Indian J. pure appl. Math., 17 (5) : 581-595, May 1986

ON THE ENTROPIC MEASURES OF STOCHASTIC DEPENDENCE

J. N. KAPUR AND M. DHANDE
Department of Mathematics, Indian Institute of Technology
Kanpur 208016

(*Received* 6 *August* 1985)

The superiority of the entropic measure of stochastic dependence over the correlation measure is demonstrated by showing that for certain distributions including the multivariate Pareto, the multivariate Dirichlet and the multivariate generalized Cauchy distributions, the former measure always gives a meaningful measure of stochastic dependence, while the latter measure does not even exist for some values of the parameters.

1. INTRODUCTION

Kapur[2] recently pointed out the following five weaknesses of the correlation coefficient ρ as a measure of stochastic dependence :

(i) ρ can vanish when the variates are not independent.

(ii) ρ may not be unity even though there may be perfect stochastic dependence between the variates.

(iii) ρ changes for non-linear transformation of variates.

(iv) ρ can be used for measuring dependence between only two variates at a time.

(v) ρ cannot be used for categorical, non-numerical data. We give below a sixth weakness.

(vi) ρ may not exist in some cases.

Kapur[2] also discussed Watanabe's[6] entropic measure of dependence among n variates $x_1, x_2, \ldots, x_n$ defined by

$$D_n = S_1 + S_2 + \ldots + S_n - S \qquad \ldots(1)$$

where S is the entropy of the joint probability distribution of $x_1, x_2, \ldots, x_n$ and S_i is the entropy of the marginal probability distribution of x_i $(i = 1, 2, \ldots, n)$.

It is easily seen that D_n is always $\geqslant 0$ and $D_n = 0$ iff the variates are independent. He showed that the entropic measure does not suffer from any of the first five weaknesses. We show below that in general it does not suffer from the sixth weakness also. At least D_n exists for many more cases than ρ.

The Multivariate Pareto Distribution

Mardia[4] introduced the multivariate Pareto distribution with density function

$$f(x_1, x_2, \ldots, x_n; a; \theta_1, \theta_2, \ldots \theta_n) = \frac{a(a+1)\ldots(a+n-1)}{\theta_1 \cdot \theta_2 \ldots \theta_n} \cdot$$

$$\left(\frac{x_1}{\theta_1} + \ldots + \frac{x_n}{\theta_n} - (n-1)\right)^{-(a+n)} \quad \ldots(2)$$

where

$$x_1 \geqslant \theta_1 > 0, x_2 \geqslant \theta_2 > 0, \ldots x_n \geqslant \theta_n > 0; a > 0. \quad \ldots(3)$$

Johnson and Kotz[1] have also described its properties. In this distribution, the marginal density functions are given by

$$g_i(x_i) = a \cdot \theta_i^a x_i^{-(a+1)}; x_i \geqslant \theta_i > 0; a > 0. \quad \ldots(4)$$

We can now easily prove that

$$E(x_i) = a \cdot \theta_i (a-1)^{-1}, a > 1 \quad \ldots(5)$$

$$E(x_i^2) = a \cdot \theta_i^2 (a-2)^{-1}, a > 2 \quad \ldots(6)$$

$$\text{Var}(x_i) = a \cdot \theta_i^2 (a-1)^{-2}(a-2)^{-1}, a > 2 \quad \ldots(7)$$

$$\text{Cov}(x_i, x_j) = \theta_i \theta_j (a-1)^{-2}(a-2)^{-1}, a > 2 \quad \ldots(8)$$

$$\rho(x_i, x_j) = \frac{1}{a}, a > 2. \quad \ldots(9)$$

Thus all the $n(n-1)/2$ correlation coefficients have the same value $1/a$ and all are independent of $\theta_1, \theta_2, \ldots, \theta_n$. This independence can be alternatively shown hy showing that the density functions for

$$y_1 = x_1/\theta_1, y_2, = x_2/\theta_2, \ldots y_n = x_n/\theta_n \quad \ldots(10)$$

is

$$h(y_1, \ldots, y_n, a) = a(a+1)\ldots(a+n-1) \times (y_1 + y_2 + \ldots + y_n - (n-1))^{-(a+n)} \quad \ldots(11)$$

so that $\rho(y_i, y_j)$ is independent of θ_i, θ_j. Also due to the linear transformation (10) with all θ_i's > 0, we have

$$\rho(y_i, y_j) = \rho(x_i, x_j). \quad \ldots(12)$$

The important fact to be observed however is that neither $\rho(x_i, x_j)$ nor $\rho(y_i, y_j)$ exists when $0 < a \leqslant 2$. It is obvious at the same time that when $0 < a \leqslant 2$, $x_1, x_2, \ldots, x_n$ are mutually dependent, but ρ in this case is not an appropriate measure of stochastic dependence, since it does not exist. We look for a measure of stochastic dependence with does not fail when $0 < a \leqslant 2$.

