Available online at www.sciencedirect.com
ScienceDirect
Journal of Multivariate Analysis 98 (2007) 328 – 349
Journal of Multivariate Analysis
www.elsevier.com/locate/jmva

Multivariate dynamic information

Nader Ebrahimi[a,*], S.N.U.A. Kirmani[b], Ehsan S. Soofi[c]

[a] *Division of Statistics, Northern Illinois University, DeKalb, IL 60155, USA*
[b] *Department of Mathematics, University of Northern Iowa, Cedar Falls, IA 50614, USA*
[c] *School of Business Administration, University of Wisconsin-Milwaukee, P.O. Box 742, Milwaukee, WI 53201, USA*

Received 10 March 2005
Available online 6 October 2005

Abstract

This paper develops measures of information for multivariate distributions when their supports are truncated progressively. The focus is on the joint, marginal, and conditional entropies, and the mutual information for residual life distributions where the support is truncated at the current ages of the components of a system. The current ages of the components induce a joint dynamic into the residual life information measures. Our study of dynamic information measures includes several important bivariate and multivariate lifetime models. We derive entropy expressions for a few models, including Marshall–Olkin bivariate exponential. However, in general, study of the dynamics of residual information measures requires computational techniques or analytical results. A bivariate gamma example illustrates study of dynamic information via numerical integration. The analytical results facilitate studying other distributions. The results are on monotonicity of the residual entropy of a system and on transformations that preserve the monotonicity and the order of entropies between two systems. The results also include a new entropy characterization of the joint distribution of independent exponential random variables.

AMS 1991 subject classification: 94A15; 60E15; 60B05

Keywords: Entropy; Independence; Kullback–Leibler information; Mutual information; Reliability; Residual life

1. Introduction

Study of duration is a subject of interest common to reliability, survival analysis, actuary, economics, business, and many other fields. In lifetime studies, consideration of the current age truncates the support of lifetime distribution progressively and leads to the past and remaining

* Corresponding author. Fax: +1 815 753 6776.
E-mail addresses: nader@math.niu.edu (N. Ebrahimi), kirmani@math.uni.edu (S.N.U.A. Kirmani), esoofi@uwm.edu (E.S. Soofi).

0047-259X/$ - see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmva.2005.08.004

lifetime distributions where the age becomes a parameter. The information measures of the truncated distributions are functions of time, and thus are dynamic. When the subject of duration study is other than lifetime (e.g., search time, unemployment period) the present time point plays the role of "age". More generally, the dynamic measures are applicable to any continuous distribution with a positive support. For example, for the distributions of wage, income, and depletable natural resources such as petroleum, the minimum wage, poverty line, and amount of oil extracted to date play the role of the current age, respectively.

Several authors have considered information functions that take age into account in the univariate case [5,7,11–13,17,18]. Consideration of age has led to some important insights about lifetime models such as an information characterization of the proportional hazards model [17] and maximum dynamic entropy characterization of various univariate lifetime models, including some mixture distributions, for which no other maximum entropy formulation is available [5]. Tests of distributional hypotheses based on the univariate dynamic information have been developed for reliability analysis [14,15].

Frequently, the subject of duration analysis is a group of items or individuals such as components of a system or members of a family. Capturing effects of the members' ages on the information about their remaining lifetimes as a group, and on the dependence between their remaining lifetimes are important in many applications. Examples in reliability and survival analysis are abundant. In the theory of multiple life functions in actuarial science, for example, the time-until-failure of a status is a function of the current lifetimes of the lives involved. Similar issues arise in other fields.

The objective of this paper is to develop measures of information for multivariate lifetime distributions when their supports are truncated at the current ages of the components of a system. The primary objectives are to introduce measures for assessing: (a) whether a distribution becomes more/less informative about prediction of the remaining lifetimes jointly, individually, and conditionally, and (b) if the components of a system become more/less dependent as they age. Such assessments are essential when using a multivariate model in an application. The entropies of joint, marginal, and conditional residual distributions provide diagnostics for the first purpose and the mutual information of the residual distribution serves the second purpose.

Residual information measures may be found in closed form for a few well-known distributions. As an example, we derive the residual entropy of Marshall–Olkin bivariate exponential distribution [29]. For some distributions, residual information measures may be studied using numerical integration. As an example, we study residual information measures of a bivariate gamma distribution. In general, residual information measures are mathematically unwieldy. Examples include bivariate Gumbel [22] and bivariate exponential conditionals (BEC) of Arnold and Strauss [4]. We develop some results that are useful for studying residual information measures of these and other bivariate lifetime models which possess mathematically unwieldy information measures.

Several lifetime models, such as a multivariate exponential and multivariate Weibull, can be obtained from multivariate Pareto distribution by transformations. The information measures for these distributions can be studied via the information measures of Pareto [10]. Pareto distribution yields closed form dynamic information measures, which are particularly useful because the moment-based measures such as variance and correlation coefficient do not exist for the entire family. We provide some results that enable one to examine the information properties of lifetime models that can be obtained by transformations of simpler models.

This paper is organized as follows. Section 2 presents information measures for the bivariate residual lifetime distribution. Section 3 gives results on the dynamic behavior of some entropy measures. To simplify the notations, we consider the bivariate case which are extendable to multivariate. Section 4 explores the information properties of five well-known bivariate lifetime models. Section 5 presents multivariate information measures, gives some results on transformations of dynamic entropy, and explores the properties of multivariate Pareto and some related distributions. Section 6 gives some concluding remarks. Throughout the paper "increasing" means "non-decreasing" and "decreasing" means "non-increasing".

2. Bivariate residual information

Let (X_1, X_2) be a vector of non-negative random variables. We may think of $X_j,\ j = 1, 2$, as the lifetimes of the members of a group or components of a system. (An alternative interpretation is given in [19]) At ages $t_1, t_2,\ t_j \geqslant 0$ of the components, the joint residual lifetime distribution is the conditional (truncated) distribution denoted by $F(x_1, x_2; t_1, t_2) = P_F(X_1 \leqslant x_1, X_2 \leqslant x_2 | X_1 > t_1, X_2 > t_2)$. The residual density function will be denoted by $f(x_1, x_2; t_1, t_2) = \frac{f(x_1,x_2)}{\bar{F}(t_1,t_2)}$ for $x_1 > t_1, x_2 > t_2$, where $f(x_1, x_2)$ is the probability density functions and $\bar{F}(t_1, t_2) = P(X_1 > t_1, X_2 > t_2)$ is the joint survival function. For new systems and minimal repairs it is reasonable to assume that $t_1 = t_2 = t$. The general case of (t_1, t_2) includes the equal ages as well as the case when the components are replaced with new components.

The discrimination information function of interest is the mean information in a vector of observations $(x_1, x_2), x_j \geqslant t_j, j = 1, 2$, for discriminating between two residual life distributions, given by the Kullback–Leibler function

$$K(f : g; t_1, t_2) = \int_{t_2}^{\infty} \int_{t_1}^{\infty} f(x_1, x_2; t_1, t_2) \log \frac{f(x_1, x_2; t_1, t_2)}{g(x_1, x_2; t_1, t_2)}\, dx_1\, dx_2, \tag{1}$$

where $g(x_1, x_2; t_1, t_2) = g(x_1, x_2)/\bar{G}(t_1, t_2)$, and $\bar{G}(t_1, t_2)$ is the survival function of G; F is absolutely continuous with respect to G.

It is clear that $K(f : g; 0, 0) = K(f : g)$ is the Kullback–Leibler function between F and G. By (1), for each (t_1, t_2), $K(f : g; t_1, t_2)$ possesses all the properties of the discrimination information function $K(f : g)$. If we consider $\mathcal{T} = \{(t_1, t_2) : t_j > 0,\ j = 1, 2\}$ as a bivariate index set, then $K(f : g; t_1, t_2)$ provides a dynamic discrimination information ranging over $\mathcal{T}$. The discrimination information function has many desirable properties [20,27]. Two properties of particular interest to us are: (a) $K(f : g; t_1, t_2) \geqslant 0$ and the equality holds if and only if $f(x_1, x_2; t_1, t_2) = g(x_1, x_2; t_1, t_2)$, almost everywhere; and (b) if $(Y_1, Y_2) = \phi(X_1, X_2)$ is a non-singular transformation, then $K[f_Y : g_Y; \phi(t_1, t_2)] = K(f_X : g_X; t_1, t_2)$.

The joint residual entropy of an absolutely continuous distribution is given by

$$\begin{aligned} H(X_1, X_2; t_1, t_2) &\equiv H[f(x_1, x_2; t_1, t_2)] \\ &= -\int_{t_2}^{\infty} \int_{t_1}^{\infty} f(x_1, x_2; t_1, t_2) \log f(x_1, x_2; t_1, t_2)\, dx_1\, dx_2 \\ &= \log \bar{F}(t_1, t_2) - \frac{1}{\bar{F}(t_1, t_2)} \int_{t_2}^{\infty} \int_{t_1}^{\infty} f(x_1, x_2) \\ &\quad \times \log f(x_1, x_2)\, dx_1\, dx_2. \end{aligned} \tag{2}$$

The residual entropy (2) measures uncertainty of the remaining lifetimes when the ages of components are t_1, t_2. Note that for the uniform distribution with rectangular support $\{x_j : a_j < x_j < b_j,\ j = 1, 2\}$, the residual life distribution is also uniform over $\{x_j : t_j < x_j < b_j,\ j = 1, 2\}$. Thus the negative entropy $-H[f(x_1, x_2; t_1, t_2)]$ measures lack of uniformity of $f(x_1, x_2; t_1, t_2)$.

The informativeness of $f(x_1, x_2; t_1, t_2)$ about the prediction of its outcomes is measured by negative entropy $I(X_1, X_2; t_1, t_2) = -H(X_1, X_2; t_1, t_2)$, which is the average log-height of the density [34].

Unlike the discrimination information (1), the joint entropy (2) is not invariant under nonsingular transformations. It can be shown that if $Y_j = \phi_j(X_j),\ j = 1, 2$, are one-to-one transformations, then

$$H[Y_1, Y_2; \phi_1(t_1), \phi_2(t_2)] = H(X_1, X_2; t_1, t_2) \\ -E[\log J(X_1, X_2)|X_1 > t_1, X_2 > t_2], \tag{3}$$

where

$$J(y_1, y_2) = \left|\frac{\partial\phi_1^{-1}(y_1)}{\partial y_1} \times \frac{\partial\phi_2^{-1}(y_2)}{\partial y_2}\right|$$

is the absolute value of the Jacobian of transformation and the expectation is taken with respect to the residual distribution.

