ON MEASURES OF ASSOCIATION AND A RELATED PROBLEM

TAKEMI YANAGIMOTO

(Received Feb. 17, 1969; revised Aug. 20, 1969)

Two measures of association $M_1(F)$ and $M_2(F)$ are discussed, which are defined by the expectations of certain rank statistics, T_1 and T_2, respectively. W. Hoeffding [1] has introduced the measure $M_1(F)$ and some of its properties. $M_i(F)$, $i=1, 2$, have desirable properties as the measures of association, for example, $M_i(F)=0$, iff $F(x, y)$ is independent, and $M_i(\Phi\rho)$ is a monotone increasing function of $|\rho|$, when $\Phi\rho$ is the d.f. of two-dimensional normal distribution with correlation coefficient ρ. In Section 2 precise properties are obtained under mild conditions. In Section 3, using these measures, we give a complete result on a relation between equiprobable rankings and independence, which is an improvement of a result by Hoeffding [1].

2. Notation and preliminaries

Let (X, Y) be a bivariate population with the distribution function (d.f.) $F(x, y)$ and its marginal d.f.'s $F_1(x)$ and $F_2(y)$. In what follows $F(x, y)$ is assumed to be continuous. Let $H_F(u, v)$ be

$$H_F(u, v)=F(F_1^{-1}(u), F_2^{-1}(v)) ,$$

where $F_1^{-1}(u)$ and $F_2^{-1}(v)$ are the uth and vth quantiles of $F_1(x)$ and $F_2(y)$, respectively. The marginal d.f.'s of $H_F(u, v)$ are uniform distribution, $U(0, 1)$.

Let $(X_1, Y_1),\cdots, (X_n, Y_n)$ be a sample of size n from (X, Y). Let

$$(x_1, x_2, x_3)=C(x_1-x_2)-C(x_1-x_3) ,$$

where $C(u)=1$ for $u\geqq 0$, and $=0$ for $u<0$. Let

$$\phi(x_1, y_1;\cdots; x_5, y_5)=(1/4)\psi(x_1, x_2, x_3)\psi(x_1, x_4, x_5) \cdot \psi(y_1, y_2, y_3)\psi(y_1, y_4, y_5) ,$$

and

$$\varphi(x_1, y_1; \cdots; x_6, y_6) = (1/4)\phi(x_1, x_3, x_4)\phi(x_1, x_5, x_6) \cdot \phi(y_2, y_3, y_4)\phi(y_2, y_5, y_6) .$$

Then T_1 and T_2 are defined by U statistics with their kernels ϕ and φ, respectively. $M_1(F)$ and $M_2(F)$ are defined by

$$M_1(F) = E(T_1) = \int (F(x, y) - F_1(x)F_2(y))^2 dF(x, y) ,$$

and

$$M_2(F) = E(T_2) = \int (F(x, y) - F_1(x)F_2(y))^2 dF_1(x)dF_2(y) .$$

For abbreviation of the subsequent sections we give two propositions. The proofs are omitted, since they are intuitively obvious. Proposition 2-(1) is seen in [2].

PROPOSITION 1. *For* $i=1, 2$, *and each set* C *of* R^1

$$P_F(T_i \in C) = P_{H_F}(T_i \in C) .$$

Hence for $i=1, 2$

$$M_i(F) = M_i(H_F) .$$

PROPOSITION 2. (1) *If* $F(x, y)$ *is continuous, then* $H_F(u, v)$ *also is continuous.*

(2) *If* $F(x, y)$ *is absolutely continuous, then* $H_F(u, v)$ *also is absolutely continuous, and its probability density function* $h_F(u, v)$ *is given by*

$$h_F(u, v) = \begin{cases} \dfrac{f(F_1^{-1}(u), F_2^{-1}(v))}{f_1(F_1^{-1}(u))f_2(F_2^{-1}(v))} , & \text{when } f_1(F_1^{-1}(u))f_2(F_2^{-1}(v)) \neq 0 ; \\ \text{arbitrary}, & \text{otherwise.} \end{cases}$$

By Propositions 1 and 2, we may assume without loss of generality that both marginal distributions of $F(x, y)$ are $U(0, 1)$.

2. Properties of the measures of association

PROPOSITION 3. *When* $F(x, y)$ *is absolutely continuous,* $M_1(F)=0$, *iff* $F(x, y) = F_1(x)F_2(y)$ *for all* $(x, y) \in R^2$ *i.e.* $F(x, y) \equiv F_1(x)F_2(y)$.