3. The Entropic Measure of Dependence

For the marginal density function (4)

$$\begin{aligned} S_i &= -\int_{\theta_i}^{\infty} g_i(x_i) \ln g_i(x_i)\, dx_i \\ &= -\int_{\theta_i}^{\infty} g_i(x_i) [\ln a + a \ln \theta_i - (a+1) \ln x_i]\, dx_i \\ &= -\ln a - a \ln \theta_i + (a+1) \int_{\theta_i}^{\infty} g_i(x_i) \ln x_i\, dx_i;\ i = 1, 2, \ldots, n. \end{aligned} \quad \ldots(13)$$

But

$$\int_{\theta_i}^{\infty} x_i^{-(a+1)}\, dx_i = \frac{1}{a}\, \theta_i^{-a}. \quad \ldots(14)$$

Differentiating under the sign of integration with respect to a, we get

$$\int_{\theta_i}^{\infty} x_i^{-(a+1)} \ln x_i\, dx_i = \theta_i^{-a}\left(\frac{1}{a^2} + \frac{1}{a} \ln \theta_i\right) \quad \ldots(15)$$

or

$$\int_{\theta_i}^{\infty} g_i(x_i) \ln x_i\, dx_i = \frac{1}{a} + \ln \theta_i \quad \ldots(16)$$

From (13) and (16)

$$S_l = \ln \theta_l - \ln a + \left(1 + \frac{1}{a}\right). \qquad \ldots(17)$$

Similarly

$$S = -\int \ldots \int f(x_1\, x_2 \ldots x_n) \left[\ln \frac{a(a+1)\ldots(a+n-1)}{\theta_1\, \theta_2 \ldots \theta_n} + \ln\left(\frac{x_1}{\theta_1} + \frac{x_2}{\theta_2} + \ldots + \frac{x_n}{\theta_n} - (n-1)\right)\right]^{-(a+n)} dx_1\, dx_2 \ldots dx_n. \qquad \ldots(18)$$

Using differentiation under the sign of integration and simplifying, we get

$$S = -\sum_{i=1}^{n} \ln(a+i-1) + \sum_{i=1}^{n} \ln \theta_i + (a+n)\left(\frac{1}{a} + \frac{1}{a+1} + \ldots + \frac{1}{a+n-1}\right) \qquad \ldots(19)$$

so that

$$D_n = S_1 + S_2 + \ldots + S_n - S = \left(\ln\left(1 + \frac{1}{a}\right) + \ln\left(1 + \frac{2}{a}\right) + \ldots + \ln\left(1 + \frac{n-1}{a}\right)\right) + n\left(1 + \frac{1}{a}\right) - (a+n)\left(\frac{1}{a} + \frac{1}{a+1} + \ldots + \frac{1}{a+n-1}\right). \qquad \ldots(20)$$

Unlike ρ, D_n exists for all values of $a > 0$ and for all positive integral values $n \geq 2$.

4. Comparison of the Two Measures

(i) ρ is defined for only $n = 2$, while D_n is defined for all positive integral values of $n \geq 2$.

(ii) ρ is defined only when $a > 2$ while D_n is defined for $a > 0$.

(iii) Both ρ and D_n are independent of $\theta_1, \theta_2, \ldots, \theta_n$.

(iv) $$\frac{d\rho}{da} = -\frac{1}{a^2}, \quad \frac{dD_2}{da} = -\frac{1}{a(a+1)^2}. \qquad \ldots(21)$$

Both ρ and D_2 decreases with a, but D_2 decreases at a faster rate. ρ decreases from a value slightly less than $1/2$ when a is slightly greater than $1/2$ and it approaches zero as a approaches infinity (Fig. 1). D_2 is not defined at $a = 0$, but it decreases from an almost infinite value at 0^+, and decreases to 0 as a approaches infinity (Fig. 1).

