terms of the entropy and Kullback–Leibler information for residual distribution and showed that these dynamic criteria provide the unique maximum entropy and minimum discrimination information models that satisfy the partial information in terms of bounds on the hazard rates and mean residual functions.

In the multivariate case, survival function $\bar{F}(\boldsymbol{x}) = P(X_1 > x_1, \ldots, X_n > x_n)$ and hazard functions are the main reliability measures for probabilistic and statistical modeling for various problems. Unlike the survival function, the multivariate extension of hazard function is not unique. The joint hazard rate of a multivariate random vector $\boldsymbol{X}$ with a joint survival function $\bar{F}$ is defined in terms of the gradient vector

$$R(\boldsymbol{x}) = (R^{(1)}(\boldsymbol{x}), \ldots, R^{(n)}(\boldsymbol{x})) = -\left(\frac{\partial}{\partial x_1} \log \bar{F}(\boldsymbol{x}), \ldots, \frac{\partial}{\partial x_n} \log \bar{F}(\boldsymbol{x})\right) \qquad (1)$$

(Johnson and Kotz [14]; Marshall [21]). The ith component of the hazard gradient can be represented as the conditional hazard rate of X_i evaluated at x_i, given that $X_j > x_j$ for all $j \neq i$:

$$R^{(i)}(\boldsymbol{x}) = \frac{f_i(x_i | X_j > x_j)}{\bar{F}_i(x_i | X_j > x_j)},$$

https://doi.org/10.1017/S0269964810000185 Published online by Cambridge University Press

where $f_i(x_i|X_j > x_j, j \neq i)$ and $\bar{F}_i(x_i|X_j > x_j, j \neq i)$ are respectively the conditional probability density function (PDF) and survival function of X_i evaluated at x_i, given that $X_j > x_j$.

The hazard gradient function is shown to be particularly useful in various problems in engineering and biostatistics (Kotz, Navarro, and Ruiz [17]). Due to its simple relationships with moments of a truncated multivariate normal distribution, the hazard gradient has been interpreted as the appropriate generalization of the univariate hazard rate (McGill [22]). Navarro and Ruiz [20] have shown that the relationships between hazard gradient and moments allow characterization of multivariate distributions. Bounds on hazard gradient have been developed that capture dependence structure for competing risks models (Karia and Deshpande [15]). Bounds on hazard gradient are partial information that do not determine a unique model for the joint distribution. We develop unique information optimal models for the joint distribution when the partial information about survival function and the hazard gradient of the system can be formulated in terms of upper or lower bounds.

Section 2 presents definitions, notations, and a brief background on information criteria used in the subsequent sections. Section 3 gives results that relate MDIC and MEIC to the upper orthant ordering and results that relate the dynamic versions of these criteria with hazard gradient ordering. This section also presents information characterizations of three bivariate models. Section 4 develops models for transformations. Section 5 gives representations of the main results that are useful for developing complicated models such as bivariate mixtures. Section 6 gives some concluding remarks.

2. PRELIMINARIES AND BACKGROUNDS

This section presents definitions, relations, and notations that are used in derivations of our results. A vector inequality $\boldsymbol{a} < \boldsymbol{b}$ denotes $a_i < b_i$ for all components. We develop models for classes of distributions defined by survival function and hazard gradient inequalities:

$$\Omega_G = \{F : \bar{F}(\boldsymbol{x}) \leq (\geq)\ \bar{G}(\boldsymbol{x})\}, \tag{2}$$

$$\Omega_Q = \{F : R(\boldsymbol{x}) \geq (\leq)\ Q(\boldsymbol{x})\}, \tag{3}$$

where $R = R_F$ and the sets are compact (i.e., $G \in \Omega_G$ and $Q = R_G, G \in \Omega_Q$). Subscript F and G for R are not used unless the context requires.

We draw from the literature of two areas: stochastic orders and information theory.

2.1. Stochastic Orders

We use the following well-known orders of multivariate distributions and their relationships. Details can be found in Shaked and Shanthikumar [23].

https://doi.org/10.1017/S0269964810000185 Published online by Cambridge University Press

DEFINITION 1: *Let $\boldsymbol{X}$ and $\boldsymbol{Y}$ be two random vectors with survival functions $\bar{F}$ and $\bar{G}$, respectively.*

(*a*) *$\boldsymbol{X}$ is said to be smaller than $\boldsymbol{Y}$ in the upper orthant order ($\boldsymbol{X} \leq_{\text{uo}} \boldsymbol{Y}$) if $\bar{F}(\boldsymbol{x}) \leq \bar{G}(\boldsymbol{x})$ for all $\boldsymbol{x} \in \Re^n$.*

(*b*) *$\boldsymbol{X}$ is said to be smaller than $\boldsymbol{Y}$ in the weak multivariate hazard rate order ($\boldsymbol{X} \leq_{\text{whr}} \boldsymbol{Y}$) if*

$$\frac{\bar{F}(\boldsymbol{x})}{\bar{G}(\boldsymbol{x})} \text{ is decreasing in } \boldsymbol{x} \in \{\boldsymbol{x} : \bar{G}(\boldsymbol{x}) > 0\}; \tag{4}$$

where $a/0 \equiv \infty$ whenever $a > 0$.

The weak hazard rate order is also referred to as *hazard gradient order* due to the following relationship. If R and Q are respectively the hazard gradients of $\boldsymbol{X}$ and $\boldsymbol{Y}$, then

$$R(\boldsymbol{x}) \geq Q(\boldsymbol{x}) \Longleftrightarrow \boldsymbol{X} \leq_{\text{whr}} \boldsymbol{Y} \tag{5}$$

[23, p. 291]. Thus, Ω_Q in (3) can be represented in terms of (4). For details on the upper orthant ordering and multivariate hazard rate orderings, see, for example [11,16,23].

DEFINITION 2: *A real-valued n-variate function $\psi(\boldsymbol{x})$ is said to be Δ-monotone of order n if for all ordered pairs of points $\boldsymbol{x} \leq \boldsymbol{y}$, the difference operator*

$$\Delta_{\boldsymbol{x}}^{\boldsymbol{y}} \psi = \sum_{\boldsymbol{\epsilon} \in \{0,1\}^n} (-1)^{S(\boldsymbol{\epsilon})} \psi(\epsilon_1 x_1 + (1-\epsilon_1) y_1, \ldots, \epsilon_n x_n + (1-\epsilon_n) y_n) \geq 0,$$

where $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_n)$ and $S(\boldsymbol{\epsilon}) = \sum_{i=1}^n \epsilon_i$. The set of all n-variate functions that are Δ-monotone of order i in its i coordinates when the other $n - i$ coordinates are held fixed for all $i = 1, \ldots, n$ will be denoted by $\mathcal{C}_n$.

If $\psi \in \mathcal{C}_n$ then

$$\boldsymbol{X} \leq_{\text{uo}} \boldsymbol{Y} \Longrightarrow E[\psi(\boldsymbol{X})] \leq E[\psi(\boldsymbol{Y})] \tag{6}$$

[23, p. 310]. The following lemma gives conditions for $-\log f(x_1, x_2)$ to be in $\mathcal{C}_2$.

LEMMA 1: *For a real-valued bivariate function $\psi(x_1, x_2) > 0$, $-\log \psi(x_1, x_2) \in \mathcal{C}_2$ if and only if the following hold:*

(*a*) *$\psi(x_1, x_2)\psi(y_1, y_2) \leq \psi(x_1, y_2)\psi(y_1, x_2)$ for all $x_1 \leq y_1$ and $x_2 \leq y_2$.*

(*b*) *$\psi(x_1, y) \geq \psi(y_1, y)$ for all $x_1 \leq y_1$ and for all y.*

(*c*) *$\psi(x, x_2) \geq \psi(x, y_2)$ for all $x_2 \leq y_2$ and for all x.*

PROOF: The inequality in condition (a) holds if and only if $-\log \psi(x_1, x_2)$ is Δ-monotone of order 2. The inequalities in conditions (b) and (c) hold if and only

https://doi.org/10.1017/S0269964810000185 Published online by Cambridge University Press

if $-\log \psi(x_1, x_2)$ is Δ-monotone of order one in each coordinate when the other coordinate is held fixed. Thus, $-\log \psi(x_1, x_2) \in \mathcal{C}_2$ if and only if conditions (a)–(c) hold. ■

When $\psi(x_1, x_2) = f(x_1, x_2)$ is a bivariate PDF, condition (a) of Lemma 1 implies that $f(x_1, x_2)$ is the reverse regular of order 2 (RR2); for more details about RR2, see [19]. The reverse of the inequality in condition (a) implies that $f(x_1, x_2)$ is total positive of order 2 (TP2). Conditions (b) and (c) of Lemma 1 are equivalent to the case when all conditional PDFs of $f(x_1, x_2)$ are decreasing.

Let $\mathcal{F} = \mathcal{F}(F_1, F_2)$ be the set of all distributions with given marginals F_i, $i = 1, 2$ (bivariate Fréchet class [13]). If $F, G \in \mathcal{F}(F_1, F_2)$ are distributions of (X_1, X_2) and (Y_1, Y_2), then

$$(X_1, X_2) \leq_{\text{uo}} (Y_1, Y_2) \iff E[\log \psi(X_1, X_2)] \leq E[\log \psi(Y_1, Y_2)] \tag{7}$$

for all $\psi : \Re^2 \to \Re$ such that

$$\frac{\partial^2}{\partial x_1 \partial x_2} \log \psi(x_1, x_2) \geq 0 \tag{8}$$

(see [23, p. 335]). If the direction of the inequality in (8) is reversed ($\leq$), then $\psi(x_1, x_2)$ satisfies condition (a) of Lemma 1 (i.e., ψ is RR2).