The *marginal residual entropy* of X_1 given $X_1 > t_1, X_2 > t_2$ is

$$H(X_1; t_1, t_2) = H[f_1(x_1; t_1, t_2)] = -\int_{t_1}^{\infty} f_1(x_1; t_1, t_2) \log f_1(x_1; t_1, t_2)\, dx_1, \tag{4}$$

where

$$f_1(x_1; t_1, t_2) = \int_{t_2}^{\infty} f(x_1, x_2; t_1, t_2)\, dx_2, \quad x_1 > t_1, x_2 > t_2 \tag{5}$$

is the marginal residual density of X_1 given $X_1 > t_1, X_2 > t_2$. The marginal residual density and entropy of X_2 given $X_1 > t_1, X_2 > t_2$ are defined similarly.

Comparison of $H(X_j; t_1, t_2)$ with the *univariate residual entropy* $H(X_j; t_j) = H[f_j(x_j; t_j)]$ provides an assessment of the effect of considering both ages vis á vis only one age t_j on the uncertainty about the residual lifetime of the component j. In general, the difference $H(X_j; t_j) - H(X_j; t_1, t_2)$, $j = 1, 2$, can be positive or negative. That is, considering both ages may decrease or increase the uncertainty about the residual lifetime of a component or leave it unchanged. A definite answer is possible under certain conditions. For example, it can be shown that if $P(X_i > x_i|X_j > x_j)$ is decreasing in x_j for all x_i (X_i is right tail decreasing dependence in $X_j,\ j \neq i$ [16]), and if the univariate marginal distributions $f_j(x_j),\ j = 1, 2$, have decreasing hazard rates, then $H(X_j; t_1, t_2) \leqslant H(X_j; t_j)$. That is, right tail decreasing dependence and marginal decreasing hazard rates are sufficient for the uncertainty reduction about the residual lifetime of one component due to consideration of the age of the other component.

The conditional residual density of X_j given that $X_j > t_j, X_i = x_i$ is

$$f_{j|i}(x_j|x_i; t_1, t_2) = \begin{cases} \dfrac{f(x_1, x_2; t_1, t_2)}{f_i(x_i; t_1, t_2)}, & x_1 > t_1, x_2 > t_2, \\ 0 & \text{otherwise.} \end{cases} \tag{6}$$

Its entropy is denoted by $H[f_{j|i}(x_j|x_i; t_1, t_2)] = H(X_j|x_i; t_1, t_2), i \neq j = 1, 2$, which in general is a function of x_i.

The *conditional residual entropy* is defined by

$$H(X_j|X_i; t_1, t_2) = \int_{t_i}^{\infty} H(X_j|x_i; t_1, t_2) f_i(x_i; t_1, t_2)\, dx_i. \tag{7}$$

This measure quantifies uncertainty about X_j on average when we know $X_i,\ i \neq j$. The conditional entropy plays important roles in entropy decomposition and information measure of dependence discussed below. In Section 3, we will show that the conditional residual entropy characterizes the independent exponential distribution.

The joint residual entropy can be decomposed as

$$H(X_1, X_2; t_1, t_2) = H(X_i; t_1, t_2) + H(X_j|X_i; t_1, t_2), \quad i \neq j, \quad j = 1, 2. \tag{8}$$

An important question in statistics is to what extent the use of a variable X_i reduces uncertainty about predicting the outcomes of another variable X_j. The worth of an outcome x_i of X_i for predicting X_j, given that $X_1 > t_1$, $X_2 > t_2$, is assessed by comparing $f_j(x_j; t_1, t_2)$ and $f_{j|i}(x_j|x_i; t_1, t_2)$. Two information measures that may be used for this purpose are $K[f_{j|i}(x_j|x_i) : f_j(x_j); t_1, t_2]$ and the entropy difference

$$\vartheta(X_j|x_i; t_1, t_2) = H(X_j; t_1, t_2) - H(X_j|x_i; t_1, t_2). \tag{9}$$

In general, both $K[f_{j|i}(x_j|x_i) : f_j(x_j); t_1, t_2]$ and $\vartheta(X_j|x_i; t_1, t_2)$ depend on x_i. The non-negative discrimination function $K[f_{j|i}(x_j|x_i) : f_j(x_j); t_1, t_2]$ quantifies the discrepancy between the marginal and conditional distributions, but it does not indicate which of the two distributions is more informative for the prediction of the outcome of X_j. The entropy difference (9) may be positive or negative depending on which of the two distributions is more informative. Interestingly, the mean values of the two measures are the same and defines the mutual information between two variables:

$$M(X_1, X_2; t_1, t_2) = \int_{t_i}^{\infty} \vartheta(X_j|x_i; t_1, t_2) f_i(x_i; t_1, t_2)\, dx_i \tag{10}$$

$$= \int_{t_i}^{\infty} K[f_{j|i}(x_j|x_i) : f_j(x_j); t_1, t_2] f_i(x_i; t_1, t_2)\, dx_i. \tag{11}$$

The Kullback–Leibler and entropy representations of mutual information between two residual lifetimes are as follows:

$$M(X_1, X_2; t_1, t_2) = K\,[f(x_1, x_2; t_1, t_2) : f_1(x_1; t_1, t_2) f_2(x_2; t_1, t_2)] \tag{12}$$

$$= H(X_1; t_1, t_2) + H(X_2; t_1, t_2) - H(X_1, X_2; t_1, t_2) \tag{13}$$

$$= H(X_j; t_1, t_2) - H(X_j|X_i; t_1, t_2), \quad i \neq j. \tag{14}$$

By (12), $M(X_1, X_2; t_1, t_2) \geqslant 0$ and the residual lifetimes are independent if and only if $M(X_1, X_2; t_1, t_2) = 0$. Thus, $M(X_1, X_2; t_1, t_2)$ measures the extent of dependence between the residual lifetimes.

It is clear that $M(X_1, X_2; 0, 0) = M(X_1, X_2)$, which is the global information measure of dependence. Anderson et al [2] suggested three desirable properties for association measures: (a) symmetry with respect to the coordinates when (X_1, X_2) are exchangeable; (b) being as free as possible from the influence of marginals (e.g., locations and scales of the marginals); and (c) amenable to interpretation. The mutual information possesses all these three properties at a more general level. As seen in (12) and (13), for any bivariate distribution the mutual information is symmetric with respect to the coordinates, $M(X_1, X_2) = M(X_2, X_1)$. An important property of $M(X_1, X_2)$ is invariance under one-to-one transformations of each component implied by (12). For more properties and statistical applications of mutual information see [33].

For many purposes, a single global measure of dependence might be satisfactory for failure time data. But for equally many cases it is useful to consider the dependence in more detail.

One consideration, for example, is whether the dependence is most marked at early or late times. For discussing such concepts, local measures of dependence are needed. The dynamic mutual information $M(X_1, X_2; t_1, t_2)$ measures the extent of dependency between the remaining lifetimes of the components when the components are already survived to times t_1, t_2. Thus it is a local measure, among other things, that can be used to address the concepts of early/late dependence and short-term and long-term dependence.

From (13) and (14) we have

$$H(X_1, X_2; t_1, t_2) \leqslant H(X_1; t_1, t_2) + H(X_2; t_1, t_2),$$
$$H(X_j|X_i; t_1, t_2) \leqslant H(X_j; t_1, t_2), \quad i \neq j. \tag{15}$$

The equalities hold if and only if the residual lifetimes are independent.

The following example illustrates various residual information measures, their interrelationships, and insights that can be gained from a dynamic information analysis.

Example 1. Consider the bivariate distribution with the following density on the unit square:

$$f(x_1, x_2) = \begin{cases} x_1 + x_2, & 0 \leqslant x_1, x_2 \leqslant 1, \\ 0 & \text{otherwise.} \end{cases}$$

The survival function is $\bar{F}(t_1, t_2) = .5(1 - t_1^2)(1 - t_2) + .5(1 - t_2)(1 - t_2^2)$, and the joint residual entropy is given by

$$H(X_1, X_2; t_1, t_2) = \log \bar{F}(t_1, t_2) + \frac{1}{\bar{F}(t_1, t_2)} \sum_{n=0}^{1} \sum_{m=0}^{1} (-1)^{m+n} (t_1^m + t_2^n)^3 \times \left[\frac{5}{6} - \log(t_1^m + t_2^n) \right].$$

The expressions for $H(X_j; t_1, t_2)$ and $H(X_j|X_i; t_1, t_2)$ are more messy. It can be shown that all residual entropies are decreasing in t_1 and t_2 and the rate of decrease is highest on the diagonal (t, t).

Fig. 1(a) shows the graphs of four residual entropies on the diagonal. The joint residual entropy is shown by the solid curve, the marginal residual entropy by the dashed curve, the univariate residual entropy $H(X_j; t)$ by the dashed-dotted and the conditional residual entropy by the dotted curve. We note that $H(X_j; t, t) < H(X_j; t)$, which indicates that considering the age of component i is informative about the residual life of the component $j \neq i$.

The mutual information increases in t_1 and t_2. Thus, by (13), the joint residual entropy decreases with a faster rate than the total marginal residual entropies, and by (14), the conditional residual entropy decreases with a faster rate than the marginal residual entropy. Fig. 1(b) shows the graph of $M(X_1, X_2; t, t)$. Although dependency between lifetimes of the two components is low to begin with, it sharply decreases with t. Note that the regression relationship is non-linear, $E(X_j|x_i) = (3x_i + 2)/(6x_i + 3)$, thus the correlation coefficient is not useful. The residual mutual information serves as a useful measure due to the fact that it captures any kind of functional dependency.

3. Dynamics of residual information

This section presents results that are useful for studying the dynamic information properties of some well-known bivariate distributions and addresses the question of dynamic information properties of parallel and series systems.