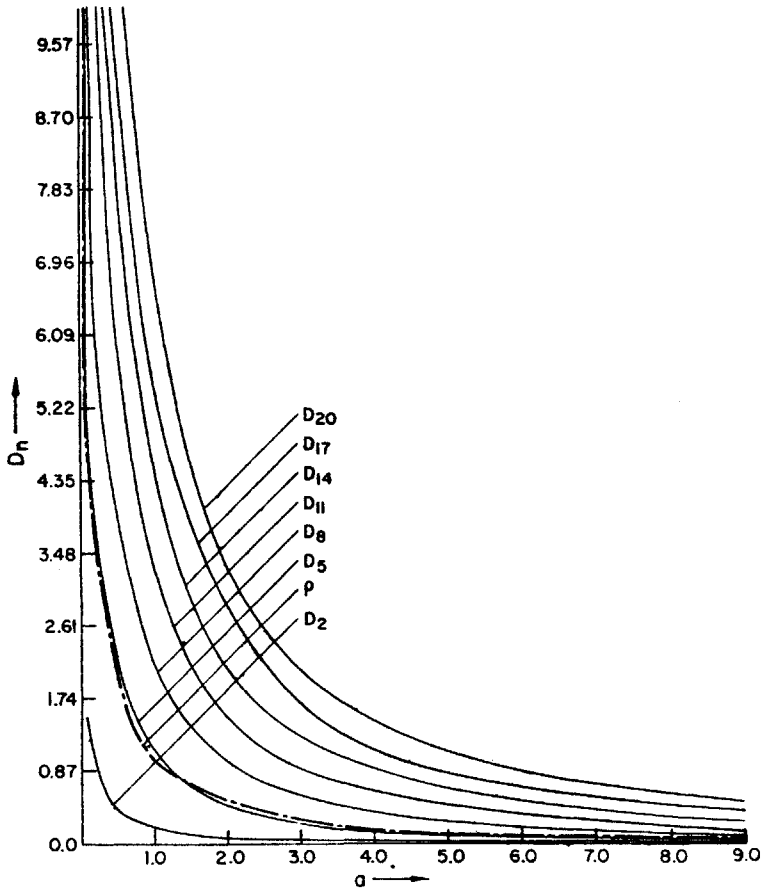

FIG. 1. Graphs of ρ and D_n (n=2,5,8,11,14,17,20) against a.

(v) $$\frac{dD_n}{da} = -\frac{n-1}{a} + (n+a)\left(\frac{1}{(a+1)^2} + \frac{1}{(a+2)^2} + \ldots + \frac{1}{(a+n-1)^2}\right) \quad \ldots(22)$$

As

$$a \to 0, \quad D_n \to \infty, \quad dD_n/da \to \infty. \quad \ldots(23)$$

As

$$a \to \infty, \quad D_n \to 0, \quad dD_n/da \to 0. \quad \ldots(24)$$

The behaviour of D_n is shown in Fig. 1. There is no ρ_n, (when $n > 2$) with which its behaviour can be compared. From Fig. 1 we see that D_n is a decreasing function of a, but is an increasing function of n. As n increases, D_n decreases at a slower rate. D_n is not defined for $a = 0$ for all n. D_n decreases from an almost infinite value at 0^+ and decreases at a fast rate approximately upto $a = 2$, then the rate of decrease decreases and finally D_n approaches zero asymptotically as a approaches infinity.

5. Multivariate Inverted Dirichlet Distribution

Tiao and Guttman[5] obtained the following distribution

$$dP = \frac{\Gamma(m_1+m_2+\ldots+m_{n+1})}{\Gamma(m_1)\ldots\Gamma(m_{n+1})} \cdot \frac{x_1^{m_1-1} x_2^{m_2-1} \ldots x_n^{m_n-1}}{(1+x_1+\ldots+x_n)^{m_1+\ldots+m_{n+1}}} dx_1 \ldots dx_n,$$

$$x_i \geqslant 0;\ i = 1, 2, \ldots, n. \qquad \ldots(25)$$

This distribution is also discussed in Johnson and Kotz[1] It was obtained as a maximum-entropy distribution by Kapur[3].

If we substitute

$$x_i = \frac{u_i}{\theta_i} - 1\ ;\ i = 1, 2, \ldots, n \qquad \ldots(26)$$

in (25), we get

$$dP = \frac{\Gamma(m_1+\ldots+m_{n+1})}{\Gamma(m_1)\ldots\Gamma(m_{n+1})} \frac{\left(\frac{u_1}{\theta_1}-1\right)^{m_1-1} \ldots \left(\frac{u_n}{\theta_n}-1\right)^{m_n-1}}{\left(\frac{u_1}{\theta_1}+\ldots+\frac{u_n}{\theta_1}-(n-1)^{m_1+\ldots+m_{n+1}}\right)}$$

$$\times \frac{1}{\theta_1\theta_2\ldots\theta_n} du_1 \ldots du_n, \qquad \ldots(27)$$

