مجله‌ی پژوهش‌های آماری ایران
J. Statist. Res. Iran 3 (2006): 91-101
دوره‌ی ۳، شماره‌ی ۱، بهار و تابستان ۱۳۸۵، صص ۹۱-۱۰۱
پژوهشکده‌ی آمار

On Classification of Bivariate Distributions Based on Mutual Information

Mohamed Habibullah[†] and Mohammad Ahsanullah[‡,*]

[†] Northeastern University
[‡] Rider University

Abstract. Among all measures of independence between random variables, mutual information is the only one that is based on information theory. Mutual information takes into account of all kinds of dependencies between variables, i.e., both the linear and non-linear dependencies. In this paper we have classified some well-known bivariate distributions into two classes of distributions based on their mutual information. The distributions within each class have the same mutual information. These distributions have been used extensively as survival distributions of two component systems in reliability theory.

Keywords. mutual information; entropy; survival distribution; bivariate distributions.

1 Introduction

In reliability theory, a number of bivariate and multivariate distributions have been developed in order to model the probability distributions of two component systems. Among these distributions are bivariate extreme value distribution (Tawn, 1988), Freund's (1961) bivariate distribution of X and Y where failure of one component changes the distribution of the other component, Gumbel's (1961) bivariate logistic distributions, Hougaard's (1986) class

* Corresponding author

of multivariate failure time distributions, Marshall and Olkin's (1967) bivariate exponential distribution. Hutchinson and Lai (1991) and Mardia (1970) provided a review of a large number of such distributions that are used in reliability theory.

Although information theory and reliability theory have been studied extensively and almost independently in the literature, many authors such as El-Sayeed (1969), Evans (1969), Soofi et al. (1995), Teitler et al. (1986), Tribus (1962) among others have used information theoretic measures in reliability analysis. In this paper we studied the mutual information of several bivariate distributions and observed that some bivariate distributions have identical mutual information. This led to the formation of some classes of bivariate distributions based on common mutual information. In section 2, we briefly explained the concept of entropy and mutual information of bivariate distributions. In section 3, we proposed two classes of distributions based on general expression of mutual information for different bivariate distributions.

2 Measures of Entropy and Mutual Information

In information theory, the uncertainty associated with the distribution of a random variable X is measured by the entropy

$$H(X) = -\int_{-\infty}^{\infty} f(x) \log f(x)\ dx, \qquad \text{if } X \text{ is continuous.}$$

If the random variable X is discrete with a probability mass function $p(x)$ in the sample space S, then the entropy $H(x)$ is given by

$$H(p) = -\sum_{x \in S} p(x) \log p(x),$$

and is equivalent to the amount of information required on the average to describe the random variable X.

For bivariate continuous density function $f(x, y)$ of the random variables X and Y, the entropies of the joint and the marginal densities are

$$H_{12}(X, Y) = -\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x, y) \ln f(x, y)\ dx\ dy,$$

$$H_1(X) = -\int_{-\infty}^{\infty} f_1(x) \ln f_1(x)\ dx,$$

and

$$H_2(Y) = -\int_{-\infty}^{\infty} f_2(y) \ln f_2(y)\ dy,$$

where $f_1(x)$ and $f_2(y)$ are the marginal densities of the random variables X and Y respectively.

The mutual information $I(X,Y)$ of the random variable X and Y is defined by

$$I(X,Y) = \iint_{R^2} f(x,y) \ln \frac{f(x,y)}{f_1(x)\, f_2(y)}\, dxdy$$
$$= H_1(X) + H_2(Y) - H_{12}(X,Y).$$

$I(X,Y) \geqslant 0$ and X and Y are independent if and only if $I(X,Y) = 0$.

We have studied the entropies and the mutual information of the following bivariate distributions with joint survival function $\bar{F}(x,y)$.