2.2. Information Criteria for Model Derivation

We use three well-known information criteria and their dynamic versions for derivation of probability models based on partial information. Ebrahimi, Soofi, and Soyer [8] gave an overview of the foundation of these information criteria, their interpretations, their relationships with other information measures, and an extensive list of references.

Information provided by a probability distribution F for the prediction of random outcomes of $\boldsymbol{X}$ is measured relative to another distribution F_0 referred to as the *reference distribution*, which is a dominating measure for F. This requires that F be absolutely continuous with respect to F_0, denoted as $F \ll F_0$. The Kullback–Leibler information of F with PDF f relative to F_0 with PDF f_0 is defined by

$$K(F : F_0) = \int_{\Re^n} f(\boldsymbol{x}) \log \frac{f(\boldsymbol{x})}{f_0(\boldsymbol{x})}\, d\boldsymbol{x} \geq 0, \tag{9}$$

provided that $F \ll F_0$ and the integral is finite. The equality in (9) holds if and only if $f(\boldsymbol{x}) = f_0(\boldsymbol{x})$ almost everywhere.

The Shannon entropy of an absolutely continuous distribution is defined by

$$H(F) = -\int_{\Re^n} f(\boldsymbol{x}) \log f(\boldsymbol{x})\, d\boldsymbol{x}, \tag{10}$$

provided that the integral is finite. The entropy can be positive or negative. It measures uniformity (lack of concentration) of the distribution and quantifies uncertainty associated with F.

https://doi.org/10.1017/S0269964810000185 Published online by Cambridge University Press

The mutual information among $X_i,\ i = 1, \ldots, n$, is defined by

$$M(F) = K(F : F_1 \cdots F_n) = \sum_{i=1}^{n} H(F_i) - H(F) \geq 0, \tag{11}$$

where $F_0(\boldsymbol{x}) = F_1(x_1) \cdots F_n(x_n)$ in (9) is the product of the marginal distributions $F_i(x_i)$. Clearly, $M(F) = 0$ if and only if $X_i,\ i = 1, \ldots, n$, are independent. The absolute continuity of $F(\boldsymbol{x})$ with respect to $F_1(x_1) \cdots F_n(x_n)$ is necessary. Otherwise, (11) can produce a negative result; see [6] for the Marshall–Olkin bivariate exponential example.

Given F_0, $K(F : F_0)$ is convex in f and the entropy is concave in f. Therefore, $K(F : F_0)$ [$H(F)$] can be minimized [maximized] with respect to f in a class of distributions denoted as $\Omega = \{F\}$. Some specific examples of Ω will be given after the definitions of the MDIC and MEIC and their interpretations.

DEFINITION 3:

(*a*) *The minimum discrimination information* (*MDI*) *model in* Ω *reference to* F_0 *is* $F^* = \arg\min_{F \in \Omega} K(F : F_0)$.

(*b*) *The maximum entropy (ME) model in* Ω *is* $F^* = \arg\max_{F \in \Omega} H(F)$.

The MEIC, proposed by Jaynes [12] , develops probability models that are most noncommittal to information other than that explicitly taken into account. The MEIC extends Laplace's principle of insufficient reason. The MDIC, also known as the minimum cross-entropy principle, proposed by Kullback [18], generalizes the MEIC by developing noncommittal models with reference to any given F_0 instead of the uniform reference. Shore and Johnson [24] provided an axiomatic foundation for the MDIC and MEIC.

When F_0 in (9) is uniform (proper or improper), the MDIC and MEIC are equivalent. The MEIC provides the model that is closest to the uniform distribution (lack of concentration) and satisfies the conditions specified in Ω. The MDIC provides the model that is closest to any reference distribution F_0 and satisfies the conditions specified in Ω. The MDIC and MEIC are widely used in statistics and many other fields for characterizing distributions, estimation, and testing, among other purposes.

Two well-known examples of Ω are the moment classes

$$\Omega_\Sigma = \{F : E(X_i^2) = \sigma_{ii}, E(X_1 X_2) = \sigma_{12},\ i = 1, 2\}, \tag{12}$$

$$\Omega_\theta = \{F : E[\log(1 + X_1 + X_2)] = \theta\}. \tag{13}$$

Both classes generalize to the multivariate case. They also can be expanded by replacing the equalities in the sets with the inequality ($\leq$); see [18,24]. The class in (12) characterizes the bivariate normal distribution. In this case, the ME model is nested because the class can be represented as $\Omega_\Sigma = \Omega_M \cap \Omega_D$, which respectively includes marginal moments $E(X_i^2) = \sigma_{ii}$, where the marginal distributions are ME models and

https://doi.org/10.1017/S0269964810000185 Published online by Cambridge University Press

the cross-product moment $E(X_1X_2) = \sigma_{12}$; see [7] for details and the latest results on characterization of probability models when $\Omega = \Omega_{\boldsymbol{\theta}}$ is a moment class of distributions. If Ω_D is null, by (11), the joint entropy attains its maximum for the independent model $F^*(\boldsymbol{x}) = F_1^*(x_1)F_2^*(x_2)$. However, Ω_M is necessary for the existence of the ME model. The class in (13) is not separable and Ω_θ characterizes the bivariate Pareto distribution. We mention in passing that, more generally, the marginal fitting where r-dimensional marginal distributions are given, $r < n$, is an important class of information-theoretic problems in statistics; see Gokhale and Kullback [10].

In the multivariate case, the MDIC and MEIC are also equivalent when $F_0 = F_1 \cdots F_n$. Let $\Omega_M = \mathcal{F} = \mathcal{F}(F_1, \cdots, F_n)$ be the set of all distributions with given marginals $F_i,\ i = 1, \ldots, n$ (Fréchet classes [13]), the Ω_D include joint (dependence) constraints, and

$$\Omega = \mathcal{F} \cap \Omega_D. \tag{14}$$

DEFINITION 4: *The minimum mutual information* (*MMI*) *model in* Ω *defined in* (14) *is*

$$F^* = \arg\min_{F \in \Omega} M(F) = \arg\max_{F \in \Omega} H(F). \tag{15}$$

At age $\boldsymbol{t} = (t_1, \ldots, t_n),\ t_j \geq 0$, of the components of a system, the joint residual distribution $F(\boldsymbol{x};\boldsymbol{t}) = F(\boldsymbol{x})/\bar{F}(\boldsymbol{t})$ and the residual PDF will be denoted by $f(\boldsymbol{x};\boldsymbol{t}) = f(\boldsymbol{x})/\bar{F}(\boldsymbol{t})$ for $\boldsymbol{x} > \boldsymbol{t}$. The information in $F(\boldsymbol{x};\boldsymbol{t})$ relative to the reference residual distribution $F_0(\boldsymbol{x};\boldsymbol{t})$ is given by $K(F : F_0;\boldsymbol{t}) \geq 0$. The joint residual entropy $H(F;\boldsymbol{t})$ is found by using $f(\boldsymbol{x};\boldsymbol{t})$ in (10). The residual mutual information $M(F;\boldsymbol{t})$ is given by (11), where $H(F_i;\boldsymbol{t})$ is given by (10) with the PDF of the marginal of the multivariate residual distribution

$$f_i(x_i;\boldsymbol{t}) = \int_{\boldsymbol{t}_{[i]}}^{\infty} f(\boldsymbol{x};\boldsymbol{t})\, d\boldsymbol{x}_{[i]}, \qquad \boldsymbol{x} > \boldsymbol{t}; \tag{16}$$

here, $\boldsymbol{x}_{[i]}$ and $\boldsymbol{t}_{[i]}$ denote the $\boldsymbol{x}$ and $\boldsymbol{t}$ without the x_i and t_i, respectively. Considering $\mathcal{S}_t = \{\boldsymbol{x} : \boldsymbol{x} > \boldsymbol{t}\}$ as a multivariate index set, $K(F : F_0;\boldsymbol{t}), H(F;\boldsymbol{t})$, and $M(F;\boldsymbol{t})$ do evolve with $\boldsymbol{t}$ and are referred to as *dynamic* information measures.

The minimum dynamic discrimination information (MDDI) and maximum dynamic entropy (MDE) models are optimal in terms of dynamic information measures $K(F : F_0;\boldsymbol{t})$ and $H(F;\boldsymbol{t})$, respectively. The MDDI (MDE) model $F^* \in \Omega$ is the one that its residual distribution $F^*(\boldsymbol{x};\boldsymbol{t})$ retains the MDI (ME) property among the residual distributions $F(\boldsymbol{x};\boldsymbol{t})$ of all $F \in \Omega$ for all $t \in \Re_+^n$. An optimal model that minimizes $M(F;\boldsymbol{t})$ can be defined similarly. However, the reference distribution is $F(\boldsymbol{x};\boldsymbol{t}) = \prod_{i=1}^n F_i(x_i;\boldsymbol{t})$, where the PDF of $F_i(x_i;\boldsymbol{t})$ as in (16), which requires the joint distribution and hence is not useful. The MDDI and MDE models are found by orderings of the dynamic information functions. Thus, the optimal models are boundary solutions.

Asadi et al. [1,3] considered $\Omega = \{F\}$ for an univariate F subject to the inequality or differential inequality constraints for the hazard rate or inequality constraints for the mean residual lifetime. They characterized numerous univariate distributions as

https://doi.org/10.1017/S0269964810000185 Published online by Cambridge University Press

MDDI and MDE models. We develop models using (2) and (3), which define classes of distributions by inequality constraints on survival and gradient hazard functions.