$$u_i \geqslant \theta_i > 0;\ i = 1, 2, \ldots, n.$$

If in (27), we put

$$m_1 = m_2 = \ldots = m_n = 1,\ m_{n+1} = a \qquad \ldots(28)$$

we get

$$dp = \frac{\Gamma(n+a)}{\Gamma(a)} \cdot \frac{1}{\theta_1 \cdot \theta_2 \ldots \theta_n} \left(\frac{u_1}{\theta_1} + \frac{u_2}{\theta_2} + \ldots + \frac{u_n}{\theta_n}(n-1)\right)^{-(a+n)}$$

$$\times\ du_1\, du_2 \ldots du_n,$$

$$u_i \geqslant \theta_i > 0,\ i = 1, 2, \ldots n \qquad \ldots(29)$$

which is the same as the multivariate Pareto distribution.

Thus (27) can be regarded as a generalization of Mardia's multivariate Pareto distribution of type 1. We can also regard multivariate Pareto distribution as a special case of multivariate inverted Dirichlet distribution with shift of origin for variates.

If $X = (x_1, x_2, \ldots, x_n)$ is distributed as a multivariate Dirichlet distribution with parameters $(1, 1, \ldots, 1, a)$, then $U = (u_1, u_2, \ldots, u_n)$ when $u_i = \theta_i (x_i + 1)$, $i = 1, 2, \ldots, n$ is distributed as multivariate Pareto distribution with parameters $\theta_1, \theta_2, \ldots, \theta_n, a$.

This relation between the two multivariate distributions does not appear to have been noted earlier. Even Johnson and Kotz[1] treated them as two independent distributions.

For the distribution (25), we easily verify that

$$E(x_i) = \frac{m_i}{m_{n+1} - 1}, \quad m_{n+1} > 1 \qquad \ldots(30)$$

$$\text{Var}(x_i) = \frac{m_i (m_{n+1} + m_i - 1)}{(m_{n+1} - 1)^2 (m_{n+1} - 2)}, \quad m_{n+1} > 2 \qquad \ldots(31)$$

$$\text{Cov}(x_i, x_j) = \frac{m_i m_j}{(m_{n+1} - 2)(m_{n+1} - 2)}, \quad m_{n+1} > 2 \qquad \ldots(32)$$

$$\rho(x_i, x_j) = \frac{(m_i m_j)^{1/2}}{(m_{n+1} + m_i - 1)^{1/2} (m_{n+1} - m_j - 1)^{1/2}}. \qquad \ldots(33)$$

If $m_1 = m_2 = \ldots = m_n = 1$, $m_{n+1} = a$, this reduces to $1/a$, as expected. Thus for the multivariate Dirichlet ρ_{ij} exists only for $m_{n+1} > 2$.

However the entropic measure of dependence exists for all values of m_{n+1}. In fact

$$S_i = -\ln \frac{\Gamma(m_i + m_{n+1})}{\Gamma(m_i)\,\Gamma(m_{n+1})} - (m_i - 1) \frac{\Gamma'(m_i)}{\Gamma(m_i)}$$

$$+ (m_i + m_{n+1}) \frac{\Gamma'(m_i + m_{n+1})}{\Gamma(m_i + m_{n+1})}$$

(equation continued on p. 588)

$$- (m_{n+1} + 1) \frac{\Gamma'(m_{n+1})}{\Gamma(m_{n+1})} \qquad \ldots(34)$$

from the marginal distribution of the inverted Dirichlet distribution which is

$$g_i(x_i) = \frac{\Gamma(m_i + m_{n+1})}{\Gamma(m_i)\,\Gamma(m_{n+1})} \; x_i^{m_i - 1} (1 + x_i)^{-(m_i + m_{n+1})}. \qquad \ldots(35)$$

Also,

$$S = -\ln \frac{\Gamma(m_1 + \ldots + m_{n+1})}{\Gamma(m_1) \ldots \Gamma(m_{n+1})} - \sum_{i=1}^{n} (m_i - 1) \frac{\Gamma'(m_i)}{\Gamma(m_i)}$$

$$- (m_{n+1} + n) \frac{\Gamma'(m_{n+1})}{\Gamma(m_{n+1})}$$

$$+ (m_1 + \ldots + m_{n+1}) \frac{\Gamma'(m_1 + \ldots + m_{n+1})}{\Gamma(m_1 + \ldots + m_{n+1})} . \qquad \ldots(36)$$