PROOF. We prove only necessity. By Propositions 1 and 2-(2), we may assume that $F_1(x)$ and $F_2(y)$ are $U(0, 1)$. Since $F(x, y)$ is absolutely continuous, $F(x, y)$ has Radon-Nikodym derivative $f(x, y)$. Let

$$E = \{(x, y) \mid f(x, y) > 0\} .$$

Then $\mu(E)>0$, where μ denotes the Lebesgue measure, and $F(x, y)=xy$ a.e. (μ) on E by our condition. Let $\{k_n\}$ and $\{h_n\}$ be two decreasing sequences with their limit 0. We define $DF(x, y)$ by

$$DF(x, y)=\lim_{n\to\infty}\frac{F(x+k_n, y+h_n)-F(x+k_n, y)-F(x, y+h_n)+F(x, y)}{k_n h_n}.$$

By using the theory of the derivatives of functions of a set (c.f. for example [3]), we see that there exists a null set $\wedge$ such that on R^2-$\wedge$ the above limit exists and $DF(x, y)$ coincides with $f(x, y)$. Let

$$E_n=E\cap[E+(0, k_n)]\cap[E+(h_n, 0)]\cap[E+(h_n, k_n)] \qquad n=1, 2, \cdots,$$

where $[E+(x, y)]=[(x'+x, y'+y):(x', y')\in E]$, and let

$$E'=\bigcap_{N=1}^{\infty}\bigcup_{n=N}^{\infty}E_n.$$

Then $\mu(E-E')=0$, and $DF(x, y)=1$ for $(x, y)\in E'$-$\wedge$. Hence $f(x, y)=1$ on E'-$\wedge$, and therefore a.e. (μ) on E. This completes the proof.

Remark 1. When we assume only continuity of $F(x, y)$, the above proposition does not hold. In fact let A, B, C, D and E be defined as shown in Fig. 1 and let

$$F_0(x, y)=\begin{cases} xy & \text{for } (x, y)\in A\cup C\cup D \\ (1/2)x-(1/2-y)^2 & \text{for } (x, y)\in B \\ y(1-y) & \text{for } (x, y)\in E. \end{cases}$$

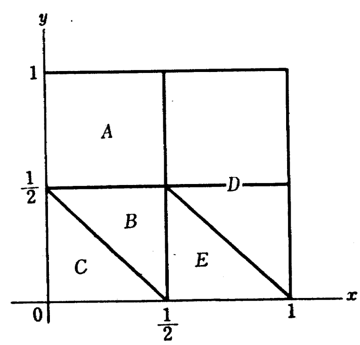

Fig. 1

$F_0(x, y)$ satisfies the conditions of distribution function, and both marginal distributions are $U(0, 1)$. We have easily $M_1(F_0)=0$, but $F_0(1/2, 1/4)\neq 1/8$.

PROPOSITION 4. *$M_2(F)=0$, iff $F(x, y)=F_1(x)F_2(y)$ for all $(x, y)\in R^2$ i.e., $F(x, y)\equiv F_1(x)F_2(y)$.*

PROPOSITION 5. (1) *If $\{F_n(x, y); n=1, 2, \cdots\}$ converges to $F_1(x)\cdot F_2(y)$, then $\lim_{n\to\infty} M_1(F_n)=0$, and* (2) *$\{F_n(x, y); n=1, 2, \cdots\}$ converges to $F_1(x)F_2(y)$, iff $\lim_{n\to\infty} M_2(F_n)=0$.*

One may conjecture that $M_1(\Phi\rho)$ and $M_2(\Phi\rho)$ are increasing functions of $|\rho|$, where $\Phi\rho(x, y)$ is the d.f. of two dimensional normal d.f. with the correlation coefficient ρ. We give this property more generally.

For this purpose we need two partial orders of association introduced in [4].

DEFINITION 1. $F(x, y)$ and $G(x, y)$ have the common marginal d.f.'s $F_1(x)$ and $F_2(y)$. $G(x, y)$ is said to have larger quadrant dependence than $F(x, y)$, if $G(x, y) > F(x, y)$ for all $(x, y) \in R^2$, and we write $G(x, y) > F(x, y)$ (Q.D.).

DEFINITION 2. $F(x, y)$ and $G(x, y)$ have the common marginal distribution functions. $F(y \mid x)$ and $G(y \mid x)$ are conditional d.f.'s of $F(x, y)$ and $G(x, y)$ given $X=x$. $G(x, y)$ is said to have larger regression dependence than $F(x, y)$, if for $x' > x$, $F^{-1}(u \mid x') \geqq F^{-1}(v \mid x)$ implies $G^{-1}(u \mid x') \geqq G^{-1}(v \mid x)$. And we write $G(x, y) > F(x, y)$ (R.D).

In [4], we have seen that $G(x, y) > F(x, y)$ (Q.D.), if $G(x, y) > F(x, y)$ (R.D.), and that for $\rho' > \rho$, $\Phi\rho'(x, y) > \Phi\rho(x, y)$ (R.D.).

PROPOSITION 6. *If either of the following properties is satisfied, then $M_1(G) \geqq M_1(F)$, where the equality is attained, iff $F(x, y) \equiv G(x, y)$,*
(i) $G(x, y) > F(x, y)$ (R.D.), *and* $F(x, y) > F_1(x)F_2(y)$ (R.D.),
(ii) $G(x, y) < F(x, y)$ (R.D.), *and* $F(x, y) < F_1(x)F_2(y)$ (R.D.).