3. MODEL DERIVATION AND CHARACTERIZATIONS

3.1. Survival Function Inequality Constraint

For two distributions F and G, the upper orthant ordering plays the fundamental role in discrimination information ordering $K(G : F_0) \leq K(F : F_0)$ and entropy ordering $H(F) \leq H(G)$. The following result gives the conditions for the MDI model.

THEOREM 1: *Let Ω_G in (2) be a compact set of distributions with $F \ll F_0$.*

(*a*) *If $-\log(g(\boldsymbol{x})/f_0(\boldsymbol{x})) \in \mathcal{C}_n$, then $G = F^*$ is the MDI model in Ω_G with $\bar{F}(\boldsymbol{x}) \leq \bar{G}(\boldsymbol{x})$.*

(*b*) *If $\log(g(\boldsymbol{x})/f_0(\boldsymbol{x})) \in \mathcal{C}_n$, then $G = F^*$ is the MDI model in Ω_G with $\bar{F}(\boldsymbol{x}) \geq \bar{G}(\boldsymbol{x})$.*

PROOF:

(*a*) Let $\boldsymbol{X}$ and $\boldsymbol{Y}$ denote random variables with distributions F and G, so $\boldsymbol{X} \leq_{\text{uo}} \boldsymbol{Y}$. Since $-\log(g(\boldsymbol{x})/f_0(\boldsymbol{x})) \in \mathcal{C}_n$, by (6),

$$E_F\left[\log \frac{g(\boldsymbol{X})}{f_0(\boldsymbol{X})}\right] \geq E_G\left[\log \frac{g(\boldsymbol{X})}{f_0(\boldsymbol{X})}\right] = K(G : F_0). \tag{17}$$

For any $F \in \Omega_G$,

$$\begin{aligned} K(F : F_0) &= \int_{\Re^n} f(\boldsymbol{x}) \log \frac{f(\boldsymbol{x})}{f_0(\boldsymbol{x})}\, d\boldsymbol{x} \\ &= \int_{\Re^n} f(\boldsymbol{x}) \log \frac{g(\boldsymbol{x})}{f_0(\boldsymbol{x})}\, d\boldsymbol{x} + \int_{\Re^n} f(\boldsymbol{x}) \log \frac{f(\boldsymbol{x})}{g(\boldsymbol{x})}\, d\boldsymbol{x} \\ &\geq E_F\left[\log \frac{g(\boldsymbol{X})}{f_0(\boldsymbol{X})}\right], \end{aligned} \tag{18}$$

where the inequality follows from the fact that the second integral in (18) is $K(F : G) \geq 0$. Then (17) gives the result.

(*b*) Part (b) can be proved similarly. ■

The next result immediately follows from Theorem 1.

COROLLARY 1: *Let Ω_G in (2) be a compact set of absolutely continuous distributions.*

(*a*) *If $-\log g(\boldsymbol{x}) \in \mathcal{C}_n$, then $G = F^*$ is the ME model in Ω_G with $\bar{F}(\boldsymbol{x}) \leq \bar{G}(\boldsymbol{x})$.*

(*b*) *If $\log g(\boldsymbol{x}) \in \mathcal{C}_n$, then $G = F^*$ is the ME model in Ω_G with $\bar{F}(\boldsymbol{x}) \geq \bar{G}(\boldsymbol{x})$.*

https://doi.org/10.1017/S0269964810000185 Published online by Cambridge University Press

PROOF: The proof follows from Theorem 1 when F_0 is uniform on its support. ■

If Ω_G in (2) is a compact set of distributions with given marginals $F_i,\ i = 1, \ldots, n$ and $F \ll F_1 \cdots F_n$, then, by (15), the ME model $G = F^*$ in Corollary 1 is also the MMI model in $\Omega = \mathcal{F}(F_1, \ldots, F_n) \cap \Omega_G$, with the corresponding direction of the inequality in Ω_G. The following result gives the MMI model for the bivariate case, based on (8).

COROLLARY 2: *Let $\Omega = \mathcal{F}(F_1, F_2) \cap \Omega_G$ such that Ω_G in (2)is a compact set of bivariate distributions with given marginals F_1 and F_2, and $F \ll F_1F_2$.*

(*a*) *If $-\log g(x_1, x_2)$ satisfies* (8)*, then $G = F^*$ is the MMI model in Ω with Ω_G defined by $\bar{F}(x_1, x_2) \le \bar{G}(x_1, x_2)$.*

(*b*) *If $\log g(x_1, x_2)$ satisfies* (8)*, then $G = F^*$ is the MMI model in Ω with Ω_G defined by $\bar{F}(x_1, x_2) \ge \bar{G}(x_1, x_2)$.*

PROOF: For $F_0 = F_1F_2$, (7) gives the inequality in (17). ■

Next, we show applications of the above results by characterizing two bivariate models.

Example 1 (Gumbel Model): The bivariate Gumbel survival function is

$$\bar{G}(x_1, x_2) = e^{-\theta_1 x_1 - \theta_2 x_2 - \delta\theta_1\theta_2 x_1 x_2}, \qquad x_1, x_2 > 0,\ \ 0 \le \delta \le 1,\ \ \theta_i > 0,\ i = 1, 2. \quad \textbf{(19)}$$

(a) It can be shown that $g(x_1, x_2)$ satisfies condition (a) of Lemma 1 for $0 \le \delta \le 1$ and conditions (b) and (c) when $0 \le \delta \le .5$. Thus, $-\log g(x_1, x_2) \in \mathcal{C}_2$. By part (a) of Corollary 1, the Gumbel distribution with parameter restriction $0 \le \delta \le 0.5$ is the ME model $\bar{G}(x_1, x_2) = \bar{F}^*(x_1, x_2)$ in Ω_G defined as in (2) with $\bar{F}(x_1, x_2) \le \bar{G}(x_1, x_2)$ for the set of all absolutely continuous bivariate distributions, inclusive of (19). In addition, Corollary 2 implies that the entropy of the Gumbel model (19) is ordered by the dependence parameter δ.

(b) It is easy to check that the Gumbel PDF $g(x_1, x_2)$ satisfies the condition of (8) with direction ≤ 0. By Corollary 2, the Gumbel distribution is the MMI model $\bar{G}(x_1, x_2) = \bar{F}^*(x_1, x_2)$ in $\Omega = \mathcal{F}(F_1, F_2) \cap \Omega_G$, where $F_i(x_i) = 1 - e^{-\theta_i x_i},\ i = 1, 2$, and Ω_G defined as in (2) with $\bar{F}(x_1, x_2) \le \bar{G}(x_1, x_2)$ for the set of all bivariate distributions, $F \ll F_1F_2$, inclusive of (19).

(c) For a more general MDI application, consider the reference distribution

$$\bar{F}_0(x_1, x_2) = e^{-\lambda_1 x_1 - \lambda_2 x_2},\ \ x_i \ge 0,\ \ \lambda_i > 0,\ i = 1, 2. \qquad \textbf{(20)}$$

https://doi.org/10.1017/S0269964810000185 Published online by Cambridge University Press

It can be shown that $g(x_1, x_2)/f_0(x_1, x_2)$ satisfies condition (a) of Lemma 1 for $0 \leq \delta < 1$ and conditions (b) and (c) for parameters restricted as

$$0 \leq \delta \leq \frac{1}{2} \min \left\{ 1 - \frac{\lambda_1}{\theta_1}, 1 - \frac{\lambda_2}{\theta_2} \right\}, \qquad \theta_i > \lambda_i,\ i = 1, 2. \tag{21}$$

Thus, $-\log(g(x_1, x_2)/f_0(x_1, x_2)) \in \mathcal{C}_2$. By part (a) of Theorem 1, the Gumbel distribution with survival function (19), and the narrower parameter space (21), is the MDI model $\bar{G}(x_1, x_2) = \bar{F}^*(x_1, x_2)$ in Ω_G defined as in (2) with $\bar{F}(x_1, x_2) \leq \bar{G}(x_1, x_2)$ for the set of all bivariate distributions $F \ll F_0$, inclusive of (19); that is, among all distributions $F \ll F_0$ absolutely continuous with respect to the independent exponential (20) and whose survival functions are dominated by (dominate) (19) with the parameter space (21), the Gumbel model (19) is the closest one, in the Kullback–Leibler information sense, to the independent exponential (20).

Example 2 [Farlie–Gumbel–Morgenstern (FGM)]: The FGM bivariate family has the PDF in the form

$$\bar{G}_\beta(x_1, x_2) = \bar{G}_1(x_1)\bar{G}_2(x_2)[1 + \beta G_1(x_1)G_2(x_2)], \qquad -1 \leq \beta \leq 1. \tag{22}$$

The PDF of this distribution does not satisfy the conditions of Lemma 1. However, it can be shown that (8) [(8) with the equality reversed] holds when $\beta \geq [\leq]\ 0$. By Corollary 2, the FGM is the MMI model in the set of all bivariate distributions with marginal distributions $F_1 = G_1$ and $F_2 = G_2$, absolutely continuous with respect to $F_0 = F_1F_2$, and whose survival functions are dominated by (dominate) the FGM survival function $\bar{G}_\beta = \bar{F}^*$ with $\beta \geq [\leq] 0$. Furthermore, using the mutual information formula for FGM from [9], the minimum dependence in the classes of distribution where FGM is the MMI model is given by

$$M(G_\beta) = \sum_{n=1}^{\infty} \frac{\beta^{2n}}{2n(2n-1)(2n+1)^2} \leq \sum_{n=1}^{\infty} \frac{1}{2n(2n-1)(2n+1)^2}.$$

The series converges quickly and computation indicates that $M(G_{|\beta|=1}) \approx 0.06$; see [9] for details.