Hence

$$D_n = \sum_{i=1}^{n} S_i - S$$

$$= \ln \frac{\Gamma(m_1 + \ldots + m_{n-1})\,\Gamma^{n+1}(m_{n+1})}{\Gamma(m_1 + m_{n+1}) \ldots \Gamma(m_n + m_{n+1})}$$

$$+ \sum_{i=1}^{n} (m_i + m_{n+1}) \frac{\Gamma'(m_i + m_{n+1})}{\Gamma(m_i + m_{n+1})}$$

$$- (m_1 + m_2 + \ldots + m_{n+1}) \frac{\Gamma'(m_1 + \ldots + m_{n+1})}{\Gamma(m_1 + \ldots + m_{n+1})}$$

$$- (n - 1)\, m_{n+1} \frac{\Gamma'(m_{n+1})}{\Gamma(m_{n+1})} . \qquad \ldots(37)$$

Substituting $m_i = 1$, $i \neq n + 1$ and $m_{n+1} = a$ in (37) we get

$$D_n = \sum_{i=1}^{n} \ln \frac{(a + i - 1)}{a} + \frac{n(a + 1)}{a}$$

$$- (a + n) \left[\frac{1}{a} + \frac{1}{a + 1} + \ldots + \frac{1}{a + n - 1} \right]$$

which is the measure of dependence for the Pareto distribution.

For the special case when

$$m_1 = m_2 = \ldots = m_n = M, \quad m_{n+1} = a$$

$$D_n = \ln \frac{\Gamma(nM + a) \cdot \Gamma^{n-1}(a)}{\Gamma^n(M + a)} + n(M + a) \frac{\Gamma'(M + a)}{\Gamma(M + a)}$$

$$- (nM + a) \frac{\Gamma'(nM + a)}{\Gamma(nM + a)}$$

$$- (n-1)\, a \frac{\Gamma'(a)}{\Gamma(a)} \qquad \ldots(38)$$

and

$$\rho = \frac{M}{a + M - 1}. \qquad \ldots(39)$$

Thus we see that ρ and D_n decrease with a, but ρ_D for the Inverted Dirichlet distribution is less than ρ_P (for the Pareto distribution) if $M < 1$. Similarly

$$(D_n)_D < (D_n)_P \text{ if } M < 1$$

as can be seen from Fig. 2. Also from Figs. 2 and 3, it can be observed that (D_n) is an increasing function of n and M and a decreasing function of a. Hence it can be said that the Inverted Dirichlet distribution behaves in a manner similar to that of the Pareto distribution.

6. Generalised Cauchy Distribution

The joint density function is given by[3]

$$f(x_1, x_2, \ldots, x_n) = \frac{\Gamma(\alpha)}{(\pi)^{n/2}\, \Gamma\left(\alpha - \frac{n}{2}\right)} \left(1 + x_1^2 + \ldots + x_n^2\right)^{-\alpha},$$

$$\alpha > \frac{n}{2} \qquad \ldots(40)$$

$E(x_i)$ exists if $\alpha > (n + 1)/2$. Thus if $\alpha > \frac{n}{2}$ but $< (n + 1)/2$ then the mean does not exist. However in this case because of symmetry of the distribution, by taking