1. Bivariate Weibull distribution (Mardia, 1970):

$$\bar{F}(x,y) = \exp\left\{-\left[\left(\frac{x}{\theta_1}\right)^{\frac{\beta_1}{c}} + \left(\frac{y}{\theta_2}\right)^{\frac{\beta_2}{c}}\right]^c\right\},$$

where $x, y > 0$; $\theta_1, \theta_2, \beta_1, \beta_2 > 0$; $0 < c \leqslant 1$.

2. Bivariate extreme value distribution (Mardia, 1970):

$$\bar{F}(x,y) = \exp\left\{-\left(\exp\left\{\frac{-x}{c}\right\} + \exp\left\{\frac{-y}{c}\right\}\right)^c\right\},$$

where $x, y \geqslant 0$; $0 < c \leqslant 1$.

3. Bivariate Gompertz distribution (Mardia, 1970):

$$\bar{F}(x,y) = \exp\left\{-\left[\{m(a^x - 1)\}^{\frac{1}{c}} + \{m(a^y - 1)\}^{\frac{1}{c}}\right]^c\right\},$$

where $x, y, m > 0$; $a > 1$; $0 < c \leqslant 1$.

4. Bivariate Makeham distribution (Mardia, 1970):

$$\bar{F}(x,y) = \exp\left\{\left[\{Ax + m(a^x - 1)\}^{\frac{1}{c}} + \{By + m(a^y - 1)\}^{\frac{1}{c}}\right]^c\right\},$$

where $x, y > 0$; $0 < c \leqslant 1$; $a > 1$; $A, B, m > 0$.

5. Bivariate copula distribution (Nelson, 2005):

$$\bar{F}(x,y) = \exp\left\{-\left[\{-\ln(1-x)\}^{\frac{1}{c}} + \{-\ln(1-y)\}^{\frac{1}{c}}\right]^c\right\},$$

where $0 < c \leqslant 1$, and x, y are uniform random variables in $(0,1)$.

6. Bivariate Gumbel's logistic distribution (Gumbel, 1961):

$$\bar{F}(x,y) = \frac{1}{(1+\exp\{-x\}+\exp\{-y\})^c}, \qquad -\infty < x,\ y < \infty;\ c > 0,$$

with a joint density function

$$f(x,y) = \frac{c\,(c+1)\exp\{-x-y\}}{(1+\exp\{-x\}+\exp\{-y\})^{c+2}}.$$

7. Bivariate Pareto distribution (Mardia, 1970):

$$\bar{F}(x,y) = \frac{1}{(1+x+y)^c}, \qquad x,y \geqslant 0; \quad c > 0,$$

with joint density function

$$f(x,y) = \frac{c\,(c+1)}{(1+x+y)^{c+2}}.$$

3 Classes of Bivariate Distributions Based on Mutual Information

In this section, we proposed two classes of distributions characterized by two theorems. The members of each class have the same mutual information.

Theorem 1 *Suppose* (X,Y) *are random variables with joint survival function*

$$\bar{F}(x,y) = \exp\left\{-\left[\left\{-\ln \bar{F}_1(x)\right\}^{\frac{1}{c}} + \left\{-\ln \bar{F}_2(y)\right\}^{\frac{1}{c}}\right]^c\right\}, \tag{1}$$

where $x,y \geqslant 0;\ 0 < c \leqslant 1,$ *and* $\bar{F}_1,\ \bar{F}_2$ *are marginal survival functions of* X *and* Y *respectively. Then the mutual information of* (X,Y) *is given by*

$$I(X,Y) = -1 - \Gamma'(1) + c\,\exp\left\{\frac{1}{c}-1\right\}\Gamma'\left(2, \frac{1}{c}-1\right),$$

where $\Gamma'\left(2,\frac{1}{c}-1\right) = \int_{\frac{1}{c}-1}^{\infty} x\ \exp\{-x\}\ln x\,dx$, $\Gamma'(1) =$ *Euler's Constant.*