3.2. Hazard Gradient Inequality Constraint

In this subsection, we show that for dynamic discrimination information ordering $K(G : F_0; \boldsymbol{t}) \leq K(F : F_0; \boldsymbol{t})$ and dynamic entropy ordering $H(F; \boldsymbol{t}) \leq H(G; \boldsymbol{t})$, the hazard gradient ordering plays the same fundamental role as that of the upper orthant ordering in the previous subsection. The following result gives the conditions for the MDDI model.

https://doi.org/10.1017/S0269964810000185 Published online by Cambridge University Press

THEOREM 2: *Let Ω_Q in* (3) *be a compact set of distributions $F \ll F_0$.*

(*a*) *If $-\log(g(\boldsymbol{x})/f_0(\boldsymbol{x})) \in \mathcal{C}_n$, then $G = F^*$ is the MDDI model in Ω_Q with $R(\boldsymbol{x}) \geq Q(\boldsymbol{x})$.*

(*b*) *If $\log(g(\boldsymbol{x})/f_0(\boldsymbol{x})) \in \mathcal{C}_n$, then $G = F^*$ is the MDDI model in Ω_Q with $R(\boldsymbol{x}) \leq Q(\boldsymbol{x})$.*

PROOF: Let $\boldsymbol{X}$ and $\boldsymbol{Y}$ denote random variables with distributions G and F and let $\boldsymbol{X_t}|\boldsymbol{X} > \boldsymbol{t}$ and $\boldsymbol{Y_t}|\boldsymbol{Y} > \boldsymbol{t}$ denote their residual lifetimes. Since $R(\boldsymbol{t}) \geq Q(\boldsymbol{t})$ for all $\boldsymbol{t} \in \Re_+^n$, using (5) we have $\boldsymbol{X_t} \geq_{\text{uo}} \boldsymbol{Y_t}$. Using this, the fact that $-\log(g(\boldsymbol{x})/f_0(\boldsymbol{x})) \in \mathcal{C}_n$, and following the same steps as in the proof of Theorem 1 give the result. ■

The following result gives the conditions for the MDE model.

COROLLARY 3: *Let Ω_Q in* (3) *be a compact set of absolutely continuous distributions.*

(*a*) *If $-\log g(\boldsymbol{x}) \in \mathcal{C}_n$, then $G = F^*$ is the MDE model in Ω_Q with $R(\boldsymbol{x}) \geq Q(\boldsymbol{x})$.*

(*b*) *If $\log g(\boldsymbol{x}) \in \mathcal{C}_n$, then $G = F^*$ is the MDE model in Ω_Q with $R(\boldsymbol{x}) \leq Q(\boldsymbol{x})$.*

PROOF: The proof follows from Theorem 2 when F_0 is uniform on its support. ■

The next result gives conditions for the MDE properties of the marginals of the bivariate residual distribution $F_i^*(x_i;\boldsymbol{t}),\ i = 1, 2$.

COROLLARY 4: *Let Ω_Q in* (3) *be a compact set of absolutely continuous distributions with $G = F^* \in \Omega_Q$.*

(*a*) *If the PDF of the marginal of the bivariate residual distribution $g_i(x_i;\boldsymbol{t})$ is decreasing in x_i in the support, then $G_i(x_i;\boldsymbol{t}) = F_i^*(x_i;\boldsymbol{t})$ is the MDE distribution in Ω_{F_i} with $R(\boldsymbol{x}) \geq Q(\boldsymbol{x})$.*

(*b*) *If the PDF of the marginal of the bivariate residual distribution $g_i(x_i;\boldsymbol{t})$ is increasing in x_i in the support, then $G_i(x_i;\boldsymbol{t}) = F_i^*(x_i;\boldsymbol{t})$ is the MDE distribution in Ω_{F_i} with $R(\boldsymbol{x}) \leq Q(\boldsymbol{x})$.*

PROOF:

(a) Let $\boldsymbol{X}$ and $\boldsymbol{Y}$ denote the bivariate random variables with distributions G and F and let $\boldsymbol{X_t} = \boldsymbol{X}|\boldsymbol{X} > \boldsymbol{t}$ and $\boldsymbol{Y_t} = \boldsymbol{Y}|\boldsymbol{Y} > \boldsymbol{t}$ denote the residual lifetimes. Since $Q^{(i)}(\boldsymbol{x}) \leq R^{(i)}(\boldsymbol{x})$, it can be shown that $X_i|\boldsymbol{x} > \boldsymbol{t} \geq_{\text{st}} Y_i|\boldsymbol{y} > \boldsymbol{t}$. Since $g_i(x_i;\boldsymbol{t})$ is

https://doi.org/10.1017/S0269964810000185 Published online by Cambridge University Press

decreasing in x_i in the support, we have

$$E_{g_i(x_i,\boldsymbol{t})}[\log g_i(X_i;\boldsymbol{t})] \le E_{f_i(x_i,\boldsymbol{t})}[\log g_i(Y_i;\boldsymbol{t})]. \tag{23}$$

Now, consider the discrimination information between the two residual lifetime distributions:

$$\begin{aligned} K(F_i : G_i;\boldsymbol{t}) &= \int_{t_i}^{\infty} f_i(v;\boldsymbol{t}) \log \frac{f_i(v;\boldsymbol{t})}{g_i(v;\boldsymbol{t})}\, dv \\ &= -H(F_i;\boldsymbol{t}) - E_{f_i(x_i,\boldsymbol{t})}[\log g_i(Y_i;\boldsymbol{t})] \ge 0. \end{aligned}$$

Using this and (23) gives $H(F_i;\boldsymbol{t}) \le H(G_i;\boldsymbol{t})$.

(b) The proof of (b) is similar.

■

Note that Corollary 3 implies Corollary 4. Therefore, marginals of all the MDE bivariate residual models obtained above are also MDE subject to the respective constraints.

Next, we show applications of the above results through characterizing and developing two bivariate models.

Example 3 (Gumbel Model): Let Ω_Q be defined by hazard gradients whose components satisfy the following linear inequalities:

$$R^{(i)}(x_1, x_2) \ge a_i + bx_j, \qquad j \ne i = 1, 2,\ a_i > 0,\ 0 \le b \le 0.5a_1a_2. \tag{24}$$

Figure 1a depicts Ω_Q in the (x_1, x_2)-plane as the solutions of the linear inequalities (24). The region depicts an envelope with angle $\alpha = \arctan[(1 - b^2)/2b]$. The

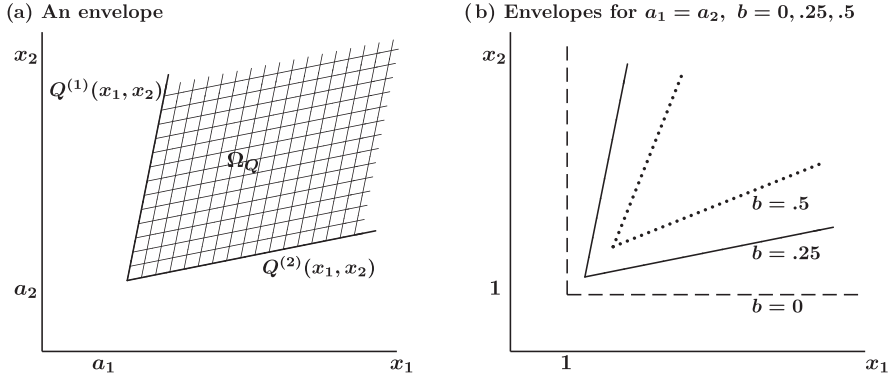


FIGURE 1. Envelopes of hazard gradients with Gumbel borders $Q^{(i)}(x_1, x_2) = a_i + bx_j$, $i \ne j$, $a_1 = a_2 = 1$, $b = 0, 0.25, 0.5$.

https://doi.org/10.1017/S0269964810000185 Published online by Cambridge University Press

hazard gradient of the information optimal model (19) defines the border of the envelope.

(a) For $b = \delta\theta_1\theta_2$, the right-hand side of (24) is the ith element of the hazard gradient of the Gumbel model (19) with parameters determined by

$$0 \le \delta = \frac{b}{a_1 a_2} \le 0.5, \qquad \theta_i = a_i,\ i = 1, 2. \tag{25}$$

Since $g(x_1, x_2)$ with parameter restrictions (25) satisfies conditions of Lemma 1, by Corollary 3 the Gumbel distribution with parameter restriction $0 \le \delta \le 0.5$ is the MDE model in Ω_Q defined as in (3) for all absolutely continuous bivariate distributions, inclusive of (19). Additionally, similarly by part (a) of Theorem 2, the Gumbel model with parameter restrictions (21) is the MDDI model relative to the independent exponential reference distribution (20). Figure 1b compares the Ω_Q envelopes for $\delta = b = 0, 0.25, 0.5$ and $a_1 = a_2 = 1$. For $\delta = 0$, the model is independent exponential and the components of its hazard gradient are orthogonal, so the envelope is rectangular and the largest. For $\delta = 0.5$, the envelope is smallest with angle $\alpha \approx 36.8°$.

(b) Alternatively, we can define Ω_Q in terms of the hazard gradients whose components satisfy the following partial differential inequalities:

$$\frac{\partial R^{(i)}(x_1, x_2)}{\partial x_i} \ge 0, \qquad a_i > 0,\ \ 0 \le b \le 0.5 a_1 a_2,\ \ i = 1, 2,$$
$$R^{(1)}(0, x_2) = a_1 + b x_2, \qquad R^{(2)}(x_1, 0) = a_2 + b x_1.$$

Then MDDI and MDE give the same model as above. This is seen by integrating both sides of the differential inequality and applying the initial conditions, which gives Inequality (24) as the solution.