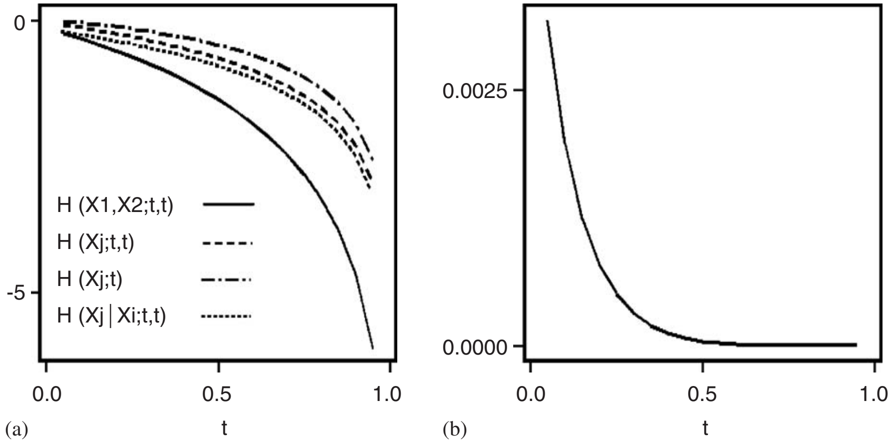

Fig. 1. Residual entropies and mutual information for the distribution of Example 1.

3.1. Memoryless information

A residual information function of a distribution is said to be memoryless if it is free from the ages of the components. In the univariate case, memoryless residual entropy characterizes the exponential distribution. In the bivariate case, some residual information, e.g., the mutual information, may be free from the ages without others being so. Also, the entire set of bivariate residual information measures may be free from the ages only in certain direction, e.g., $t_1 = t_2 = t$, with or without the marginals being exponential.

The next theorem characterizes the independent exponential model in terms of memoryless conditional entropies.

Theorem 1. *The conditional residual entropies* $H(X_j|X_i; t_1, t_2)$, $j = 1, 2$, *are constants free from* t_1, t_2 *if and only if* X_1 *and* X_2 *are independently exponentially distributed*,

$$f(x_1, x_2) = \lambda_1\lambda_2 e^{-\lambda_1 x_1 - \lambda_2 x_2}, \quad x_j > 0, \ \ \lambda_j > 0.$$

Proof. For the sufficient condition, by (14) for two independent random variables X_1 and X_2, $H(X_j|X_i; t_1, t_2) = H(X_j; t_1, t_2)$ and by the result of [13] on dynamic entropy of an exponentially distributed random variable being constant independent of time.

The proof of the necessary condition follows from the fact that

$$\frac{\partial}{\partial t_1} H(X_1|X_2; t_1, t_2) = \frac{H(X_1|x_2; t_1, t_2) - H(X_1|X_2; t_1, t_2)}{\bar{F}(t_1, t_2)} \int_{t_2}^{\infty} f(t_1, s)\, ds = 0.$$

Hence, $H(X_1|X_2; t_1, t_2) = H(X_1|x_2; t_1, t_2) = a_1$, which by assumption is constant free from $t_1, t_2 \geqslant 0$. It follows from [13] that for some $\lambda_1 > 0$,

$$f_{1|2}(x_1|x_2; t_1, t_2) = \lambda_1 e^{-\lambda_1 x_1}, \quad x_1 \geqslant 0, \ \ t_1, t_2 \geqslant 0.$$

Since the conditional residual density $f_{j|i}(x_j|x_i; t_1, t_2)$ is free from x_2 and t_j, $j = 1, 2$, the marginal residual density is $f_1(x_1; t_1, t_2) = \lambda_1 e^{-\lambda_1 x_1}$. Similarly, we obtain $H(X_2|X_1; t_1, t_2) =$

$H(X_2|x_1;t_1,t_2)=a_2$ for all $x_1,t_1,t_2\geqslant 0$ and $f_2(x_2;t_1,t_2)=\lambda_2 e^{-\lambda_2 x_2}$. Hence,

$$f(x_1,x_2;t_1,t_2)=\lambda_1 e^{-\lambda_1 x_1}\lambda_2 e^{-\lambda_2 x_2}, \quad x_1,x_2\geqslant 0. \qquad \square$$

Since for the independent exponential model $M(X_1,X_2;t_1,t_2)=0$ for all t_1 and t_2, from Theorem 1 and (14) we have $H(X_j;t_1,t_2)$, $j=1,2$, free of (t_1,t_2). We also have from (13), $H(X_1,X_2;t_1,t_2)$ free of (t_1,t_2).

A bivariate distribution is said to have the bivariate lack of memory (BLM) property if [8,29]

$$\bar{F}(s_1+t,s_2+t)=\bar{F}(s_1,s_2)\bar{F}(t,t). \tag{16}$$

The BLM property (16) implies that $\bar{F}(s_1+t,t)=\bar{F}_1(s_1)\bar{F}(t,t)$ and for an absolutely continuous distribution we have $-\frac{\partial}{\partial s_1}\bar{F}(s_1+t,t)=f_1(s_1)\bar{F}(t,t)$. Consequently $H(X_j;t,t)=H(X_j)$, $j=1,2$, and $H(X_1,X_2;t,t)=H(X_1,X_2)$ for all t. Furthermore, by (8), $H(X_j|X_i;t,t)$, and by (13), $M(X_1,X_2;t,t)$ are also free from t. However, $H(X_j;t)$ may be time-dependent; see Section 4.1. These properties of information are in accord with the intuition that under BLM, information quantities that are derived from the bivariate residual distribution are memoryless along the diagonal (t,t). Note that BLM is sufficient, but not necessary for obtaining these properties; see Section 4.2.

3.2. Monotone entropy

The following theorem gives sufficient conditions for monotonicity of $H(X_1,X_2;t_1,t_2)$ in terms of the monotonicity of the hazard rate functions $\lambda_j(x_j;t_1,t_2)$, $j=1,2$, and the conditional entropies.

Theorem 2. *If* (a) $\lambda_j(x_j;t_1,t_2)$ *is increasing* (*decreasing*) *in* x_j, $j=1,2$, *and* (b) $H(X_j|x_i;t_1,t_2)$, $i\neq j$, $j=1,2$, *is decreasing* (*increasing*) *in* x_i *for each* $j=1,2$, *then* $H(X_1,X_2;t_1,t_2)$ *is decreasing* (*increasing*) *in* t_1,t_2.

Proof. For any random variable Z with survival function $\bar{F}_Z$, density f_Z, and hazard function λ_Z, the residual entropy has the following representation:

$$H(Z;t)=1-\frac{1}{\bar{F}_Z(t)}\int_t^\infty f_Z(s)\log\lambda_Z(s)\,ds. \tag{17}$$

From (2), for $j=1,2$,

$$\frac{\partial}{\partial t_j}H(X_1,X_2;t_1,t_2)=\lambda_j(t_j;t_1,t_2)[H(X_1,X_2;t_1,t_2)-1+\log\lambda_j(t_j;t_1,t_2) -H(X_j|t_i;t_1,t_2)]. \tag{18}$$

Since $\lambda_j(x_j;t_1,t_2)$, $j=1,2$, is increasing, from (17) we have $H(X_j;t_1,t_2)\leqslant 1-\log\lambda_j(x_j;t_1,t_2)$. Noting that $P(X_i\geqslant t_i)=1$, by assumption (b) we have $H(X_j|t_i;t_1,t_2)\geqslant H(X_j|X_i;t_1,t_2)$. Now using (8), we have

$$H(X_1,X_2;t_1,t_2)\leqslant 1-\log\lambda_j(t_j;t_1,t_2)+H(X_j|t_i;t_1,t_2). \tag{19}$$

It follows from (18) and (19) that $\frac{\partial H(X_1,X_2;t_1,t_2)}{\partial t_j}\leqslant 0$, for $j=1,2$. This completes the proof for the increasing part. The proof for the decreasing part is similar. $\square$

Heuristically, Theorem 2 says that under conditions (a) and (b), the joint residual distribution progressively become more (less) informative as the individuals age, and thus prediction of lifetimes becomes easier (or more difficult). An analogue of Theorem 2 under the weaker condition when both individuals have the same age can be proved similarly.

Next we address the important question of whether the residual entropies of lifetimes of series and parallel systems of two components inherit monotonicity of the joint residual entropy.

Theorem 3. *Suppose that X_1 and X_2 are independent and identically distributed random variables.*

(a) *If $H(X_1, X_2; t, t)$ is decreasing (increasing) in t, then $H[\min(X_1, X_2); t, t]$ is decreasing (increasing) in t.*
(b) *If $H(X_1, X_2; t, t)$ is decreasing in t, then $H[\max(X_1, X_2); t, t]$ is decreasing in t.*

Proof. (a) We will prove for decreasing. From (2),

$$H[\min(X_1, X_2); t, t] = 1 - \log 2 - \frac{2}{\bar{F}^2(t)} \int_t^\infty f(s)\bar{F}(s) \log \lambda_F(s)\, ds,$$

where λ_F is the hazard function of X_j. It is sufficient to show that $\psi(t) = 1 - \log 2 - H[\min(X_1, X_2); t, t]$ is increasing in t. Integration by parts gives

$$\begin{aligned}\psi(t) &= -2[H(X_j; t) - 1] + \int_t^\infty \left[H(X_j; s) - 1\right] \frac{2f(s)}{\bar{F}(t)}\, ds \\ &= 2 - 2H(X_j; t) + \int_t^\infty \frac{2f(s)\bar{F}(s)}{\bar{F}^2(t)} \left[-\frac{1}{\bar{F}(s)} \int_s^\infty f(v) \log \lambda_F(v)\, dv\right] ds \\ &\geqslant 1 - H(X_j; t) \geqslant \log \lambda_F(t). \end{aligned} \tag{20}$$

The last two inequalities in (20) come from (17) and the fact that $H(X_1, X_2; t, t) = 2H(X_j, t)$ is decreasing in t. Taking the derivative

$$\psi'(t) = 2\lambda_F(t)[\psi(t) - \log \lambda_F(t)]$$

and using (20) we get $\psi'(t) \geqslant 0$.