Proof. We have

$$I(X,Y) = \int_0^\infty \int_0^\infty f(x,y)\ln\frac{f(x,y)}{f_1(x)f_2(y)}\ dxdy$$

and

$$\frac{f(x,y)}{f_1(x)f_2(y)} = \frac{\bar{F}(x,y)}{\bar{F}_1(x)\bar{F}_2(y)}\left[\left\{-\ln \bar{F}_1(x)\right\}^{\frac{1}{c}-1}\left\{-\ln \bar{F}_2(y)\right\}^{\frac{1}{c}-1}\right]$$
$$\times\left[\left\{-\ln \bar{F}_1(x)\right\}^{\frac{1}{c}}+\left\{-\ln \bar{F}_2(y)\right\}^{\frac{1}{c}}\right]^{c-2}$$
$$\times\left(\left[\left\{-\ln \bar{F}_1(x)\right\}^{\frac{1}{c}}+\left\{-\ln \bar{F}_2(y)\right\}^{\frac{1}{c}}\right]^{c}+\frac{1}{c}-1\right).$$

Let $uw^{1/c} = \left\{-\ln \bar{F}_1(x)\right\}^{1/c}$ and $(1-u)\,w^{1/c} = \left\{-\ln \bar{F}_2(y)\right\}^{1/c}$, $0 \leqslant u \leqslant 1$; $w > 0$. This gives $\bar{F}(x,y) = \exp\{-w\}$, $\bar{F}_1(x) = \exp\{-u^c w\}$, $\bar{F}_2(y) = \exp\{-(1-u)^c w\}$, $w = \{(-\ln \bar{F}_1(x))^{1/c} + (-\ln \bar{F}_2(y))^{1/c}\}^c$

$$\frac{\bar{F}(x,y)}{\bar{F}_1(x)\,\bar{F}_2(y)} = \exp\{w\,(u^c + (1-u)^c - 1)\},$$

and after simplification, we get

$$\frac{f(x,y)}{f_1(x)\,f_2(y)} = u^{1-c}(1-u)^{1-c}\,w^{-1}\left(w+\frac{1}{c}-1\right)\exp\left\{w\,(u^c+(1-u)^{c}-1)\right\}$$

and

$$f(x,y)\ dxdy = (1-c+cw)\ \exp\{-w\}\ dudw, \qquad 0 \leqslant u \leqslant 1; \quad w > 0.$$

Substituting these in $I(X,Y)$, we get

$$\begin{aligned} I(X,Y) &= \int_0^\infty \int_0^\infty f(x,y)\ln\frac{f(x,y)}{f_1(x)f_2(y)}\ dxdy \\ &= \int_0^1 \int_0^\infty \ln\left[u^{1-c}(1-u)^{1-c}\,w^{-1}\left(w+\frac{1}{c}-1\right)\right. \\ &\quad \left.\times\exp\{w(\,u^c+(1-u)^c-1)\}\right](1-c+cw)\exp\{-w\}\ dudw \\ &= -1-\Gamma'(1)+c\,\exp\left\{\frac{1}{c}-1\right\}\Gamma'\left(2,\frac{1}{c}-1\right), \qquad 0< c \leqslant 1. \end{aligned}$$

If $c = 1$, then $I(X,Y) = 0$, which means that X and Y are independent. The bivariate distributions whose survival functions are given by (1) will be called to form a class C_1.

There are many such distributions who are members of this class. Some of them are (a) bivariate Weibull distribution, (b) bivariate extreme value distribution, (c) bivariate Gompertz distribution, (d) bivariate Makeham distribution, and (e) bivariate copula distribution.

For example, if we consider

$$\bar{F}_1(x) = \exp\left\{-\left(\frac{x}{\theta_1}\right)^{\beta_1}\right\}, \qquad \bar{F}_2(y) = \exp\left\{-\left(\frac{y}{\theta_2}\right)^{\beta 2}\right\},$$

then (1) will reduce to

$$\bar{F}(x,y) = \exp\left\{-\left(\left(\frac{x}{\theta_1}\right)^{\frac{\beta_1}{c}} + \left(\frac{y}{\theta_2}\right)^{\frac{\beta_2}{c}}\right)^c\right\},$$

which is the bivariate Weibull distribution.