It is noteworthy to mention in passing that we can draw an analogy between the information characterizations of the bivariate Gumbel and the bivariate normal distributions. The equalities in (24) play the same role as the linear regression functions of the bivariate normal distribution. The geometry of the bivariate Gumbel information set Ω_Q shown in Figure 1 is analogous to the geometry of the bivariate normal information set Ω_Σ in (24) expanded in terms of the inequalities and represented in terms of the second moments and regression functions. The Gumbel dependence parameter δ determines the angle between the components of the hazard gradient exactly in the same way as the standard normal correlation parameter ρ does for the regression functions $\alpha = \arctan[(1 - \rho^2)/2\rho]$. However, there is no restriction such as (25) on ρ. Another notable dissimilarity is that unlike Ω_Σ, Ω_Q is not separable into the marginal and joint information sets, so in this aspect it is similar to Ω_θ defined by (13).

Next, example shows application of Corollary 3 to the Lomax distribution.

https://doi.org/10.1017/S0269964810000185 Published online by Cambridge University Press

Example 4 (Lomax Model): Let Ω_Q be defined by hazard gradients whose components satisfy the following partial differential inequalities:

$$\frac{\partial R^{(i)}(x_1,x_2)}{\partial x_i} \geq -a[R^{(i)}(x_1,x_2)]^2, \qquad a > 0.5,\ i = 1,2$$
$$R^{(1)}(0,x_2) = \frac{b_1(1+b_2x_2)}{1+ab_2x_2}, \qquad R^{(2)}(x_1,0) = \frac{b_2(1+b_1x_1)}{1+ab_1x_1}, \quad b_i > 0,\ i = 1,2. \tag{26}$$

This differential inequality can be represented as

$$\frac{\partial R^{(i)}(x_1,x_2)/\partial x_i}{R_i^2(x_1,x_2)} \geq -a, \qquad a > 0,\ i = 1,2.$$

Integrating both sides and applying the initial conditions give the following solutions:

$$R^{(i)}(x_1,x_2) \geq \frac{b_i(1+b_jx_j)}{1+a(b_1x_1+b_2x_2+b_1b_2x_1x_2)}, \qquad j \neq i = 1,2.$$

For $\delta = a^{-1} > 0$ and $\theta_i = ab_i > 0,\ i = 1,2$, the right-hand side is the ith element of the hazard gradient of the bivariate Lomax distribution with the survival function

$$\bar{G}(x_1,x_2) = (1+\theta_1x_1+\theta_2x_2+\delta\theta_1\theta_2x_1x_2)^{-\delta},$$
$$x_1,\ x_2, > 0, \quad 1 \leq \delta = \frac{1}{a} \leq 2,\ \ \theta_i = ab_i,\ i = 1,2. \tag{27}$$

It can be shown that $g(x_1,x_2)$ satisfies condition (a) of Lemma 1 for $\delta \geq 1$ and conditions (b) and (c) when $0 < \delta \leq 2$. Thus, $-\log g(x_1,x_2) \in \mathcal{C}_2$. By Corollary 3, the bivariate Lomax distribution (27), the MDE model $\bar{G}(x_1,x_2) = \bar{F}^*(x_1,x_2)$ in Ω_Q defined as in (3) with $\bar{F}(x_1,x_2) \leq \bar{G}(x_1,x_2)$ for the set of all absolutely continuous bivariate distributions, inclusive of (27).

The next result shows that the set of distributions that satisfy the given partial information might not include distributions with positive dependence.

THEOREM 3: *Let* $\Omega_Q = \{F : R(\boldsymbol{x}) \geq Q(\boldsymbol{x})\}$ *and* $Q(x_1,x_2) = (Q^{(1)}(x_1,x_2), Q^{(2)}(x_1,x_2))$. *If there is a distribution* $F \in \Omega_Q$ *such that its PDF is TP2, then there exist functions* $R_i(x_i)$ *such that*

$$\lim_{x_j\to\infty} Q^{(i)}(x_1,x_2) \leq R_i(x_i), \qquad \forall x_i \in \Re_+,\ \ i,j = 1,2, i \neq j, \tag{28}$$

where $R_i(x_i),\ i = 1,2$, *are the marginal hazard rates.*

https://doi.org/10.1017/S0269964810000185 Published online by Cambridge University Press

PROOF: We will prove inequality for $i = 1$ and $j = 2$. Let $F \in \Omega_Q$ with hazard gradient $R(x_1, x_2) = (R^{(1)}(x_1, x_2), R^{(2)}(x_1, x_2))$ and PDF f, which is TP2. Since $F \in \Omega_Q$, we have

$$R^{(i)}(x_1, x_2) \geq Q^{(i)}(x_1, x_2), \qquad \forall (x_1, x_2) \in \Re_+^2. \tag{29}$$

Since f is TP2, we have

$$R^{(1)}(x_1, x_2) \leq R^{(1)}(x_1, 0) \equiv R_1(x_1), \qquad \forall (x_1, x_2) \in \Re_+^2.$$

Using this and (29), we get $Q^{(1)}(x_1, x_2) \leq R_1(x_1), \forall\ (x_1, x_2) \in \Re_+^2$. Taking the limit of both sides as $x_2 \to \infty$ proves the result. ■

Using Theorem 3, we can conclude that the set of bivariate distributions Ω_Q in Example 3 does not contain any distribution that its survival function is TP2 because in these cases, $\lim_{x_j \to \infty} Q^{(i)}(x_1, x_2) = \infty,\ \ i, j = 1, 2, i \neq j$.

4. TRANSFORMATION

The Kullback–Leibler function is invariant under all one-to-one transformations of $\boldsymbol{X}$ and the mutual information is invariant under all coordinatewise one-to-one transformations. However, the entropy is not invariant under all one-to-one transformations of $\boldsymbol{X}$ or X_i.

Let F be the distribution of (X_1, X_2) with PDF $f(x_1, x_2)$ and $Y_i = \phi_i(X_i),\ i = 1, 2$, where $\phi_i : \Re_+ \to \Re_+,\ i = 1, 2$. From [23, pp. 292 and 310], when $\phi_i,\ i = 1, 2$, are increasing functions, we have the following transformations of (2) and (3):

$$\Omega_G \to \Omega_{G^\phi} = \left\{F^\phi : \bar{F}^\phi(y_1, y_2) \leq (\geq) \bar{G}^\phi(y_1, y_2)\right\}, \tag{30}$$

$$\Omega_Q \to \Omega_{Q^\phi} = \left\{F^\phi : R^\phi(y_1, y_2) \geq (\leq) Q^\phi(y_1, y_2)\right\}, \tag{31}$$

where the superscript ϕ stands for transformed $\bar{F}, \bar{G}, R$, and Q.

Membership in $\mathcal{C}_n$ is not invariant under all transformations. The following lemma gives the ingredients for developing bivariate models and the maximum entropy characterizations of bivariate distributions under transformations.

LEMMA 2: *Let F be the distribution of (X_1, X_2) with PDF $f(x_1, x_2)$ and $Y_i = \phi_i(X_i),\ i = 1, 2$, where $\phi_i : \Re_+ \to \Re_+,\ i = 1, 2$.*

(*a*) *If $\phi_i, i = 1, 2$, are increasing and $-\log(f(x_1, x_2)/f_0(x_1, x_2)) \in \mathcal{C}_2$, then $-\log(f^\phi(x_1, x_2)/f_0^\phi(x_1, x_2)) \in \mathcal{C}_2$.*

(*b*) *If $\phi_i, i = 1, 2$, are increasing and convex and $-\log f(x_1, x_2) \in \mathcal{C}_2$, then $-\log f^\phi(y_1, y_2) \in \mathcal{C}_2$.*

https://doi.org/10.1017/S0269964810000185 Published online by Cambridge University Press

PROOF: The Jacobian of transformation is

$$J(y_1, y_2) = |J_1(y_1) \times J_2(y_2)|, \qquad J_i(y_i) = \frac{\partial}{\partial y_i}\phi_i^{-1}(y_i),\ i = 1, 2. \tag{32}$$

Let $\mathcal{X} = \mathcal{X}_1 \times \mathcal{X}_2$ and $\mathcal{Y} = \mathcal{Y}_1 \times \mathcal{Y}_2$ be the supports of the F and F^{ϕ}, respectively. For all $(y_1, y_2), (y_1^*, y_2^*) \in \mathcal{Y}$, there are two vectors $(x_1, x_2), (x_1^*, x_2^*) \in \mathcal{X}$ such that $y_i = \phi_i(x_i), i = 1, 2$, and $y_i^* = \phi_i(x_i^*), i = 1, 2$. Since $\phi_i,\ i = 1, 2$, are increasing, $y_1 \le y_1^*$ and $y_2 \le y_2^*$ implies that $x_1 \le x_1^*$ and $x_2 \le x_2^*$.

(a) For $\psi(x_1, x_2) = f(x_1, x_2)/f_0(x_1, x_2)$, (32) gives $\psi^{\phi}(y_1, y_2) = f(\phi_1^{-1}(y_1), \phi_2^{-1}(y_2))/f_0(\phi_1^{-1}(y_1), \phi_2^{-1}(y_2))$. Since $-\log \psi(x_1, x_2) \in \mathcal{C}_2$, we obtain the result by Lemma 1.