(b) From (2),

$$\begin{aligned}H[\max(X_1, X_2); t] = &-\frac{2}{1 - F^2(t)} \int_t^\infty f(x)F(x) \log \frac{F(x)}{1 + F(t)}\, dx \\ &-\frac{2}{1 - F^2(t)} \int_t^\infty f(x)F(x) \log \frac{f(x)}{\bar{F}(t)}\, dx - \log 2.\end{aligned}$$

It is sufficient to show that $\psi(t) = -H[\max(X_1, X_2); t] - \log 2$ is increasing in t. That is,

$$\psi'(t) = \psi(t) - \log F(t) + \log[1 + F(t)] - \log \lambda_F(t) \geqslant 0.$$

Integration by parts gives

$$\begin{aligned}&\psi(t) - \log F(t) + \log[1 + F(t)] \\ &= -\frac{1}{2} - \frac{\log F(t)}{1 - F^2(t)} + \int_t^\infty \frac{2F(s)}{1 + F(t)} H(X_j; s)\, ds \\ &= -\frac{1}{2} - \frac{\log F(t)}{1 - F^2(t)} - H(X_j; t)\frac{2F(t)}{1 + F(t)} - \int_t^\infty H(X_j; s)\frac{2f(s)}{1 + F(t)}\, ds\end{aligned}$$

$$
\begin{aligned}
&= -\frac{1}{2} - \frac{\log F(t)}{1-F^2(t)} - H(X_j;t)\frac{2F(t)}{1+F(t)} \\
&\quad + \int_t^\infty \frac{2f(s)}{1+F(t)}\left[\int_s^\infty \frac{f(u)}{\bar{F}(t)} \log \frac{f(u)}{\bar{F}(t)}\, du\right] ds \\
&= -\frac{1}{2} - \frac{\log F(t)}{1-F^2(t)} - H(X_j;t)\frac{2F(t)}{1+F(t)} \\
&\quad - \int_t^\infty \left[\frac{2f(s)\bar{F}(s)}{\bar{F}(t)[1+F(t)]} - \frac{1}{\bar{F}(s)}\int_s^\infty \log \frac{f(u)}{\bar{F}(s)} f(u)\, du\right] ds \\
&\quad + \int_t^\infty \frac{2f(s)\bar{F}(s)}{1-F^2(t)} \log \frac{\bar{F}(s)}{\bar{F}(t)}\, ds \\
&\geqslant -H(X_j;t) - \frac{1}{1+F(t)} - \frac{\log F(t)}{1-F^2(t)} \\
&\geqslant \log \lambda_F(t) - \frac{\log F(t)}{1-F^2(t)} - \frac{2+F(t)}{1+F(t)} \geqslant \log \lambda_F(t).
\end{aligned} \tag{21}
$$

From (21) we get that $\psi'(t) \geqslant 0$. This completes the proof. □

It may be noted that $H(\max(X_1, X_2); t)$ need not be increasing even when $H(X_1, X_2; t, t)$ is increasing. For example, if X_1 and X_2 are independent with a common exponential distribution then $H[\max(X_1, X_2); t]$ is decreasing. It should also be noted that the assumption of identical distributions for X_1 and X_2 is essential in Theorem 4.

4. Examples: some bivariate lifetime models

This section illustrates applications of the dynamic information measures for some well-known bivariate lifetime models.

4.1. Marshall–Olkin model

The joint survival function of Marshall–Olkin bivariate exponential is given by

$$\bar{F}(x_1, x_2) = e^{-\lambda_1 x_1 - \lambda_2 x_2 - \lambda_{12} \max(x_1, x_2)}, \quad x_1, x_2 \geqslant 0, \ \lambda_1, \ \lambda_2 > 0, \ \lambda_{12} \geqslant 0. \tag{22}$$

The joint survival function (22) may be represented as

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

where $\lambda = \lambda_1 + \lambda_2 + \lambda_{12}$, $\bar{F}_c(x_1, x_2)$ is an absolutely continuous part with a bivariate density function

$$f_c(x_1, x_2) = \begin{cases} \dfrac{\lambda}{\lambda_1 + \lambda_2} \lambda_1(\lambda_2 + \lambda_{12})\, e^{-\lambda_1 x_1 - (\lambda_2 + \lambda_{12}) x_2}, & x_1 < x_2, \\ \dfrac{\lambda}{\lambda_1 + \lambda_2} \lambda_2(\lambda_1 + \lambda_{12})\, e^{-(\lambda_1 + \lambda_{12}) x_1 - \lambda_2 x_2}, & x_1 > x_2, \end{cases} \tag{23}$$

and $\bar{F}_s(x_1, x_2)$ is a singular part,

$$\bar{F}_s(x_1, x_2) = e^{-\lambda \max(x_1, x_2)}.$$

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, see [29].

The absolutely continuous bivariate exponential distribution (ACBED) of Block and Basu [8] is defined by the joint density (23). The ACBED distribution has the BLM property (16). Thus, $H_c(X_1, X_2; t, t)$, $H_c(X_j; t, t)$, $H_c(X_j|X_i; t, t)$, and $M_c(X_1, X_2; t, t)$ are all free from t.

For computing the entropy of (22) we use the partitioning property of Shannon entropy and a more general representation than (2). The entropy of a random vector $\boldsymbol{X}$ with a probability distribution F is defined by

$$H(\boldsymbol{X}) \equiv H(F) = -\int_{\mathcal{S}} \log[dF(\boldsymbol{x})]\, dF(\boldsymbol{x}),$$

where $\mathcal{S}$ is the support.

Let $\mathcal{A}_i,\ i = 1, \ldots, n$, be a partition of $\mathcal{S}$ and $p_i = \int_{\mathcal{A}_i} dF(\boldsymbol{x})$. Let $H_i(\boldsymbol{X})$ denote the entropy of the truncated distribution $F_i(\boldsymbol{x}) = F(\boldsymbol{x})/p_i$ for $\boldsymbol{x} \in \mathcal{A}_i$. Then

$$H(\boldsymbol{X}) = H(p_1, \ldots, p_n) + \sum_{i=1}^{n} p_i H_i(\boldsymbol{X}), \tag{24}$$

where $H(p_1, \ldots, p_n) = -\sum_{i=1}^{n} p_i \log p_i$ is the entropy of the partition (mixing) probability vector.

For the residual entropy of ACBED, we partition $\mathcal{A}_c$ by $\mathcal{A}_{c1} = \{(x_1, x_2) : x_1 < x_2\}$ and $\mathcal{A}_{c2} = \{(x_1, x_2) : x_1 > x_2\}$, where $\omega_j = P_c(\mathcal{A}_{cj}) = \lambda_j/(\lambda_1 + \lambda_2),\ j = 1, 2$, is the probability (weight) given to $\mathcal{A}_{cj}$ by ACBED. Application of (24) gives the entropy of (23) as

$$\begin{aligned} H_c(X_1, X_2; t, t) &= H(\omega_1, \omega_2) + \sum_{j=1}^{2} \omega_j H_{cj}(X_1, X_2; t, t) \\ &= H(\omega_1, \omega_2) + 2 - \log \lambda + \log \left(\sigma_1^{\omega_2} \sigma_2^{\omega_1}\right), \end{aligned} \tag{25}$$

where $H(\omega_1, \omega_2)$ is the entropy of mixing probability vector, $H_{cj}(X_1, X_2; t, t) = 2 - \log[\lambda(\lambda - \lambda_j)]$ is the residual entropy of the truncated distribution, and $\sigma_j = (\lambda_j + \lambda_{12})^{-1}$ is the standard deviation (also mean) of X_j. (Ahsanullah and Habibullah [1] computed $H_c(X_1, X_2)$ by integration. Application of (24) simplifies the computation.)

Following the probabilistic argument of Marshall and Olkin [29, p. 35] we find the entropy of (22) via partitioning the positive quadrant by $\mathcal{A}_s = \{(x_1, x_2) : x_1 = x_2\}$ and $\mathcal{A}_c = \{(x_1, x_2) : x_1 \neq x_2\}$. The distribution gives $P(\mathcal{A}_s) = \rho = \lambda_{12}/\lambda$. This probability is distributed along the line $x_1 = x_2$ according to the one-dimensional exponential density

$$f_s(x) = \lambda \bar{F}_s(x, x) = \lambda \exp\{-\lambda x\}.$$

The residual entropy of the singular part is simply $H_s(X; t, t) = 1 - \log \lambda$.

The absolutely continuous part has density $f_c(x_1, x_2)$ over the subset $\mathcal{A}_c$ with $P(\mathcal{A}_c) = (\lambda_1 + \lambda_2)/\lambda = 1 - \rho$. Application of (24) gives the entropy of (22) as

$$H(X_1, X_2; t, t) = H(\rho, 1 - \rho) + \rho H_s(X; t, t) + (1 - \rho) H_c(X_1, X_2; t, t), \tag{26}$$

where $H(\rho, 1-\rho)$ is the entropy of mixing probability vector $(\rho, 1-\rho)$, $H_s(X; t, t)$ is the residual entropy of the exponential distribution for the singular part, and $H_c(X_1, X_2; t, t)$ is the entropy of ACBED (25). The parameter $\rho = \lambda_{12}/\lambda$ is also the correlation coefficient.

Thus, $H(X_1, X_2; t, t)$ is free from t. Also $H(X_j; t, t) = 1 + \log \sigma_j$, $j = 1, 2$, is free from t. The entropy difference (13) is free from t and for the independent case is zero. But (12) is not well-defined because the joint distribution is not absolutely continuous with respect to the product measure $F_1 F_2$. As such, the mutual information is not defined.

Thus, for both, Marshall–Olkin and ACBED, when components are at the same age and that is taken into account, the informativeness of the joint residual distribution of the two components as well as the residual lifetime distribution of each component remains unchanged as they age. The dynamics of information for Marshall–Olkin and ACBED differ when the components are considered singly. The marginal distributions of Marshal–Olkin distribution are exponential, so the univariate residual entropy $H(X_j; t)$ are free from t, and the entropy difference $H(X_j; t) - H(X_j; t, t) = 0$ for all $t \geqslant 0$. Thus if both components are at the same age, there is no change of uncertainty when ages of one or both components are considered. However, since the marginal distributions of ACBED have increasing failure rates, and a result of [13] implies that, $H_c(X_j; t_j)$, $j = 1, 2$, is decreasing in t_j. Consequently, $H_c(X_j; t) - H_c(X_j; t, t)$ is decreasing in t. Thus when a component is considered singly, as it ages its residual lifetime distribution becomes more informative (concentrated).

4.2. McKay bivariate gamma

Consider McKay bivariate gamma distribution [25,30] with joint density

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

The marginal distributions are gamma with shape parameters α and β, respectively, and scale δ.