Lemma 1 *Suppose X and Y are two continuous random variables. Then the following result holds.*

$$\begin{aligned} I &= c(c+1)\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \ln\left(1+\exp\{-x\}+\exp\{-y\}\right) \\ &\quad \times \frac{\exp\{-x\}\exp\{-y\}}{(1+\exp\{-x\}+\exp\{-y\})^{c+2}}\,dxdy \\ &= \frac{2c+1}{c(c+1)}. \end{aligned}$$

Proof. Let $u = \exp\{-x\}/(1+\exp\{-y\})$ and $v = \exp\{-y\}$. Then $1 + \exp\{-x\}+\exp\{-y\} = (1+u)(1+v)$ and $\exp\{-x\}\exp\{-y\}\,dxdy = (1+v)\,dudv$, where $u, v \geqslant 0$. Then it follows that

$$\begin{aligned} I &= c(c+2)\int_0^{\infty}\int_0^{\infty}\{\ln(1+u)+\ln(1+v)\}\frac{1+v}{(1+u)^{c+2}(1+v)^{c+2}}\,dudv \\ &= c(c+1)\int_0^{\infty}\int_0^{\infty}\frac{\ln(1+u)}{(1+u)^{c+2}(1+v)^{c+1}}\,dudv \\ &\quad + c(c+1)\int_0^{\infty}\int_0^{\infty}\frac{\ln(1+v)}{(1+u)^{c+2}(1+v)^{c+1}}\,dudv \\ &= (c+1)\int_0^{\infty}\frac{\ln(1+u)}{(1+u)^{c+2}}\,du + c\int_0^{\infty}\frac{\ln(1+v)}{(1+v)^{c+1}}\,dv. \end{aligned}$$

Let $w = \ln(1+u)$ and $z = \ln(1+v)$. Then

$$\begin{aligned} I &= (c+1)\int_0^\infty w\,\exp\{-(c+2)w\}\,\exp\{w\}\,dw \\ &\quad + c\int_0^\infty z\,\exp\{-(c+1)z\}\,\exp\{z\}\,dz \\ &= \frac{1}{c+1} + \frac{1}{c} \\ &= \frac{2c+1}{c(c+1)}. \end{aligned}$$

I tends to zero as c tends to infinity.

Theorem 2 *Let X, Y be continuous random variables with the joint probability density function given by*

$$f(x,y) = \frac{c\,(c+1)\left[g_1(x)\,g_2(y)/\{\bar{G}_1(x)\,\bar{G}_2(y)\}\right]}{\left\{1 - \ln\bar{G}_1(x) - \ln\bar{G}_2(y)\right\}^{c+2}} \tag{2}$$

and joint survival function given by

$$\bar{F}(x,\,y) = \frac{1}{\{1 - \ln\bar{G}_1(x) - \ln\bar{G}_2(y)\}^c}, \qquad c > 0,$$

subject to the condition

$$1 - \ln\bar{G}_1(x) - \ln\bar{G}_2(y) \geqslant 0 \quad \text{and} \quad \bar{G}_i(x) = 1 - G_i(x), \quad i = 1,2,$$

where $G_i(x)$, $i = 1,2$, are distribution functions.

The marginal densities and the survival functions of X and Y are given respectively by

$$f_1(x) = \frac{cg_1(x)/\bar{G}_1(x)}{\{1 - \ln\bar{G}_1(x)\}^{c+1}} \quad \text{and} \quad \bar{F}_1(x) = \frac{1}{\{1 - \ln\bar{G}_1(x)\}^c},$$

$$f_2(y) = \frac{cg_2(y)/\bar{G}_2(y)}{\{1 - \ln\bar{G}_2(y)\}^{c+1}} \quad \text{and} \quad \bar{F}_2(y) = \frac{1}{\{1 - \ln\bar{G}_2(y)\}^c},$$

where $g_i(x) = dG_i(x)/dx$, $i = 1,2$.