(b) For $\psi(x_1, x_2) = f(x_1, x_2)$, (32) gives $\psi^{\phi}(y_1, y_2) = f^{\phi}(y_1, y_2) = f(\phi_1^{-1}(y_1), \phi_2^{-1}(y_2))J(y_1, y_2)$. Since

$$\psi(x_1, x_2)\psi(x_1^*, x_2^*) \le \psi(x_1, x_2^*)\psi(x_1^*, x_2),$$

we obtain

$$f^{\phi}(y_1, y_2)f^{\phi}(y_1^*, y_2^*) \le f^{\phi}(y_1, y_2^*)f^{\phi}(y_1^*, y_2), \qquad \forall (y_1, y_2) \le (y_1^*, y_2^*).$$

Also by Lemma 1, $f(x_1, x_2) \ge f(x_1^*, x_2)$. Since ϕ_1 is convex, we have $f^{\phi}(y_1, y_2) \ge f^{\phi}(y_1^*, y_2)$. The condition $f^{\phi}(y_1, y_2) \ge f^{\phi}(y_1, y_2^*)$ and Lemma 1 gives the result. ■

The following theorem encapsulates results of Lemma 2, Theorems 1 and 2, and Corollaries 1 and 3 for transformations of the MDI, MDDI, ME, and MDE models.

THEOREM 4: *Let Ω_G and Ω_Q as defined in* (2) *and* (3) *be compact sets of absolutely continuous bivariate distributions. Let $Y_i = \phi_i(X_i),\ i = 1, 2$, where $\phi_i : \Re_+ \to \Re_+,\ i = 1, 2$.*

(*a*) *If $\phi_i,\ i = 1, 2$, are increasing and $-\log g(x_1, x_2)/f_0(x_1, x_2) \in \mathcal{C}_2$, then the distribution induced by $\phi_i,\ i = 1, 2$, on the MDI [MDDI] model $G(x_1, x_2) = F^*(x_1, x_2)$ of Ω_G (Ω_Q) defined in* (2) [(3)] *is the MDI [MDDI] model $G^{\phi}(y_1, y_2) = F^{\phi *}(y_1, y_2)$ in the set of all transformed distributions $\Omega_{G^{\phi}}$ ($\Omega_{Q^{\phi}}$) defined in* (30) [(31)].

(*b*) *If $\phi_i,\ i = 1, 2$, are increasing and convex and $-\log g(x_1, x_2) \in \mathcal{C}_2$, then the distribution induced by $\phi_i,\ i = 1, 2$, on the ME [MDE] model $G(x_1, x_2) = F^*(x_1, x_2)$ of Ω_G [Ω_Q] defined in* (2) [(3)] *is the ME [MDE] model $G^{\phi}(y_1, y_2) = F^{\phi *}(y_1, y_2)$ in the set of all transformed distributions $\Omega_{G^{\phi}}$ [$\Omega_{Q^{\phi}}$] defined in* (30) [(31)].

Table 1 gives examples of a few important transformations of exponential variables X_i. The last column shows the information criteria applicable to the transformations listed in Table 1. Applications of Theorem 4 provide assortments of bivariate

https://doi.org/10.1017/S0269964810000185 Published online by Cambridge University Press

https://doi.org/10.1017/S0269964810000185 Published online by Cambridge University Press

TABLE 1. Distributions Related to Exponential Distribution via Transformations

Distribution, Support	Survival Function	Transformation	Criteria Applicable
Pareto Type I, $[\alpha_i, \infty)$	$\bar{G}_i(y_i) = \alpha_i^{\theta_i} y_i^{-\theta_i}$	$y_i = \alpha_i e^{x_i},\ \alpha_i > 0$	All
Pareto Type II, $\Re_+$	$\bar{G}_i(y_i) = (1 + y_i)^{-\theta_i}$	$y_i = e^{x_i} - 1$	All
Pareto Type IV, $\Re_+$, (Type III, $\theta_i = 1$)	$\bar{G}_i(y_i) = \left(1 + y_i^{\alpha_i}\right)^{-\theta_i}$	$y_i = (e^{x_i} - 1)^{1/\alpha_i}$	$\begin{cases}\text{MDI and MDDI } \alpha_i > 0, \\ \text{ME and MDE } \alpha_i \leq 1\end{cases}$
Weibull, $\Re_+$	$\bar{G}_i(y_i) = \exp\left(-\theta_i y_i^{\alpha_i}\right)$	$y_i = x_i^{1/\alpha_i}$	$\begin{cases}\text{MDI and MDDI } \alpha_i > 0, \\ \text{ME and MDE } \alpha_i \leq 1\end{cases}$
Linear failure rate, $\Re_+$	$\bar{G}_i(y_i) = e^{-\theta_i(y_i + \alpha_i y_i^2/2)}$	$y_i = \frac{1}{\alpha_i}\left[(1 + 2\alpha_i x_i)^{1/2} - 1\right]$	MDI and MDDI $\alpha_i > 0$
Extreme value, $\Re_+$	$\bar{G}_i(y_i) = \exp(1 - \theta_i e^{y_i})$	$y_i = \log(1 + x_i)$	MDI

models using combinations of marginal distributions shown in Table 1. The next example shows a few of these applications.

Example 5 (Generalized Gumbel Model): Let $\bar{F}_i(x_i) = e^{-\theta_i x_i}$, $x_i \geq 0$, $\theta_i > 0$, $i = 1, 2$, $Y_i = \phi_i(X_i)$, and ϕ_i be one-to-one. Applying these transformations to the bivariate Gumbel model (19) gives the bivariate generalized Gumbel distribution with survival function

$$\bar{G}^{\phi}(y_1, y_2) = \bar{G}_1(y_1)\bar{G}_2(y_2)e^{-\delta \log \bar{G}_1(y_1) \log \bar{G}_2(y_2)}, \qquad y_i > \phi_i(0).\ i = 1, 2,\ 0 \leq \delta \leq 1. \tag{33}$$

(a) If ϕ_i, $i = 1, 2$, are increasing, then (33) with parameter restrictions (21) is the MDI model $\bar{G}^{\phi}(y_1, y_2) = \bar{F}^{\phi *}(y_1, y_2)$ relative to the independent distribution

$$\bar{F}_0^{\phi}(y_1, y_2) = e^{-\lambda_1 \phi_1^{-1}(y_1) - \lambda_2 \phi_2^{-1}(y_2)}, \qquad 0 < \lambda_i < \theta_i,\ i = 1, 2, \tag{34}$$

in the class of distributions $\Omega_{G^{\phi}}$ defined in (30), with Ω_G being the class where (19) as the MDI model.

(b) If ϕ_i are increasing and convex, then (33) with $0 < \delta < 0.5$ is the ME model $\bar{G}^{\phi}(x_1, x_2) = \bar{F}^{\phi *}(x_1, x_2)$ in the set of all absolutely continuous bivariate distributions $\Omega_{G^{\phi}}$ defined in (30) with Ω_G being the class where (19) is the ME model.

(c) Let $\Omega_{Q^{\phi}}$ in (31) be the set of all bivariate distributions with hazard gradients satisfying the inequalities

$$R^{(i)\phi}(y_1, y_2) \geq R_i^{\phi}(y_i)[1 - b \log \bar{F}_j^{\phi}(y_j)], \quad b > 0,\ j \neq i = 1, 2, \tag{35}$$

where $\bar{F}_i^{\phi}(y_i) = e^{-\theta_i \phi_i^{-1}(y_i)}$ and $R_i^{\phi}(y_i) = \theta_i J_i(y_i)$, $i = 1, 2$, are survival functions and hazard functions of Y_i, respectively, and $J_i(y_i)$ is defined in (32). If ϕ_i, $i = 1, 2$, are increasing, then (33) with parameter restrictions (25) is the MDDI model $\bar{G}^{\phi}(x_1, x_2) = \bar{F}^{\phi *}(x_1, x_2)$, relative to the bivariate residual distribution of (34), in the class of distributions $\Omega_{Q^{\phi}}$ defined in (31), with Ω_Q being the class where (19) as the MDDI model. If ϕ_i are increasing and convex, then (33), with $0 < \delta < 0.5$, is the MDE model $\bar{G}^{\phi}(x_1, x_2) = \bar{F}^{\phi *}(x_1, x_2)$ in the set of all absolutely continuous bivariate distributions $\Omega_{Q^{\phi}}$ defined in (31), with Ω_Q being the class where (19) is the MDE model.

(d) For each of the models shown in Table 1, transformation of the linear inequality constraints (35) define nonlinear volumes for the hazard gradient of the model. As an example, let $Y_i = e^{X_i}$. Then $\bar{G}_i(y_i) = y_i^{-\theta_i}$, and the hazard gradient inequalities (35) become

$$R^{(i)}(y_1, y_2) \geq \theta_i y_i^{-1}(1 + b\theta_j \log y_j), \qquad b > 0,\ j \neq i = 1, 2.$$

Since $\phi_i, i = 1, 2$, are increasing and convex, the MDE model in the transformed class is the generalized Gumbel distribution with Pareto marginals

https://doi.org/10.1017/S0269964810000185 Published online by Cambridge University Press

having survival function

$$\bar{G}^{\phi}(y_1, y_2) = y_1^{-\theta_1} y_2^{-\theta_2} e^{-\delta\theta_1\theta_2 \log y_1 \log y_2}, \qquad y_1, y_2 \geq 1, \ 0 \leq \delta = b \leq 0.5. \tag{36}$$

The MDDI model, relative to the independent Pareto distribution $\bar{G}_0(y_1, y_2) = y_1^{-\lambda_1} y_2^{-\lambda_2}$, $0 < \lambda_i \leq \theta_i$, $i = 1, 2$, in the transformed class is (36) with the narrower parameter space (21).

By Theorem 3, the set of bivariate distributions $\Omega_{Q^{\phi}}$ in Example 5 does not contain any distribution for which its survival function is TP2.

5. MIXTURE OF HAZARD GRADIENTS

This section presents models for mixtures of hazard gradients and mixtures of distribution functions. The following result is used for developing models based on mixtures of hazard gradients.