Let $\alpha = 1$ and without loss of generality we can let $\delta = 1$. Then we have a one-parameter subfamily of McKay bivariate gamma distributions where one marginal is exponential and one is gamma with shape parameter β. For this subfamily, $\bar{F}(t, t) = e^{-t}$, $f_1(x_1; t, t)$ is exponential with shift parameter t for $x_1 > t$, $f_2(x_2; t, t)$ is gamma with shift parameter t and shape parameter $\beta + 1$ for $x_2 > t$, and $f_{2|1}(x_2|x_1; t, t)$ is gamma with shape parameter β for $x_2 > x_1$, free from t. Since entropy is not a function of shift parameter, $H(X_1; t, t)$ and $H(X_2|X_1; t, t)$ are free from t. Using the exponential and gamma entropy expressions in (8) and (13) gives $H(X_1, X_2; t, t)$ and $M(X_1, X_2; t, t)$ as functions of β only. Thus, this subfamily has memoryless bivariate residual information measures along the diagonal (t, t), without having BLM property (16). However, $H(X_2; t)$ is decreasing in t because $f_2(x_2)$ is gamma with shape parameter $\beta + 1$ for $x_2 > x_1$, which is an increasing failure rate distribution. (Note that $f_2(x_2; t)$ is truncated gamma which is not free from t.)

4.3. Bivariate Gumbel model

Consider X_1 and X_2 with the bivariate Gumbel distribution [22] with joint density

$$f(x_1, x_2) = [(1 + \delta x_1)(1 + \delta x_2) - \delta] e^{-x_1 - x_2 - \delta x_1 x_2}, \quad x_1, x_2 > 0, \ \ 0 \leqslant \delta \leqslant 1.$$

The joint residual entropy for this density is not available in closed form. However, the results of the preceding section allow us to learn about the behavior of $H(X_1, X_2; t_1, t_2)$. The information properties of this distribution are as follows:

(a) $H(X_1, X_2; t_1, t_2) \leqslant 2$ and is decreasing in t_j, $j = 1, 2$.
(b) $H(X_j; t_1, t_2) = 1 - \log(1 + \delta t_i)$, which is decreasing in t_i, $i \neq j = 1, 2$.
(c) $H(X_j; t_1, t_2) \leqslant H(X_j; t_j) = 1$.

Property (a) is obtained from Theorem 2. Since $\lambda_{j|i}(x_j; t_1 + \Delta, t_2) \geqslant \lambda_{j|i}(x_j; t_1, t_2)$ for $x_j > 0$, it follows from [18] that $H(X_j|x_i; t_1, t_2) \geqslant H(X_j|x_i; t_1 + \Delta, t_2)$ for every $\Delta > 0$. Because of this and the exponentiality of the marginals, both assumptions of Theorem 2 are satisfied. Property (b) is obtained directly

Property (c) is noted from (b). The bound is attainable when marginal distributions are independent exponential variables. We also have $H(X_j; t_i) - H(X_j; t_1, t_2) = \log(1 + \delta t_i)$, $i \neq j$, which is increasing in t_i. That is, under the Gumbel model, the age of component i becomes progressively more informative about the residual life distribution of the other component j, $i \neq j$.

Thus, unlike the ACBED case, for the bivariate Gumbel model when a component is considered singly, there is no age effect on informativeness of its residual lifetime distribution. However, when the ages of both components are taken into account, each component's residual lifetime distribution and the joint residual lifetime distribution of the two components become more informative.

As a final note, despite the fact that the entropy of conditional distributions $H(X_j|x_i; t_1, t_2)$, $i \neq j$, are decreasing in $x_i > t_i$, analytical results for the conditional entropy $H(X_j|X_i; t_1, t_2)$, $i \neq j$, and mutual information are not available. Numerical integration, however, suggests that these information functions are decreasing in t_1 and t_2.

4.4. BEC model

The standard form (scale of $X_j = 1$, $j = 1, 2$) of the joint density of the BEC family of distributions of Arnold and Strauss [4] is given by

$$f(x_1, x_2) = C(\delta) e^{-x_1 - x_2 - \delta x_1 x_2}, \quad x_1, x_2 > 0, \ \ \delta \geqslant 0, \tag{27}$$

where $C(\delta) = \left[\int_0^\infty (1 + \delta s)^{-1} e^s \, ds\right]^{-1}$. The conditional distribution $f_{j|i}$ of X_j given $X_i = x_i > 0$ is exponential with hazard rate function $\lambda_{j|i}(x_i) = \delta x_i + 1$, $i \neq j$ $j = 1, 2$.

The normalizing constant $C(\delta)$ must be evaluated numerically, therefore the joint and marginal residual dynamic entropies are not available in closed form. However, the results of the preceding section allow us to determine following information properties of this distribution.

(a) $H(X_1, X_2; t_1, t_2)$ is decreasing in t_j.
(b) $H(X_j|X_i; t_1, t_2)$, $i \neq j$, is increasing in t_i.
(c) $H(X_j; t_1, t_2)$, $j = 1, 2$, is decreasing in t_j, $j = 1, 2$.
(d) $M(X_1, X_2; t_1, t_2)$ is decreasing in t_j, $j = 1, 2$.
(e) $H(X_j; t_j)$ is decreasing in t_j.

Property (a) is obtained from Theorem 2 as follows. For every (x_1, y_1) and (x_2, y_2) such that $x_1 < x_2$ and $y_1 < y_2$, for the density (27) we have $f(x_1, y_1) f(x_2, y_2) - f(x_1, y_2) f(x_2, y_1) \leqslant 0$. A result of [16] implies that $P(X_j > x_j + \Delta | X_1 > t_1, X_2 > t_2)$ is decreasing in x_j and therefore $\lambda_j(x_j; t_1, t_2)$, $j = 1, 2$, is increasing in x_j. Hence, both assumptions of Theorem 2 are satisfied. Property (b) is obtained from the facts that the conditionals are exponential with hazard rates $\lambda_{j|i}(x_i) = \delta x_i + 1$, $i \neq j$, so $H(X_j|x_i; t_1, t_2) = -\delta x_i$ for $x_i > t_i$. Using this in (7) we find the result. Property (c) follows from using (a) and (b) in (8) and (d) follows from using (a) and (b) in (14). Property (e) follows from (a) when $t_i = 0$.

4.5. Bivariate gamma

Consider the bivariate gamma distribution with the following density:

$$f(x_1, x_2) = \frac{e^{-(x_1+x_2)}}{\Gamma(\theta_0)\Gamma(\theta_1)\Gamma(\theta_2)} \int_0^{\tilde{x}} y_0^{\theta_0-1}(x_1 - y_0)^{\theta_1-1}(x_2 - y_0)^{\theta_2-1} e^{y_0}\, dy_0,$$
$$x_1, x_2 \geqslant 0, \quad \theta_0, \theta_1, \quad \theta_2 > 0, \tag{28}$$

where $X_j = Y_0 + Y_j,\ j = 1, 2, Y_i,\ i = 0, 1, 2$, are distributed as gamma with shape parameters θ_i and common scale 1, and $\tilde{x} = \min(x_1, x_2)$. The marginal distribution of $X_j,\ j = 1, 2$, is gamma with shape parameter $\theta_0 + \theta_j$ and scale 1.

Kotz et al. [25] give an expression for the density when $\theta_1 = \theta_2 = 1$ and θ_0 an integer. In general, for integer values of $\theta_i,\ i = 0, 1, 2$, the integral in (28) can be evaluated recursively. For example, when $\theta_1 = \theta_2 = 2$ and θ_0 an integer, we obtain the density

$$\begin{aligned} f(x_1, x_2) = \frac{e^{-(x_1+x_2)+\tilde{x}}}{(\theta_0 - 1)!} \Bigg[& \tilde{x}^{\theta_0+1} - (x_1 + x_2 + \theta_0 + 1)\tilde{x}^{\theta_0} \\ & + (-1)^{\theta_0}(\theta_0 - 1)!\{x_1 x_2 + \theta_0(x_1 + x_2 + \theta_0 + 1)\} \\ & \times \left(e^{-\tilde{x}} - \sum_{k=0}^{\theta_0-1} \frac{(-1)^k \tilde{x}^k}{k!} \right) \Bigg], \quad x_1, x_2 \geqslant 0. \end{aligned} \tag{29}$$

Closed form expressions for dynamic entropies of (29) are not available. However, for any $\theta_0 = 1, 2, \cdots$ the information functions can be obtained by numerical integration.

Fig. 2a shows the plots of residual entropies $H(X_1, X_2; t, t),\ H(X_j; t, t),\ H(X_j|X_i; t, t)$, $H(X_j; t)$, and the total marginal residual entropy $H(X_1; t, t) + H(X_2; t, t)$ of (29) for $\theta_0 = 2$. Note that X_1 and X_2 are exchangeable. Fig. 2b shows the plot of residual mutual information $M(X_1, X_2; t, t)$. These plots suggest the following dynamic behaviors.

(a) The joint and marginal distributions and, on average, the conditional distributions progressively become more concentrated as the components age.
(b) $H(X_j; t) < H(X_j; t, t)$ for all $t > 0$. Thus under this model considering both ages increases uncertainty about the residual life of a component, compared with the univariate case. Furthermore, the gap between $H(X_j; t, t)$ and $H(X_j; t)$ increases with t, indicating that consideration of both ages induces progressively more uncertainty about the residual lifetime of component j.
(c) The joint residual entropy (solid) decreases with a slower rate than the total marginal residual entropy (dashed-double dotted curve). The same phenomenon can be seen by comparing the rates of decrease of the conditional residual entropy (dotted) and the marginal residual entropy (dashed). Thus, by (13) and (14) $M(X_1, X_2; t, t)$ decreases in t. These plots also illustrate the inequalities in (15).
(d) The mutual information plot reveals that dependence between the lifetime of components reduces as they age and that the lifetimes eventually become independent of each other.

Plots of joint residual entropy and residual mutual information of (29) for $\theta_0 = 1, 2, 3$ (not shown here) indicated that all these information functions are decreasing in t. These plots also

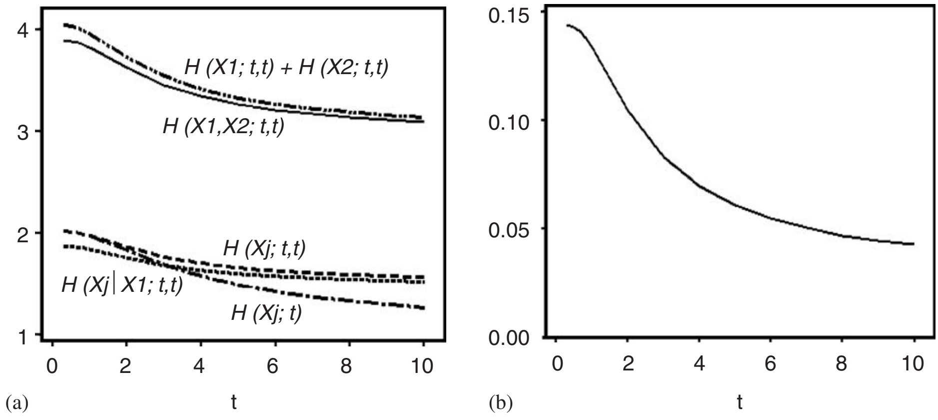

Fig. 2. Residual entropies and residual mutual information for the bivariate gamma distribution of Example 1 when $\theta_i = 2,\ i = 0, 1, 2$.

indicated that the joint entropy and mutual information are ordered by θ_0. Similar behaviors were shown for $\theta_1 = \theta_2 = 1$ and $\theta_0 = 1, 2, 3$.