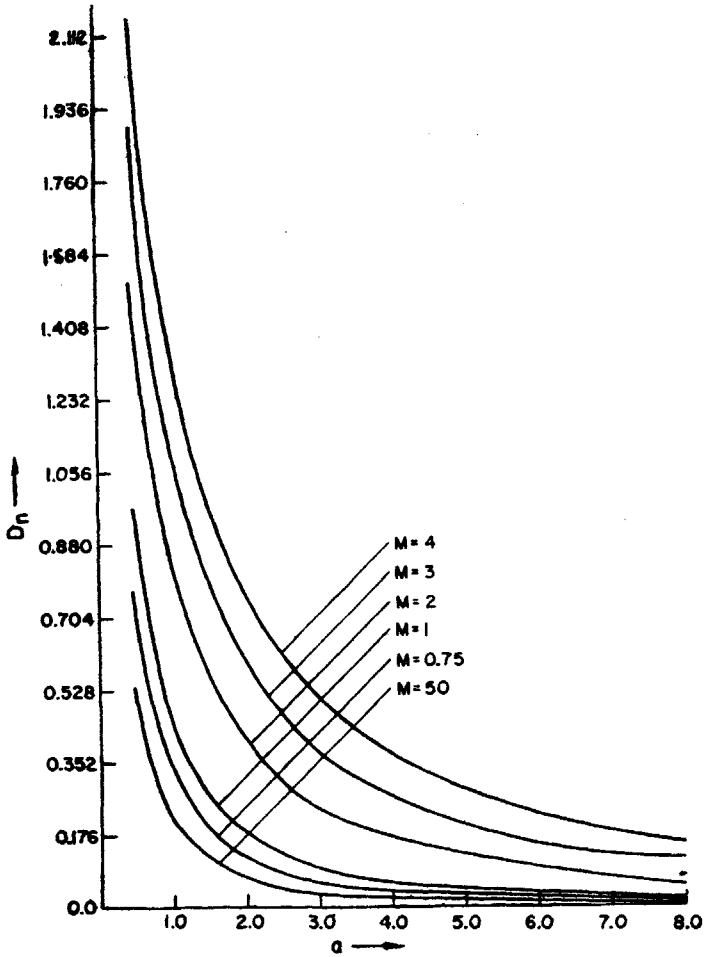

FIG. 2. D_n against a; $(n = 3)$.

Cauchy's principal values of the integral, we can say that in this sense $E(x_i)$ exists and has the value zero.

$E\left(x_i^2\right)$ exists if $\alpha > (n+2)/2$. Thus if $\alpha > n/2$ but $< n/2 + 1$, then $E\left(\sum x_i^2\right)$ does not exist. In this case $E(x_i\, x_j)$ does not exist in the general sense, but if we consider Cauchy principal values of the integral, it exists and its value is zero.

Thus if α lies between $n/2$ and $n/2 + 1$, strictly speaking $\rho(x_i, x_j)$ does not exist. However if we take only principal values of integrals we can say that Cov $(x_i, x_j) = 0$. Further if we are prepared to treat infinity as a number, we can say that Var (x_i) = Var $(x_j) = \infty$, and $\rho(x_i, x_j) = 0$.

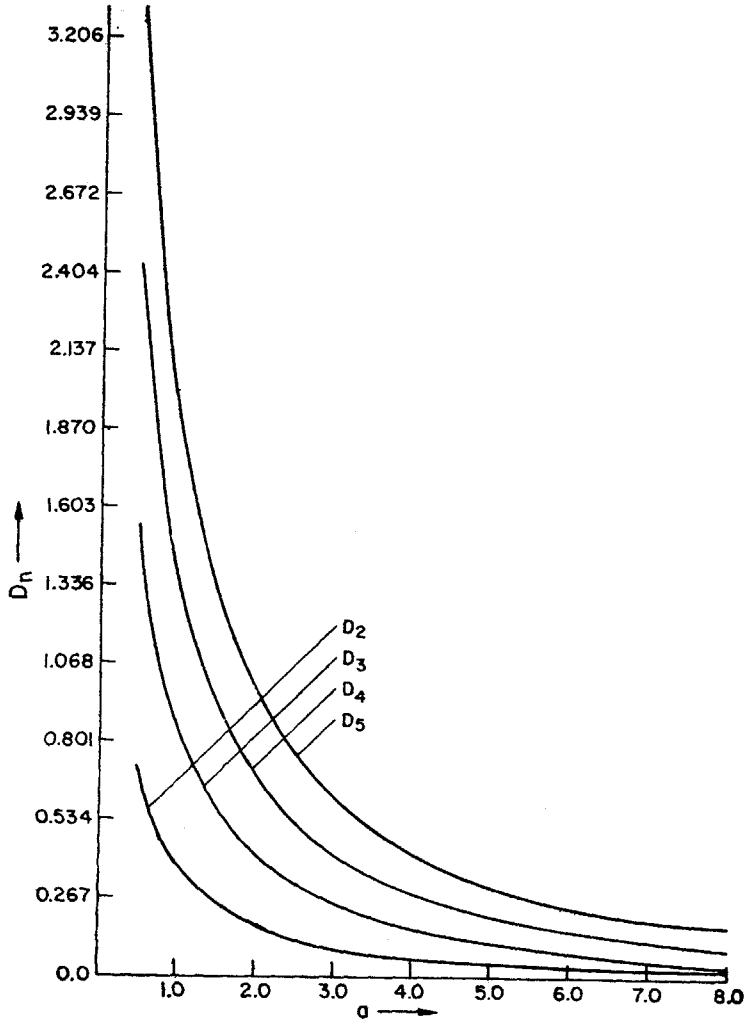

FIG. 3. D_n (n= 2, 3, 4, 5,) against α M = 2.0.

Thus in a relatively loose sense, we can say that all the correlation coefficients are zero. However, even when they all exist are zero, this gives a false picture of dependence since we easily see that the variates following the generalised Cauchy distribution are very much dependent on one another.