LEMMA 3: *Let* $R(\boldsymbol{x}) = \sum_{k=1}^{m} \pi_k R_k(\boldsymbol{x})$, $\pi_k \geq 0$, $\sum_{k=1}^{m} \pi_k = 1$, *where* $R_k(\boldsymbol{x})$, $k = 1, \ldots, m$, *are hazard gradients. Then* $R(\boldsymbol{x})$ *is the hazard gradient of*

$$\bar{F}(\boldsymbol{x}) = \prod_{k=1}^{m} \left[\bar{F}_k(\boldsymbol{x})\right]^{\pi_k}. \tag{37}$$

PROOF: Letting $\log \bar{F}(\boldsymbol{x}) = \sum_{k=1}^{m} \pi_k \log \bar{F}_k(\boldsymbol{x})$ in (1) gives the result. ■

Since $\bar{F}(\boldsymbol{x})$ is the geometric mean of $\bar{F}_k(\boldsymbol{x})$, $k = 1, \ldots, m$, with respect to the mixing distribution $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_m)$, we have stochastic dominance $\bar{F}(\boldsymbol{x}) \leq \sum_{k=1}^{m} \pi_k \bar{F}_k(\boldsymbol{x})$ for all $\boldsymbol{x} \in \Re_+^n$. However, unlike the mixture of distributions $\sum_{k=1}^{m} \pi_k \bar{F}_k(\boldsymbol{x})$, in general, the PDF of (37) is unwieldy. A simplified case is when $R_k(\boldsymbol{x})$, $k = 1, \ldots, m$, are hazard gradients of members of a family of distributions $F_k \in \mathcal{F}_{\boldsymbol{\theta}}$, $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_q)$, which is closed under convex combinations. In this case, $R(\boldsymbol{x}) = \sum_{k=1}^{m} \pi_k R_k(\boldsymbol{x})$ is the hazard gradient of a distribution $F \in \mathcal{F}_{\boldsymbol{\theta}}$. The following example illustrates this case.

Example 6 (Gumbel and Generalized Gumbel): Let Ω_Q be defined by

$$R(x_1, x_2) \geq \pi Q_1(x_1, x_2) + (1 - \pi) Q_2(x_1, x_2), \ 0 \leq \pi \leq 1, \tag{38}$$

where $Q_k(x_1, x_2)$, $k = 1, 2$, are hazard gradients of two bivariate distributions.

https://doi.org/10.1017/S0269964810000185 Published online by Cambridge University Press

(a) As in part (a) of Example 3, let

$$Q_k^{(i)}(x_1, x_2) = a_{ki} + b_k x_j, \ \ j \neq i = 1, 2, \ a_{k1}, \ a_{k2}, \ b_k \geq 0. \qquad \textbf{(39)}$$

Then the components $Q^{(i)}(x_1, x_2), \ i = 1, 2$, of the mixture of hazard gradients (38) are in the same form as (39), where a_{ki} and b_k are replaced with

$$a_i = \pi a_{1i} + (1 - \pi) a_{2i} \geq 0 \quad \text{and} \quad b = \pi b_1 + (1 - \pi) b_2 \geq 0; \qquad \textbf{(40)}$$

that is, $Q^{(i)}(x_1, x_2), \ i = 1, 2$, are components of the hazard gradient of the Gumbel distribution (19) with $\theta_i = a_i$ and $\delta = \delta^* = b/a_1 a_2$. Thus, all of the results noted in Example 3 for the Gumbel model apply here with Ω_Q, where the components of the hazard gradient of $G(x_1, x_2) = F^*(x_1, x_2)$ are given by (39). Figure 2 shows the hazard envelope when $\pi = 0.75, \ a_1 = a_2 = 1, \ \delta_1 = 0.1$, and $\delta_2 = 0.5$, hence, $\delta^* = b = 0.2$ given by (40).

(b) In (39), let $Y_i = \phi_i(X_i)$, where X_i are exponential $X_i \sim \theta_i e^{-\theta_i x_i}, \ x_i \geq 0, \ i = 1, 2$, and $\phi_i, \ i = 1, 2$, be as in Example 5. Then the components $Q^{(i)}(y_1, y_2), \ i = 1, 2$, of the mixture hazard gradient (39) are in the same form as (35) where b_k is replaced with (40), that is, $Q^{\phi(i)}(y_1, y_2) = \phi\left(Q^{(i)}(x_1, x_2)\right), \ i = 1, 2$, are components of the hazard gradient of the generalized Gumbel distribution (33) with parameters as in part (a). Thus, all of the results noted in Example 5 for the generalized Gumbel model apply here with Ω_Q where the components of the hazard gradient of $G^{\phi}(y_1, y_2) = F^{\phi *}(y_1, y_2)$ are given by the transformed (39). This result is due to the fact that the generalized Gumbel family with parameters $\delta_k \ k = 1, 2$, and the same marginals $\bar{G}_{k1} = \bar{G}_1, \ \bar{G}_{k2} = \bar{G}_2, \ k = 1, 2$, is closed under convex combinations of the gradient hazards,

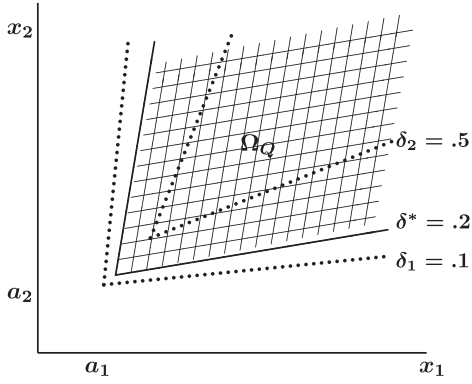


FIGURE 2. The Ω_Q envelope (shaded region) with the Gumbel borders $b = \delta^* = 0.2$ obtained as the mixture of two Gumbel hazard gradients, $Q^{(i)}(x_1, x_2) = a_i + b x_j$, $j \neq i = 1, 2, \ a_i = a_{ki}, \ k, \ i = 1, 2, \ \delta_1 = 0.1, \ \delta_2 = 0.5$, and $\pi = 0.75$.

https://doi.org/10.1017/S0269964810000185 Published online by Cambridge University Press

and by (41),

$$\bar{G}^{\phi}(y_1, y_2) = \bar{G}_1(y_1)\bar{G}_2(y_2)e^{-\delta^* \log \bar{G}_1(y_1) \log \bar{G}_2(y_2)}, \quad y_i > \phi_i(0), \quad i = 1, 2,$$

with $0 \leq \delta^* = \pi\delta_1 + (1-\pi)\delta_2 \leq 1$. Thus, mixtures of hazard gradients can be constructed for each transformation shown in Table 1: Pareto I ($\alpha_{ki} = \alpha_i$,

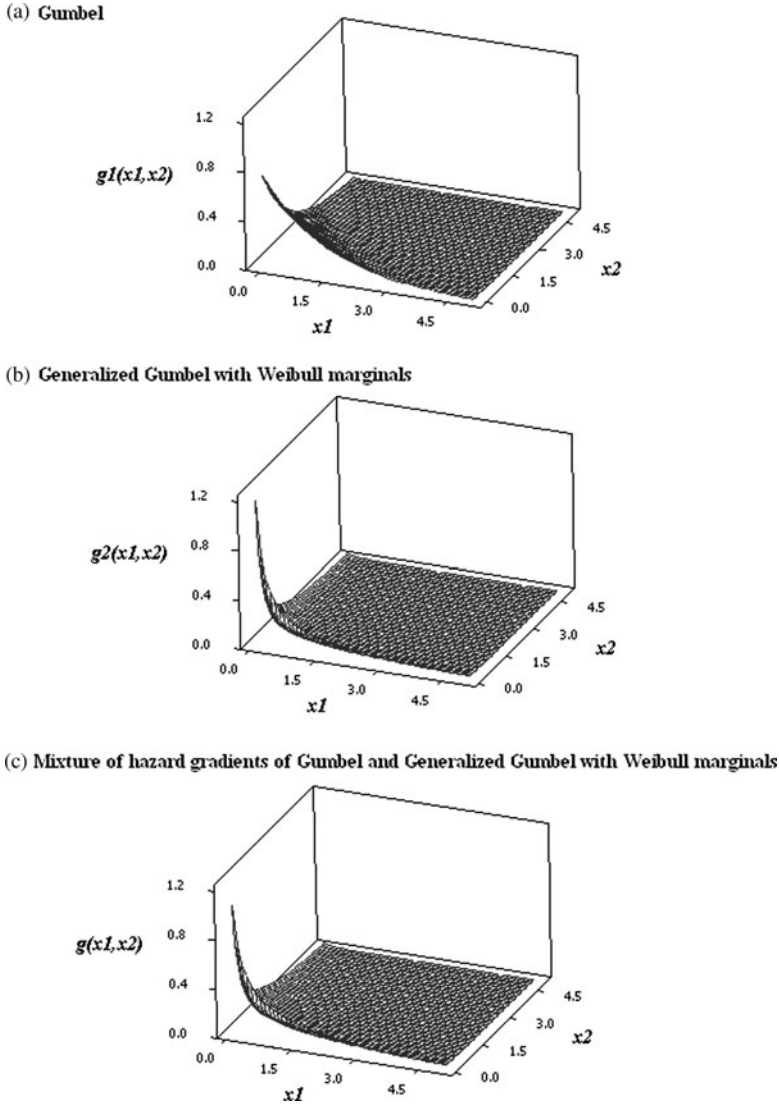


FIGURE 3. PDFs of Gumbel $g_1(x_1, x_2)$, generalized Gumbel with Weibull marginals $g_2(x_1, x_2)$ (shape = 0.5), and $g(x_1, x_2)$ found by the mixture of their hazard gradients.

https://doi.org/10.1017/S0269964810000185 Published online by Cambridge University Press

$k = 1, 2$), Pareto II, Pareto IV ($\alpha_{ki} = \alpha_i,\ k = 1, 2$), Weibull ($\alpha_{ki} = \alpha_i,\ k = 1, 2$), linear failure rate ($\alpha_{ki} = \alpha_i, k = 1, 2$), and extreme value.