5. Multivariate information

For the case of more than two variables, extensions of the dynamic joint, marginal, and conditional entropies are straightforward. Let $\boldsymbol{X} = (X_1, \ldots, X_d)$ and $\boldsymbol{t} = (t_1, \ldots, t_d)$. Then $K(f : g; \boldsymbol{t})$ and $H(\boldsymbol{X}; \boldsymbol{t})$ are given by (1) and (2) with the double integrals replaced by d-fold integrals. The marginal residual entropy of a subvector of length d_a, denoted by $H(\boldsymbol{X}_a; \boldsymbol{t})$, is defined by the d_a-fold extension of (4) where the marginal residual density of $\boldsymbol{X}_a$, given $X_j > t_j,\ j = 1, \ldots, d$, is found by the $(d - d_a)$-dimensional counterpart of (5). The conditional residual density of $\boldsymbol{X}_a$, given a subvector $\boldsymbol{X}_b$, is defined by extension of (6) to $f_{a|b}(\boldsymbol{x}_a|\boldsymbol{x}_b; \mathbf{t})$ and the conditional residual entropy $H(\boldsymbol{X}_a|\boldsymbol{X}_b; \boldsymbol{t})$ is obtained by extension of (7) accordingly. The joint residual entropy can be decomposed in terms of $H(\boldsymbol{X}_a; \boldsymbol{t})$ and $H(\boldsymbol{X}_b|\boldsymbol{X}_a; \boldsymbol{t})$. A generalization of (8) provides a chain rule decomposition,

$$H(\boldsymbol{X}; \boldsymbol{t}) = \sum_{j=1}^{d} H(X_j|X_1, \ldots, X_{j-1}; \boldsymbol{t}), \tag{30}$$

where $H(X_j|X_1, \ldots, X_{j-1}; \boldsymbol{t})$ is the conditional residual entropy of X_j given $X_1, \ldots, X_{j-1}$, $X_j > t_j$, $j = 1, \ldots, d$ and $H(X_1|X_0) = H(X_1)$.

In the multivariate case, various mutual information functions can be defined for measuring various dependencies between the components. The residual mutual information between two subvectors $\boldsymbol{X}_a$ and $\boldsymbol{X}_b$ has Kullback–Leibler and entropy representations (12)–(14) where i and j are replaced with a and b. For example, consider the trivariate case which is sufficiently general. The mutual information between the residual lifetime of a component, say X_3, and the lifetimes of the other two components jointly is given by

$$M(X_3, [X_1, X_2]; \boldsymbol{t}) = K\left(f(\boldsymbol{x}; \boldsymbol{t}) : f_3(x_3; \boldsymbol{t}) f_{1,2}(x_1, x_2; \boldsymbol{t})\right) \tag{31}$$

$$= H(X_3; \boldsymbol{t}) - H(X_3|X_1, X_2; \boldsymbol{t}) \tag{32}$$

$$= H(X_1, X_2; \boldsymbol{t}) - H(X_1, X_2|X_3; \boldsymbol{t}) \tag{33}$$
$$= M(X_3, X_1; \boldsymbol{t}) + M(X_3, X_2|X_1; \boldsymbol{t}), \tag{34}$$

where $f_{1,2}(x_1, x_2; \boldsymbol{t})$ is the bivariate marginal residual density, $H(X_3; \boldsymbol{t})$, $H(X_1, X_2; \boldsymbol{t})$, $H(X_3|X_1, X_2; \boldsymbol{t})$ and $H(X_1, X_2|X_3; \boldsymbol{t})$ are marginal and conditional residual entropies defined similarly to (7), and $M(X_3, X_2|X_1; \boldsymbol{t})$ is the dynamic partial mutual information measures defined similarly to (31). Note that $M(X_3, X_2|X_1; \boldsymbol{t})$ measures the conditional dependence between the residual lifetimes of components 2 and 3, given the residual lifetime of the first component.

Let $\boldsymbol{X}_a$ contain a single component, say X_d, and $\boldsymbol{X}_b = \boldsymbol{X}_{\bar{d}}$ contain all other components. Then the mutual information has a chain rule decomposition,

$$M(X_d, \boldsymbol{X}_{\bar{d}}; \boldsymbol{t}) = \sum_{j=1}^{d-1} M(X_d, X_j|X_1, \ldots, X_{j-1}; \boldsymbol{t}),$$

where $M(X_d, X_j|X_1, \ldots, X_{j-1}; \boldsymbol{t})$ is the dynamic partial mutual information, and $M(X_d, X_1|X_0) = M(X_d, X_1)$.

Mutual information functions between more than two subvectors can be defined similarly. In particular, the mutual information between the residual lifetimes of all d components is given by

$$\begin{aligned} M(\boldsymbol{X}, \boldsymbol{t}) &= K[f(\boldsymbol{x}; \boldsymbol{t}) : f_1(x_1\boldsymbol{t}) \cdots f_d(x_d; \boldsymbol{t})] \\ &= \sum_{j=1}^{d} H(X_j; \boldsymbol{t}) - H(\boldsymbol{X}; \boldsymbol{t}). \end{aligned} \tag{35}$$

Clearly, $M(\boldsymbol{X}, \boldsymbol{t}) = 0$ if and only if the residual lifetimes of all components are mutually independent. The mutual information (35) measures complexity of the system in terms of the interdependency between the lifetimes of its components, where the simplest system consists of independently distributed components.

5.1. Transformations

The following notion of dependence [23] is used to identify transformations that preserve monotonicity of the joint dynamic entropy.

Definition 1. The random variables $X_1, \ldots, X_d$ are said to be right corner set increasing (RCSI) if $P(X_1 > x_1, \ldots, X_d > x_d|X_1 > t_1, \ldots, X_d > t_d)$ is increasing in $t_1, \ldots, t_d$ for every choice of $x_1, \ldots, x_d$.

See Barlow and Proschan [6] for relationships between RCSI and other notions of dependence.

Theorem 4. *Suppose that* $X_1, \ldots, X_d$ *are RCSI and let* $Y_j = \phi_j(X_j)$, $\tau_j = \phi_j(t_j)$, $j = 1, \ldots, d$, $\boldsymbol{X} = (X_1, \ldots, X_d)$, $\boldsymbol{Y} = (Y_1, \ldots, Y_d)$, *and* $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_d)$.

(a) *If* $\phi_j(x_j)$, $j = 1, \ldots, d$, *are non-negative increasing and concave, and if* $H(\boldsymbol{X}; \boldsymbol{t})$ *is decreasing in* t_j, $j = 1, \ldots, d$, *then* $H(\boldsymbol{Y}; \boldsymbol{\tau})$ *is decreasing in* τ_j, $j = 1, \ldots, d$.
(b) *If* $\phi_j(x_j)$, $j = 1, \ldots, d$, *are non-negative increasing and convex, and if* $H(\boldsymbol{X}; \boldsymbol{t})$ *is increasing in* t_j, $j = 1, \ldots, d$, *then* $H(\boldsymbol{Y}; \boldsymbol{\tau})$ *is increasing in* τ_j, $j = 1, \ldots, d$.

Proof. (a) By (3),

$$H(\boldsymbol{Y};\boldsymbol{\tau}) = H(\boldsymbol{X};\boldsymbol{t}) + \sum_{j=1}^{d} E\left[\log \frac{d}{dX_j}\phi_j^{-1}(Y_j)|(X_1 > t_1, \ldots, X_d > t_d)\right]. \tag{36}$$

Without loss of generality we give the proof for $d = 2$. Since X_1, X_2 are RCSI and $\phi_j(x_j)$, $j = 1, 2$, are increasing, then $E\{\phi_j(X_j)|X_1 > t_1, X_2 > t_2\}$ is increasing in t_j, $j = 1, 2$. It follows from (36) that if the joint entropy of X_1 and X_2 is decreasing in t_1, t_2, then $H(Y_1, Y_2; \tau_1, \tau_2)$ is decreasing in τ_1, τ_2. Proof of (b) is similar. □

Consider the joint residual entropies $H(\boldsymbol{X};\boldsymbol{t})$ and $H(\boldsymbol{Y};\boldsymbol{t})$ when $\boldsymbol{X} = (X_1, \ldots, X_d)$ and $\boldsymbol{Y} = (Y_1, \ldots, Y_d)$ are components of two systems, where $Y_j = \phi_j(X_j)$, $j = 1, \ldots, d$. A question of interest is that under what conditions the transformation increases/decreases the residual entropy. We use the notion of dispersion ordering between random variables (see, e.g., [31]) for this purpose.

Definition 2. The random variable Y_j is said to be larger than X_j in dispersion ordering, denoted as $Y_j \overset{D}{\geqslant} X_j$, if and only if $Y_j = \phi(X_j)$ where ϕ is a "dilation" function; i.e., the condition $\phi(x_j) - \phi(x_j^*) \geqslant x_j - x_j^*$ holds for every $x_j \geqslant x_j^*$.

The next theorem gives sufficient conditions for $H(\boldsymbol{Y};\boldsymbol{t})$ to be more (less) than $H(\boldsymbol{X};\boldsymbol{t})$.

Theorem 5. (a) *If* $Y_j \overset{D}{\geqslant} X_j$, $j = 1, \ldots, d$, *and if* $H(\boldsymbol{X};\boldsymbol{t})$ *is decreasing in* t_j, $j = 1, \ldots, d$, *then* $H(\boldsymbol{Y};\boldsymbol{t}) \geqslant H(\boldsymbol{X};\boldsymbol{t})$.

(b) *If* $Y_j \overset{D}{\leqslant} X_j$, $j = 1, \ldots, d$, *and if* $H(\boldsymbol{X};\boldsymbol{t})$ *is decreasing in* t_j, $j = 1, \ldots, d$, *then* $H(\boldsymbol{Y};\boldsymbol{t}) \leqslant H(\boldsymbol{X};\boldsymbol{t})$.