Then the mutual information is given by

$$I(X,Y) = \ln\frac{c+1}{c} - \frac{1}{c+1}, \qquad c > 0. \tag{3}$$

Proof. The entropies of the marginal and joint distribution of distributions of X and Y are

$$H_1(x) = -E\{\ln f_1(x)\} = -\ln c - E\left(\ln \frac{g_1(x)}{\bar{G}_1(x)}\right) + I_1,$$

$$H_2(y) = -E\{\ln f_2(y)\} = -\ln c - E\left(\ln \frac{g_2(y)}{\bar{G}_2(y)}\right) + I_2,$$

$$\begin{aligned} -H_{1,2}(x,y) &= E\{\ln f(x,y)\} \\ &= \ln c + \ln(c+1) + E\left(\ln \frac{g_1(x)}{\bar{G}_1(x)}\right) + E\left(\ln \frac{g_2(y)}{\bar{G}_2(y)}\right) - I_3, \end{aligned}$$

where

$$I_1 = (c+1)\int_{-\infty}^{\infty} \ln\{1-\ln \bar{G}_1(x)\} \frac{cg_1(x)/\bar{G}_1(x)}{\{1-\ln \bar{G}_1(x)\}^{c+1}}\ dx,$$

$$I_2 = (c+1)\int_{-\infty}^{\infty} \ln\{1-\ln \bar{G}_2(y)\} \frac{cg_2(y)/\bar{G}_2(y)}{\{1-\ln \bar{G}_2(y)\}^{c+1}}\ dy,$$

$$\begin{aligned} I_3 = (c+2)c(c+1)&\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \ln\left\{1-\ln \bar{G}_1(x) - \ln \bar{G}_2(y)\right\} \\ &\times \frac{\{g_1(x)g_2(y)\}/\{\bar{G}_1(x)\bar{G}_2(y)\}}{\{1-\ln \bar{G}_1(x)-\ln \bar{G}_2(y)\}^{c+2}}\ dxdy. \end{aligned}$$

Letting $\ln\{1-\ln \bar{G}_1(x)\} = \exp\{u\}$, we get

$$\begin{aligned} I_1 &= (c+1)\int_{-\infty}^{\infty} u\ \exp\{u\}\exp\{-(c+1)u\}\ du \\ &= c(c+1)\int_{-\infty}^{\infty} u\ \exp\{-cu\}\ du \\ &= \frac{c+1}{c}. \end{aligned}$$

Similarly, using $\ln\{1-\ln \bar{G}_2(y)\} = \exp\{v\}$, we get

$$\begin{aligned} I_2 &= (c+1)\int_{-\infty}^{\infty} v\ \exp\{v\}\exp\{-(c+1)v\}\ dv \\ &= c(c+1)\int_{-\infty}^{\infty} v\ \exp\{-cv\}\ dv \\ &= \frac{c+1}{c}. \end{aligned}$$

Also putting $-\ln \bar{G}_1(x) = \exp\{-x\}$ and $-\ln \bar{G}_1(y) = \exp\{-y\}$, we can see that

Table 1. Mutual information $I(X, Y)$ as a function of c

c	0.5	1	2	10	100
$I(X, Y)$	1.7853	0.1931	0.0721	0.0044	0.0000

$$\begin{aligned} I_3 &= (c+2)c(c+1)\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \ln\left(1+\exp\{-x\}+\exp\{-y\}\right) \\ &\quad \times \frac{\exp\{-x\}\exp\{-y\}}{(1+\exp\{-x\}+\exp\{-y\})^{c+2}}\, dxdy \\ &= (c+2)\frac{2c+1}{c(c+1)}, \qquad \text{by Lemma 1.} \end{aligned}$$

Thus the mutual information is given by

$$\begin{aligned} I(X, Y) &= H_1(X) + H_2(Y) - H_{12}(X, Y) \\ &= \ln(c+1) - \ln c + \frac{2(c+1)}{c} - (c+2)\frac{2c+1}{c(c+1)} \\ &= \ln\frac{c+1}{c} - \frac{1}{c+1}. \end{aligned}$$

Table 1 gives the values of $I(X, Y)$ for some selected values of c. Thus $I(X, Y)$ is small for large values of c.