The marginal density functions for x_i are

$$g_i(x_i) = \frac{\Gamma\left(\alpha - \frac{n-1}{2}\right)}{\Gamma\left(\alpha - \frac{n-2}{2}\right)} \cdot \frac{1}{\left(1 + x_i^2\right)} \quad \alpha - \frac{n-1}{2}. \qquad \ldots(41)$$

The measure of entropy and stochastic dependence are given by

$$S_I = -\int_{-\infty}^{\infty} g_I(x_I) \ln g_I(x_i)\, dx_I$$

$$= -\ln \frac{\Gamma\left(\alpha - \frac{n-1}{2}\right)}{\sqrt{\pi}\,\Gamma\left(\alpha - \frac{n}{2}\right)}$$

$$- \left(\frac{\Gamma'\left(\alpha - \frac{n}{2}\right)}{\Gamma\left(\alpha - \frac{n}{2}\right)} - \frac{\Gamma'\left(\alpha - \frac{n-1}{2}\right)}{\Gamma\left(\alpha - \frac{n-1}{2}\right)}\right) \cdot \left(\alpha - \frac{n-1}{2}\right) \quad \ldots(42)$$

$$S = -\ln \frac{\Gamma(\alpha)}{(\pi)^{n/2}\,\Gamma\left(\alpha - \frac{n}{2}\right)}$$

$$- \left(\frac{\Gamma'\left(\alpha - \frac{n}{2}\right)}{\Gamma\left(\alpha - \frac{n}{2}\right)} - \frac{\Gamma'(\alpha)}{\Gamma(\alpha)}\right) \cdot \alpha \quad \ldots(43)$$

and

$$D_n = \ln \frac{\Gamma(\alpha)}{\Gamma\left(\alpha - \frac{n}{2}\right)} \frac{\Gamma^n\left(\alpha - \frac{n}{2}\right)}{\Gamma^n\left(\alpha - \frac{n-1}{2}\right)}$$

$$- n \cdot \left(\alpha - \frac{n-1}{2}\right)\left(\frac{\Gamma'\left(\alpha - \frac{n}{2}\right)}{\Gamma\left(\alpha - \frac{n}{2}\right)} - \frac{\Gamma'\left(\alpha - \frac{n-1}{2}\right)}{\Gamma\left(\alpha - \frac{n-1}{2}\right)}\right)$$

$$+ \alpha \cdot \left(\frac{\Gamma'\left(\alpha - \frac{n}{2}\right)}{\Gamma\left(\alpha - \frac{n}{2}\right)} - \frac{\Gamma'(\alpha)}{\Gamma(\alpha)}\right). \quad \ldots(44)$$

This always exists when $\alpha > n/2$.

Thus the entropic measure of dependence always exists, but the correlation measures do not exist in some cases and even when they exist, they give a false idea of the dependence among the variates.

7. Existence of Measure

(*a*) *Existence of Correlation Coefficient* ρ

ρ exists, if $E(x)$, $E(x^2)$, $E(y)$, $E(y^2)$, $E(xy)$ exist and if $\text{Var}(x) \neq 0$, $\text{Var}(y) \neq 0$. We consider non-degenerate distributions only.

Since $\int g_1(x)\,dx$, $\int g_2(y)\,dy$, $\iint f(x, y)\,dx\,dy$ exist all the expected values will exist if the domains are bounded. For unbounded domains, we have to consider the behaviour of the integrands when x and y are large. In particular if $g_i(x) \sim x^r$, then $\int g_i(x)\,dx$ is finite if $r + 1 < 0$. Now $\int x\, g_i(x)\,dx$ will exist if $r + 2 < 0$. and $\int x^2 g_i(x)\,dx$ will exist if $r + 3 < 0$. Thus if $r + 1 < 0$, but $r + 3 > 0$, neither the mean nor the variance exist. If $r + 1 < 0$, $r + 2 < 0$ but $r + 3 > 0$. then mean exists but variance does not exist. This is the situation we have encountered in the Pareto distribution case.

(*b*) *Existence of the Entropy Measure*

Since $S_i = -\int g_i(x) \ln g_i(x)\,dx$, if $|\ln g(x)| < K$, then S_i exists. If $g_i(x)$ can vanish then $\ln g_i(x)$ is not bounded, but $g_i(x) \ln g_i(x)$ still approaches zero as $g_i(x)$ approaches zero. Thus the entropy measure exists if the density functions are bounded. Even if $g_i(x)$ is unbounded, the entropy may still be finite. For example in the case when

$$g_i(x) = \frac{x^{m-1}(1-x)^{n-1}}{B(m, n)}, \quad 0 < x < 1 \qquad \ldots(45)$$

where $0 < m < 1$, $0 < n < 1$, the function $g_i(x)$ is unbounded, but the entropy measure is finite. In the case of Pareto distribution, the density functions are always bounded.