It can be shown that for the bivariate case with $m = 2$, the PDF can be represented as

$$f(x_1, x_2) = \bar{F}(x_1, x_2)\left[\pi_1\Lambda_1(x_1, x_2) + \pi_2\Lambda_2(x_1, x_2) - \pi_1\pi_2 D^{(1)}(x_1, x_2)D^{(2)}(x_1, x_2)\right],$$
$$\pi_1, \pi_2 \geq 0,\ \ \pi_1 + \pi_2 = 1, \tag{41}$$

where $\Lambda_k(x_1, x_2) = f_k(x_1, x_2)/\bar{F}_k(x_1, x_2),\ k = 1, 2$, and $D^{(i)}(x_1, x_2) = R_1^{(i)}(x_1, x_2) - R_2^{(i)}(x_1, x_2), i = 1, 2$, is the difference between the ith elements of the hazard gradients of F_1 and F_2. For given $\bar{F}_k(x_1, x_2),\ k = 1, 2$, representation (41) facilitates the study of $f(x_1, x_2)$ computationally. Simple codes can be written to verify conditions of Lemma 1 numerically. For example, representation (41) can be used to study the generalized Gumbel family having the same family of the marginals when $\alpha_{1i} \neq \alpha_{2i}$. Let

$$Q_1^{(i)}(x_1, x_2) = 1 + \delta x_j,\ \ j \neq i = 1, 2,$$
$$Q_2^{(i)}(x_1, x_2) = \alpha x_i^{\alpha-1}\left(1 + \delta x_j^{\alpha}\right),\ \ j \neq i = 1, 2.$$

Here, $\bar{G}_1(x_1, x_2) = e^{-x_1 - x_2 - \delta x_1 x_2}$ is the Gumbel distribution (19) with $\theta_i = 1,\ i = 1, 2$, and $\bar{G}_2(x_1, x_2) = e^{-x_1^{\alpha} - x_2^{\alpha} - \delta x_1^{\alpha} x_2^{\alpha}}$ is the generalized Gumbel distribution (33) with Weibull marginals. For $\alpha \neq 1,\ Q_1(x_1, x_2)$ and $Q_2(x_1, x_2)$ are not closed under convex combinations. We use representation (41) to study $g(x_1, x_2)$ computationally and find that $g(x_1, x_2)$ satisfies conditions of Lemma 1. Figure 3 shows plots of $g(x_1, x_2)$ computed numerically using (41) for $\delta = 0.25,\ \alpha = 0.5$, and $\pi = 0.25$ along with PDFs $g_1(x_1, x_2)$ and $g_2(x_1, x_2)$. The density plots indicate that $g(x_1, x_2)$ has a pattern similar to $g_1(x_1, x_2)$ and $g_2(x_1, x_2)$.

6. CONCLUDING REMARKS

The literature on developing information optimal multivariate models and the ME characterizations of multivariate distributions based on partial information formulated in terms of moments constraints is extensive [7]. This article is the first attempt in providing results for developing information optimal multivariate models and the ME characterizations of bivariate distributions based on partial information formulated in terms of inequality constraints for the survival function. These results enabled us to obtain the ME and information characterizations of the bivariate Gumbel and FGM families for which no other information characterization is available.

The multivariate dynamic information measures developed in [6] set the stage for extending the univariate MDE and MDDI procedures proposed in [1,3] to higher dimensions. This article took the first step in this direction and provided results for

https://doi.org/10.1017/S0269964810000185 Published online by Cambridge University Press

developing multivariate models based on partial information formulated in terms of inequality constraints for the hazard gradient. These results enabled us to obtain the MDE and MDDI characterizations of several models including the bivariate Gumbel and Lomax models.

Using the properties of entropy and information under transformation of the variables, we developed results that enabled us to obtain the MDE and MDDI characterizations of a generalized bivariate Gumbel model based on partial information about the hazard gradient in terms of marginal hazard rates. The result is applicable when the marginals are some important univariate distributions that are transformations of exponential such as Pareto, Weibull, and extreme value distributions. The marginals need not to be in the same univariate family. Other examples included the mixture of generalized bivariate Gumbel having Pareto and Weibull marginals.

The results and methodologies developed in this article can be extended in a few directions. A natural next step is to extend our results based on the inequality constraints for the bivariate survival function to higher dimensions and develop multivariate models and characterizations of multivariate distributions with such constraints. Another natural next step is developing models with the dynamic information criteria and constraints in terms of reversed hazard rate function and the vector of mean residuals. The dynamic information criteria used in this article are Shannon entropy, Kullback–Leibler information, and the mutual information function of the multivariate residual distribution. It is worth to considering developing multivariate models using other dynamic information criteria such as the Shannon entropy and Kullback–Leibler information of the past lifetime distributions through extensions of univariate measures developed by Di Crescenzo and Longobardi [4,5]. Finally, it is worth studying developing multivariate models using dynamic versions of other information measures such a Rényi entropy and information divergence through extensions of univariate results given in [2,3].

Acknowledgments

Majid Asadi's research is supported by the Ordered and Spatial Data Center of Excellence of Ferdowsi University of Mashhad, Iran. Ehsan Soofi's research is supported by a Lubar School's Business Advisory Council Summer Research Fellowship.

References

1. Asadi, M., Ebrahimi, N., Hamedani, G.G., & Soofi, E.S. (2004). Maximum dynamic entropy models. *Journal of Applied Probability* 41: 379–390.
2. Asadi, M., Ebrahimi, N., & Soofi, E.S. (2005). Dynamic generalized information measures. *Statistics and Probability Letters* 71: 85–98.
3. Asadi, M., Ebrahimi, N., Hamedani, G.G., & Soofi, E.S. (2005). Dynamic minimum discrimination information models. *Journal of Applied Probability* 42: 643–660.
4. Di Crescenzo, A. & Longobardi, M. (2002). Entropy-based measure of uncertainty in past lifetime distributions. *Journal of Applied Probability* 39: 434–440.

https://doi.org/10.1017/S0269964810000185 Published online by Cambridge University Press

5. Di Crescenzo, A. & Longobardi, M. (2004). A measure of discrimination between past life-time distributions. *Statistics and Probability Letters* 67: 173–182.
6. Ebrahimi, N., Kirmani, S.N.U.A., & Soofi, E.S. (2007). Dynamic multivariate information. *Journal of Multivariate Analysis* 98: 328–349.
7. Ebrahimi, N., Soofi, E.S., & Soyer, R. (2008). Multivariate maximum entropy identification, transformation, and dependence. *Journal of Multivariate Analysis* 99: 1217–1231.
8. Ebrahimi, N., Soofi, E.S., & Soyer, R. (2010). Information measures in perspective. *International Statistical Review*, 78: 1018–1019.
9. Ebrahimi, N., Hamedani, G.G., Soofi, E.S., & Volkmer, H. (2010). A class of models for uncorrelated random variables. *Journal of Multivariate Analysis* 101: 1859–1871.
10. Gokhale, D.V. & Kullback, S. (1978). *The information in contingency tables*. New York: Marcel Dekker.
11. Hu, H., Khaledi, B., & Shaked, M. (2003). Multivariate hazard rate orders. *Journal of Multivariate Analysis* 84: 173–189.
12. Jaynes, E.T. (1957). Information theory and statistical mechanics. *Physics Review* 106: 620–630.
13. Joe, H. (1997). *Multivariate models and multivariate dependence concepts*. New York: Chapman Hall.
14. Johnson, N.L. & Kotz, S. (1975). A vector multivariate hazard rate. *Journal of Multivariate Analysis* 5: 53–66.
15. Karia, S.R. & Deshpande, J.V. (1999). Bounds for the hazard gradients in the competing risks set up. *Journal of Statistical Planning and Inference* 75: 363–377.
16. Khaledi, B.E. & Kochar, S. (2005). Dependence, dispersiveness, and multivariate hazard rate ordering. *Probability in the Engineering and Informational Sciences* 19: 427–446.
17. Kotz, S., Navarro, J., & Ruiz, J.M. (2007). Characterizations of Arnold and Strauss' and related bivariate exponential models. *Journal of Multivariate Analysis* 98: 1494–1507.
18. Kullback, S. (1959). *Information theory and statistics*. New York: Wiley. (Reprinted in 1968 by Dover.)
19. Lee, M.L.T. (1985). Dependence by reverse regular rule. *Annals of Probability* 13: 583–591.
20. Marshall, A.W. (1975). Some comments on the hazard gradient. *Stochastic processes and their Applications* 3: 293–300.
21. McGill, J.I. (1992). The multivariate hazard gradient and moments of the truncated multinormal distribution. *Communications in Statistics, Theory, and Methods* 21: 3053–3060.
22. Navarro, J. & Ruiz, J.M. (2004). A characterization of the multivariate normal distribution by using the hazard gradient. *Annals of the Institute of Statistical Mathematics* 56: 361–367.
23. Shaked, M. & Shanthikumar, J.G. (2007). *Stochastic orders*. Amsterdam: Springer.
24. Shore, J.E. & Johnson, R.W. (1980). Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy. *IEEE Transactions on Information Theory* 26: 26–37.

https://doi.org/10.1017/S0269964810000185 Published online by Cambridge University Press