Proof. (a) Since $Y_j \overset{D}{\geqslant} X_j$, there exists a dilation function such that $Y_j = \phi_j(X_j)$, where ϕ_j is non-negative and differentiable because Y_j is lifetime of a component, and is non-negative and absolutely continuous. The dilation property implies that $\phi_j'(x_j) \geqslant 1$ implies that $\phi_j(t_j) \geqslant t_j$, and $t_j \geqslant \phi_j^{-1}(t_j) = s_j,\ 1, \ldots, d$. Now by (36) $H(\boldsymbol{Y};\boldsymbol{t}) \geqslant H(\boldsymbol{X};\boldsymbol{s})$, where $\boldsymbol{s} = (s_1, \ldots, s_d)$. The result follows from the assumption that $H(\boldsymbol{X};\boldsymbol{t})$ is decreasing.

(b) In this case $X_j = \phi_j^{-1}(Y_j)$, where ϕ_j^{-1} is also a dilation function, and the result follows using similar argument as the proof of part (a). □

Theorem 5 is applicable to the linear transformations. For two systems where the components are related as $Y_j = a_j X_j + b_j, 0 < a_j \leqslant 1$, $j = 1, \ldots, d$, if $H(\boldsymbol{X};\boldsymbol{t})$ is decreasing (increasing) in t_j, $j = 1, \ldots, d$, then $H(\boldsymbol{Y};\boldsymbol{t}) \leqslant H(\boldsymbol{X};\boldsymbol{t})$; the inequality is reversed for $a_j \geqslant 1$. When the linear transformation is applied to a single system, the ages will be transformed $\tau_j = a_j t_j + b_j$, $j = 1, \ldots, d$. Part (a) of Theorem 5 is also applicable to exponential transformation $\phi_j(x_j) = e^{x_j},\ x_j \geqslant 0$, and Part (b) to log transformations, see Section 5.2.

We should emphasize that the residual entropy ordering is not a partial ordering between random vectors since the equality in entropy does not imply the equality in distribution. The relation $H(\boldsymbol{X};\boldsymbol{t}) \leqslant H(\boldsymbol{Y};\boldsymbol{t})$ only indicates that the residual distribution of $\boldsymbol{X}$ is more concentrated (informative) than the residual distribution of $\boldsymbol{Y}$, hence prediction of the remaining lifetimes of $\boldsymbol{Y}$ is more difficult than the remaining lifetimes of $\boldsymbol{X}$.

Another question of interest is that under what conditions an uncertainty ordering $H(\boldsymbol{X}; \boldsymbol{t}) \leqslant H(\boldsymbol{V}; \boldsymbol{t})$ is preserved when the random variables X_j and V_j are transformed to $\phi_j(X_j)$ and $\phi_j(V_j)$, $j = 1, \ldots, d$. An answer will be provided using the notion of joint hazard rate of $(X_1, \ldots, X_d)$ defined as the gradient vector

$$-\nabla \log \bar{F}(x_1, \ldots, x_d) = \left(-\frac{\partial}{\partial x_1}, \ldots, -\frac{\partial}{\partial x_d}\right) \log \bar{F}(x_1, \ldots, x_d).$$

See Johnson and Kotz [24] for a detailed study of vector multivariate hazard rates.

Theorem 6. *Let* $\mathbf{X}$ *and* $\mathbf{V}$ *denote two bivariate random vectors with survival functions* $\bar{F}_{\mathbf{X}}$ *and* $\bar{F}_{\mathbf{V}}$. *Let* $Y_j = \phi_j(X_j)$ *and* $W_j = \phi_j(V_j)$, $j = 1, \ldots, d$, *where* ϕ_j, $j = 1, \ldots, d$, *are nonnegative convex and increasing. If* (a) $-\nabla \log \bar{F}_{\mathbf{X}}(s_1, \ldots, s_d) \leqslant -\nabla \log \bar{F}_{\mathbf{V}}(s_1, \ldots, s_d)$ for all $(s_1, \ldots, s_d)$ *and* (b) $H(\boldsymbol{X}; \boldsymbol{t}) \leqslant H(\boldsymbol{V}; \boldsymbol{t})$, *then* $H(\boldsymbol{Y}; \boldsymbol{\tau}) \leqslant H(\boldsymbol{W}; \boldsymbol{\tau})$, *where* $\tau_j = \phi_j^{-1}(t_j)$, $j = 1, \ldots, d$.

Proof. By (3),

$$\begin{aligned}
&H(\boldsymbol{W}; \boldsymbol{\tau}) - H(\boldsymbol{Y}; \boldsymbol{\tau}) \\
&\quad = H(\boldsymbol{V}; \boldsymbol{t}) - H(\boldsymbol{X}; \boldsymbol{t}) \\
&\qquad + \sum_{j=1}^{d} \left\{ E\left[\log \frac{d}{dV_j} \phi_j^{-1}(W_j) | (W_1 > t_1, \ldots, W_d > t_d)\right]\right. \\
&\qquad \left. - E\left[\log \frac{d}{dX_j} \phi_j^{-1}(Y_j) | (X_1 > t_1, \ldots, X_d > t_d)\right]\right\}.
\end{aligned}$$

When condition (a) holds, the conditional distribution of W_j given $W_1 > s_1, \ldots, W_d > s_d$ is stochastically larger than that of X_j given $X_1 > s_1, \ldots, X_d > s_d$. It follows that for non-negative convex and increasing ϕ_j, $j = 1, \ldots, d$, the quantity enclosed by the braces is non-negative. The theorem then follows from using condition (b). □

5.2. Multivariate Pareto and related distributions

Multivariate Pareto families, in addition to having applications in many fields [3], have a major role in derivations of entropies and mutual information of some other important multivariate distributions that can be obtained by coordinate-wise transformations of a random vector of Pareto variables [10].

Let $\boldsymbol{X}$ have the multivariate Pareto I distribution, $PI(0, \alpha, d)$, with joint density

$$f(\boldsymbol{x}) = \prod_{j=1}^{d} (\alpha + j - 1) \left(1 + \sum_{j=1}^{d} x_j\right)^{-(\alpha+d)}, \quad x_j > 0, \ \alpha > 0. \tag{37}$$

Table 1
Multivariate families related to Pareto I distribution by transformations

Multivariate family	Survival function	Pareto I transformation
Pareto Type IV	$\bar{F}_{\boldsymbol{Y}}(\boldsymbol{y}) = \left[\sum_{j=1}^{d} y_j^{\delta_j} + 1\right]^{-\alpha}$	$\phi_j(x_j) = x_j^{1/\delta_j}$
Exponential	$\bar{F}_{\boldsymbol{Y}}(\boldsymbol{y}) = \left[\sum_{j=1}^{d} e^{y_j} - d + 1\right]^{-\alpha}$	$\phi_j(x_j) = \log(1 + x_j)$
Weibull	$\bar{F}_{\boldsymbol{Y}}(\boldsymbol{y}) = \left[\sum_{j=1}^{d} \exp\left(y_j^{\delta_j}\right) - d + 1\right]^{-\alpha}$	$\phi_j(x_j) = \log(1 + x_j)^{1/\delta_j}$

The joint entropy of this distribution is given by

$$H(\boldsymbol{X}) = \sum_{j=1}^{d} h_1(\alpha + j - 1) + \sum_{j=1}^{d} \frac{d - j}{\alpha + j - 1}, \tag{38}$$

where $h_1(\cdot)$ is the entropy of univariate Pareto I, $PI(0, \alpha, 1)$, distribution,

$$h_1(\alpha) \equiv H(X_j) = 1 - \log\alpha + \frac{1}{\alpha}. \tag{39}$$

It is easy to see that $H(\boldsymbol{X})$ is decreasing in α and increasing in d.

The mutual information between two subvectors $\boldsymbol{X}_a = (X_1, \ldots, X_{d_a})$ and $\boldsymbol{X}_b = (X_{d_a+1}, \ldots, X_d)$ of dimension d_a and $d_b = d - d_a$ is given by

$$\begin{aligned} M(\boldsymbol{X}_a, \boldsymbol{X}_b) = & \sum_{j=1}^{d_a} h_1(\alpha + j - 1) - \sum_{j=d_b+1}^{d} h_1(\alpha + j - 1) \\ & + \sum_{j=1}^{d_a} \frac{d_a - j}{\alpha + j - 1} + \sum_{j=1}^{d_b} \frac{d_b - j}{\alpha + j - 1} - \sum_{j=1}^{d} \frac{d - j}{\alpha + j - 1}. \end{aligned} \tag{40}$$

The entropy and mutual information expressions are given in [10], but not in the forms of recursive representations (38) and (40).

Multivariate Pareto distributions of Types II–IV can be obtained from (37) by coordinate-wise transformations (see [3]). Several other distributions, including the multivariate exponential and multivariate Weibull shown in Table 1 can be obtained from (37) via coordinate-wise transformations, as well. The entropies of these distributions are related and obtainable from (38). Because of its invariance property, the mutual information for all distributions shown in Table 1 is (40); see [10].

The residual information measures for multivariate Pareto I are as follows:

(a)

$$H(\boldsymbol{X}; \boldsymbol{t}) = H(\boldsymbol{X}) + d\ \log\left(1 + \sum_{j=1}^{d} t_j\right).$$

(b)

$$H(\boldsymbol{X}_a;\boldsymbol{t}) = H(\boldsymbol{X}_a) + a\log\left(1+\sum_{j=1}^{d} t_j\right).$$

(c)

$$H(\boldsymbol{X}_a|\boldsymbol{X}_b;\boldsymbol{t}) = \sum_{j=d_b+1}^{d} h_1(\alpha+j-1) + \sum_{j=1}^{d}\frac{d-j}{\alpha+j-1} - \sum_{j=1}^{d_b}\frac{d_b-j}{\alpha+j-1} + d_a\log\left(1+\sum_{j=1}^{d_b} t_j\right).$$

(d)

$$M(\boldsymbol{X}_a,\boldsymbol{X}_b;\boldsymbol{t}) = M(\boldsymbol{X}_a,\boldsymbol{X}_b).$$

The joint residual entropy in (a) is obtained by noting the following relationship between the residual and the lifetime variables for (37):

$$x_j = \phi(x_j;t_j) = \frac{x_j^* - t_j}{1+\sum_{j=1}^{d} t_j}, \quad x_j^* > t_j. \tag{41}$$

If the distribution of $\boldsymbol{X}$ is $PI(0,\alpha,d)$ with density (37), then the distribution of $\boldsymbol{X}^*$ is $PI(\sum_{j=1}^{d} t_j,\alpha,d)$ and the result follows from (3). The expression in (b) comes from the fact that the marginal distribution of any subvector of $\boldsymbol{X}$ is in the form of (37). These entropies are increasing in t_j. Thus, as the components age, the respective distributions become less informative about the remaining lifetimes. From (a) and (b) we have $H(X_j;t_j) = h_1(\alpha)+\log(1+t_j)$ and $H(X_j;\boldsymbol{t}) = h_1(\alpha)+\log\left(1+\sum_{j=1}^{d} t_j\right)$. Hence

$$H(X_j;t_j) - H(X_j;\boldsymbol{t}) = -\log\left(1+\frac{\sum_{i=1,i\neq j}^{d} t_i}{1+t_j}\right),$$

which is negative and decreasing in t_i.