Again if $g_i(x)$ is unbounded but

$$g_i(x) < Ae^{Kx}, \; K > 0$$

then

$$\begin{aligned} S_i &= \left| \int g_i(x) \ln g_i(x)\,dx \right| \\ &\leqslant \int |g_i(x) \ln g_i(x)|\,dx \\ &= \int |g_i(x)|\,|\ln g_i(x)|\,dx \\ &\leqslant \int g_i(x)\,K(x)\,dx. \end{aligned} \qquad \ldots(46)$$

Similarly if

$$g_i(x) < A\,e^{Kx^2}, \; K > 0$$

then

$$S_I < K \int g_I(x)\, x^2\, dx$$

so that if variance exists, the entropy also exists.

(c) *Comparison of the Condition of Existence*

From the above discussion it appears that under very general conditions, the existence of the correlation coefficient will imply the existence of entropic measure, but the converse is not always true.

8. Invariance of ρ and D_n for Transformations

(i) If $x_1 = a_1 u_1 + b_1$, $x_2 = a_2 u_2 + b_2$, $a_1, a_2 > 0$

then

$$\rho(x_1, x_2) = \rho(u_1, u_2) \qquad \ldots(47)$$

$$D_2(x_1, x_2) = D_2(u_1, u_2). \qquad \ldots(48)$$

Both measures are invariant for the above linear transformation.

(ii) If $x_1 = f_1(u_1)$, $x_2 = f_2(u_2)$ $\qquad \ldots(49)$

$f_1(.), f_2(.)$ are non-linear functions then

$$\rho(x_1, x_2) \neq \rho(u_1, u_2) \qquad \ldots(50)$$

$$D_2(x_1, x_2) = D_2(u_1, u_2). \qquad \ldots(51)$$

For non-linear transformations of this type, the entropic measure is invariant but correlation measure is not.

(iii) If $x_1 = f_1(u_1)$, $x_2 = f_2(u_2)$, ..., $x_n = f_n(u_n)$, $\qquad \ldots(52)$

all functions are non-linear

then

$$\rho(x_I, x_J) \neq \rho(u_I, u_J) \qquad \ldots(53)$$

$$D_n(x_1, x_2, \ldots, x_n) = D_n(u_1, u_2, \ldots, u_n). \qquad \ldots(54)$$

For this transformation, no correlation coefficient is invariant, but entropic measure of dependence is invariant.

(iv) If $x_1 = f_1(u_1, u_2, \ldots, u_n)$, $x_2 = f_2(u_1, u_2, \ldots, u_n)$,

$$\ldots x_n = f_n(u_1, u_2, \ldots, u_n) \qquad \ldots(55)$$

then

$$\rho(x_i, x_j) \neq \rho(u_i, u_j) \qquad \ldots(56)$$

and

$$D_n(x_1, x_2, \ldots, x_n) \neq D_n(u_1, u_2, \ldots, u_n). \qquad \ldots(57)$$

Thus according to the entropic measure of dependence, the dependence among $x_1, x_2, \ldots, x_n$ is the same as the dependence among $f_1(x_1), f_2(x_2), \ldots, f_n(x_n)$.

For a discrete random variate if the random variate takes values $x_1, x_2, \ldots, x_n$ with probabilities $p_1, p_2, \ldots, p_n$; then the probability that $f_i(x)$ takes the values $f_i(x_i)$ is also p_i $(i = 1, 2, \ldots, n)$. Thus D_n is invariant for a larger class of transformations than ρ.

References

1. N. L. Johnson, and S. Kotz, *Multi-variate continuous Distributions*. Wiley, New York, 1972.
2. J. N. Kapur, *New measures of Stochastics Dependence*. I. I. T./K. Rep., 1985.
3. J. N. Kapur, *Maximum Entropy Models in Science and Engineering*. South Asia Publishers, New Delhi, 1985.
4. K. V. Mardia, *Ann. Math. Stat.* 33 (1962), 1008–15.
5. G. C. Tiao, and I. Guttman, 'The Inverted Dirichlet distribution with applications, *J. Am. Stat. Assoc.* 60 (1965), 793–805 (correction : 60, 1251–52).
6. S. Wantabe, *Knowing and Guessing*, Wiley, New York, 1969.