The bivariate distributions whose joint probability function are given by (2) will be called to form a class C_2. If we put $\bar{G}_1(x) = \exp\{-\exp(-x)\}$ and $\bar{G}_2(y) = \exp\{-\exp(-y)\}$ in (2), we will get the bivariate Gumbel's distribution, and its mutual information is given by (3). Similarly, if we consider $\bar{G}_1(x) = \exp\{-x\}$ and $\bar{G}_2(y) = \exp\{-y\}$ in (2), we get the bivariate Pareto distribution, and its mutual information is also given by (3).

The authors could not find any reason why the continuous bivariate distribution developed by Block and Basu (1974) does not belong to either of class C_1 or C_2. The entropy and the mutual information of this bivariate distribution was studied by Ahsanullah and Habibullah (1996).

Acknowledgment

The authors thank the referees and the editor for helpful suggestions which resulted in the improved presentation of the paper.

References

Ahsanullah, M.; Habibullah, M. (1996). On the entropies and mutual information of some bivariate continuous distributions used in reliability. In *Lifetime Data: Models in Reliability and Survival Analysis.* N.P. Jewell, A.C. Kimber, M.T. Lee, and G.A. Whitmore, eds., Kluwer, Amsterdam, Netherlands.

Block, H.B.; Basu, A.P. (1974). A continuous bivariate exponential extension. *J. Amer. Statist. Assoc.* **69**, 1031-1037.

Cover, M.; Thomas, J.A. (1991). *Elements of Information Theory.* Wiley, New York.

El-Sayeed, G.M. (1969). Information and sampling from the exponential distribution. *Technometrics* **11**, 41-45.

Evans, R.A. (1969). The principle of minimum information. *IEEE Trans. Reliab.* **R-18**, 87-90.

Gumbel, E.J. (1961). Bivariate logistic distributions. *J. Amer. Statist. Assoc.* **56**, 335-349.

Freund, J.E. (1961). A bivariate extension of the exponential distribution. *J. Amer. Statist. Assoc.* **56**, 971-977.

Hougaard, P. (1986). A class of multivariate failure time distributions. *Biometrika* **73**, 671-678.

Hutchinson, T.P.; Lai, C.D. (1991). *The Engineering Statistician's Guide to Continuous Bivariate Distributions.* Rumsby, Adelaide, Australia.

Mardia, K.V. (1970). *Families of Bivariate Distributions.* Griffin, London.

Marshall, A.W.; Olkin, I. (1967). A mltivariate exponential distribution. *J. Amer. Statist. Assoc.* **62**, 30-44.

Nelson, R.B. (2005). *An Introduction to Copula.* Springer, New York.

Soofi, E.; Ebrahimi, N.; Habibullah, M. (1995). Information distinguishability with application to analysis of failure data. *J. Amer. Statist. Assoc.* **90**, 657-668.

Teitler, S.; Rajagopal, A.K.; Ngai, K.L. (1986). Maximum entropy and reliability distributions. *IEEE Trans. Reliab.* **R-35**, 391-395.

Tawn, J.A. (1988). Bivariate extreme value theory. *Biometrika* **75**, 397-415.

Tribus, M. (1962). The use of maximum entropy estimate in the estimation of reliability. In *Recent Developments in Information and Decision Processes,* R.E. Machol and P. Gray, eds., Macmillan, New York.

Mohamed Habibullah
Department of Information, Operations and Analysis,
Northeastern University,
Boston, MA 02115
e-mail:

Mohammad Ahsanullah
Department of Managment Sciences,
Rider University,
Lawrenceville, NI 08648
e-mail: *ahsan@rider.edu*