The conditional residual entropy in (c) is obtained using the decomposition of the joint entropy and taking the difference between (a) and (b).

Property (d) is due to the one-to-one transformation (41). The residual mutual information being free from $\boldsymbol{t}$ is an interesting finding because the level of dependency between the residual lifetimes remains unchanged irrespective of the current ages of the components. This example also indicates that the BLM property is not necessary for the memoryless dependency. Note that the rate of increase of the joint residual entropy in $t_j,\ j=1,\ldots,d$, is exactly d times the rate of increase of each marginal residual entropy $H(X_j;\boldsymbol{t})$.

The information properties of the distributions shown in Table 1 can be obtained from Theorem 3. Pareto distribution (37) is RCSI and the transformations are increasing and concave ($\delta_j > 1$ for the case of Pareto IV). Also by Theorem 5, the log-transformation to exponentiality satisfies the conditions of Part (b) and $H(\boldsymbol{Y};\boldsymbol{t}) \leqslant H(\boldsymbol{X};\boldsymbol{t})$. The mutual information for all these families are the same as (40), which is free from $\boldsymbol{t}$.

6. Concluding remarks

This paper has introduced joint, marginal, and conditional residual entropies and residual mutual information between the components. These information measures are functions of the current ages of all the components, and thus are dynamic. The residual entropies are useful for assessing if the residual distributions become more/less informative (concentrated) as the components age. The residual mutual information is useful for assessing expected information about the lifetime of an individual from the knowledge of the ages of other individuals in the group as they age.

We gave several results for dynamic information measures. The bivariate exponential distribution with independent marginals is characterized in terms of memoryless conditional entropy. McKay bivariate gamma showed that the BLM is not necessary for bivariate information measures being memoryless on the diagonal (t, t). We have explored some interplay between monotonicity of the hazard rate function of marginal residual distributions and time evolution of some dynamic information functions. The monotonicity (increasing or decreasing) behavior of dynamic entropy of a system formed by series of independent and identically distributed components inherits the monotonicity of dynamic entropy of its components. Interestingly, for a system of parallel components, the result holds only when the uncertainty (entropy) decreases over time. Some results on residual entropy under transformations of lifetime variables have been given.

The concepts and results developed in this paper can be extended in several directions. Di Crescenzo and Longobardi [11,12] have studied the univariate past lifetime entropy. Unlike the univariate case, where the support of the lifetime distribution decomposes into the past and residual lifetime, the bivariate support provides four subsets for decomposition of information: a subset that represents the residual lifetimes of both components, a subset that represents the past lifetimes of both components, and two other subsets, one for the residual lifetime of one component and the past lifetime of the other. Dynamic information measures may be derived from the joint distribution $f(x_1, x_2)$ for these subsets, which lead to dynamic decompositions of $K(f : g)$, $H(X_1, X_2)$, and $M(X_1, X_2)$.

The stage is now set for developing multivariate maximum dynamic entropy (MDE) modeling. Multivariate MDE will combine ideas from the maximum entropy characterization of multivariate distributions [35] and the univariate MDE procedure developed by [5]. The MDE can provide new characterizations of known models as well as providing new models based on the partial knowledge about evolutions of various multivariate hazard functions that can be represented in a system of differential inequalities.

Estimation of entropy and mutual information is an active research area, see, e.g., [9,26]. Developing dynamic statistics based on various entropy estimation procedures and studying their properties are promising areas of research. Other topics for future research include developing dynamic versions of Bayesian information measures [28,32,34] and their applications in system design, experimental design, sampling schemes, and minimal information priors. Ebrahimi et al. [19] provide some examples of sample and Bayesian dynamic statistics. Extensions of the present results in terms of various types of multivariate dispersion orderings (see, e.g., [21,31]) and exploring dynamic mutual information in terms of copula are under study by the authors.

References

[1] M. Ahsanullah, M. Habibullah, On the entropies and mutual information of some bivariate continuous distributions used in reliability theory, in: N.P. Jewell, A.C. Kimber, M.L.T. Lee, G.A. Whitemore (Eds.), Lifetime Data: Models in Reliability and Survival Analysis, Kluwer Academic Publishers, Dordrecht, 1996, pp. 1–5.

[2] J.E. Anderson, T.A. Louis, N.V. Holm, B. Harvald, Time-dependent association measures for bivariate survival distributions, J. Amer. Statist. Assoc. 87 (1992) 641–650.
[3] B.C. Arnold, Pareto Distributions, International Co-operative Publication House, Fairland, MD, 1983.
[4] B.C. Arnold, D. Strauss, Bivariate distribution with exponential conditionals, J. Amer. Statist. Assoc. 83 (1988) 522–527.
[5] M. Asadi, N. Ebrahimi, G.G. Hamedani, E.S. Soofi, Maximum dynamic entropy models, J. Appl. Probab. 41 (2004) 379–390.
[6] R.E. Barlow, F. Proschan, Statistical Theory of Reliability and Life Testing: Probability Models to Begin With, Silver Spring, MD, 1981.
[7] F. Belzunce, J. Navarro, J. M. Ruiz, Some results on residual entropy functions, Metrika (2004) 147–161.
[8] H.B. Block, A.P. Basu, A continuous bivariate exponential extension, J. Amer. Statist. Assoc. 69 (1974) 1031–1037.
[9] G.A. Darbellay, An estimator of the mutual information based on a criterion for independence, Comput. Statist. Data Anal. 32 (2000) 1–17.
[10] G.A. Darbellay, I. Vajda, Entropy expressions for multivariate continuous distributions, IEEE Trans. Inform. Theory, IT 46 (2000) 709–712.
[11] A. Di Crescenzo, M. Longobardi, Entropy-based measure of uncertainty in past lifetime distributions, J. Appl. Probab. 39 (2002) 434–440.
[12] A. Di Crescenzo, M. Longobardi, A measure of discrimination between past life-time distributions, Statist. Probab. Lett. 67 (2004) 173–182.
[13] N. Ebrahimi, How to measure uncertainty in the residual lifetime distributions, Sankhya A 58 (1996) 48–57.
[14] N. Ebrahimi, Testing for exponentiality of the residual lifetime based on dynamic Kullback–Leibler information, IEEE Trans. Reliability R-47 (1998) 197–201.
[15] N. Ebrahimi, Testing for uniformity of the residual lifetime based on dynamic Kullback–Leibler information, Ann. Inst. Statist. Math. 53 (2001) 325–337.
[16] N. Ebrahimi, M. Ghosh, Multivariate negative dependence, Comm. Statist. A10 (1981) 307–337.
[17] N. Ebrahimi, S.N.U.A. Kirmani, A characterization of the proportional hazards model through a measure of discrimination between two residual life distributions, Biometrika 83 (1996) 233–235.
[18] N. Ebrahimi, S.N.U.A. Kirmani, Some results on ordering of survival functions through uncertainty, Statist. Probab. Lett. 29 (1996) 167–176.
[19] N. Ebrahimi, S.N.U.A. Kirmani, E.S. Soofi, Multivariate Sample and Bayesian dynamic information measures, Technical Report, 2005.
[20] N. Ebrahimi, E.S. Soofi, Information measures for reliability, in: R. Soyer, T.A. Mazzuchi, N.D. Singpurwalla (Eds.), Mathematical Reliability: An Expository Perspective, Kluwer, The Netherlands, 2004, pp. 127–159.
[21] A. Giovagnoli, H.P. Wynn, Multivariate dispersion ordering, Statist. Probab. Lett. 22 (1995) 325–332.
[22] E.G. Gumbel, Bivariate exponential distributions, J. Amer. Statist. Assoc. 55 (1960) 698–707.
[23] R. Harris, A multivariate definition for increasing hazard rate distributions, Ann. Inst. Statist. Math. 41 (1970) 713–717.
[24] N.L. Johnson, S. Kotz, A vector multivariate hazard rate, J. Multivariate Anal. 5 (1975) 53–66.
[25] S. Kotz, N. Balakrishnan, N.L. Johnson, Continuous Multivariate Distributions, vol. 1: Models and Applications, second ed., Wiley, New York, 2000.
[26] A. Kraskov, H. Stögbauer, P. Grassberger, Estimating mutual information, Phys. Rev. E 69 (2004) 066138.
[27] S. Kullback, Information Theory and Statistics, Wiley, New York, 1959 (reprinted in 1968 by Dover).
[28] D.V. Lindley, On a measure of information provided by an experiment, Ann. Math. Statist. 27 (1956) 986–1005.
[29] A.W. Marshall, I. Olkin, A multivariate exponential distribution, J. Amer. Statist. Assoc. 62 (1967) 30–44.
[30] A.T. McKay, Sampling from batches, J. Roy. Statist. Soc., Suppl. 1 (1934) 207–216.
[31] M. Shaked, J.G. Shanthikumar, Stochastic Orders and Their Applications, Academic press, New York, 1994.
[32] N.D. Singpurwalla, Entropy and information in reliability, in: D. Berry, K. Chaloner, J. Geweke (Eds.), Bayesian Analysis of Statistics and Econometrics, Wiley, New York, 1997.
[33] E.S. Soofi, Capturing the intangible concept of information, J. Amer. Statist. Assoc. 89 (1994) 1243–1254.
[34] A. Zellner, Bayesian Analysis in Econometrics and Statistics: The Zellner View and Papers, Edward Elgar, Cheltenham UK & Lyme, US, 1997.
[35] K. Zografos, On maximum entropy characterization of Pearson's Type II and VII multivariate distributions, J. Multivariate Anal. 71 (1999) 67–75.