PROOF. We prove only in the case of (i).

$$\begin{aligned} M_1(F) &= \int (F(x, y) - F_1(x)F_2(y))^2 dF(x, y) \\ &= \int (F(x, y) - F_1(x)F_2(y))^2 dF(y \mid x) dF_1(x) \\ &= \int [F(x, F^{-1}(u \mid x)) - F_1(x)F_2(F^{-1}(u \mid x))]^2 du\, dF_1(x) . \end{aligned}$$

By our condition, we get for any x and u

$$G(x, G^{-1}(u \mid x)) \geqq F(x, F^{-1}(u \mid x)) ,$$

and that

$$\begin{aligned} 1 - G_1(x) - G_2(G^{-1}(u \mid x)) + G(x, G^{-1}(u \mid x)) \\ \geqq 1 - F_1(x) - F_2(F^{-1}(u \mid x)) + F(x, F^{-1}(u \mid x)) . \end{aligned}$$

Hence for any x and u

$$\begin{aligned} G(x, G^{-1}(u \mid x)) - G_1(x)G_2(G^{-1}(u \mid x)) \\ \geqq F(x, F^{-1}(u \mid x)) - F_1(x)F_2(F^{-1}(u \mid x)) \geqq 0 . \end{aligned}$$

This completes the proof.

PROPOSITION 7. *If either of the two following properties is satisfied,*

then $M_2(G) \geqq M_2(F)$, *where the equality is attained, iff* $G(x, y) \equiv F(x, y)$,
(i) $G(x, y) > F(x, y)$ (Q.D.), *and* $F(x, y) > F_1(x)F_2(y)$ (Q.D.),
(ii) $G(x, y) < F(x, y)$ (Q.D.), *and* $F(x, y) < F_1(x)F_2(y)$ (Q.D.).

Remark 2. We can not replace the condition of Proposition 6 by that of Proposition 7. In fact let

$$F_\alpha(x, y) = (1-\alpha)F_0(x, y) + \alpha xy \qquad 0 \leqq \alpha \leqq 1,\ 0 \leqq x,\ y \leqq 1,$$

then

$$F_0(x, y) > F_{1/2}(x, y) > F_1(x, y) \text{ (Q.D.)}.$$

On the other hand

$$M_1(F_0) = M_1(F_1) = 0, \qquad \text{but } M_1(F_{1/2}) > 0.$$

The following proposition (1) is given in [1], and we can prove (2) similarly to (1).

PROPOSITION 8. (1) $\sup M_1(F) = 1/30$, *and* (2) $\sup M_2(F) = 1/90$, *where the supremum is taken over all continuous d.f.'s. The supremum is attained, iff* $H_F(u, v) \equiv \mathrm{Min}\{u, v\}$, *or* $H_F(u, v) \equiv \mathrm{Max}\{0, u+v-1\}$.

3. Equiprobable rankings and independence

The relation between equiprobable rankings and independence was considered by Hoeffding in [1]. In this section we give a more precise result. Let $(X_1, Y_1), \cdots, (X_n, Y_n)$ be a sample of size n from (X, Y) with the d.f. $F(x, y)$ and let Π_n be the set of all permutations of $(1, 2, \cdots, n)$. The statistic T is defined by

$$T(X_1, Y_1; \cdots; X_n, Y_n) = (r_1, \cdots, r_n)^{-1} \cdot (s_1, \cdots, s_n) \qquad (\in \Pi_n),$$

when $X_{r_1} > \cdots > X_{r_n}$ and $Y_{s_1} > \cdots > Y_{s_n}$. The tie is neglected, since $F(x, y)$ is assumed to be continuous. It is well-known that $P_F(T = (t_1, \cdots, t_n)) = 1/n!$ for all $(t_1, \cdots, t_n) \in \Pi_n$, if $F(x, y) \equiv F_1(x)F_2(y)$. Conversely, Hoeffding proved, in the appendix of his paper [1], that for $n \geqq 5$ $P_F(T = (t_1, \cdots, t_n)) = 1/n!$ for all $(t_1, \cdots, t_n) \in \Pi_n$ implies $F(x, y) \equiv F_1(x)F_2(y)$, but not for $n=2$, when $F(x, y)$ has the continuous derivative $f(x, y)$. The following proposition gives a complete result under a more general condition such as the one that $F(x, y)$ is continuous.

PROPOSITION 9. *For* $n \geqq 4$, $P_F(T = (t_1, \cdots, t_n)) = 1/n!$ *for all* $(t_1, \cdots, t_n) \in \Pi_n$ *implies* $F(x, y) \equiv F_1(x)F_2(y)$, *but not for* $n=2$ *or* 3.

PROOF. Since $P_F(T = (t_1, \cdots, t_{n+1})) = 1/(n+1)!$ for all $(t_1, \cdots, t_{n+1}) \in$

Π_{n+1} implies that $P_F(T=(t_1,\cdots,t_n))=1/n!$ for all $(t_1,\cdots,t_n)\in\Pi_n$, we may only prove for $n=4$, and give a counter example for $n=3$.

Case 1. $n=4$. It follows by definition that

$$M_1(F)+2M_2(F)=\int F^2(x,y)dF(x,y)$$
$$-2\int F(x,y)F_1(x)F_2(y)dF(x,y)$$
$$+2\int F^2(x,y)dF_1(x)dF_2(y)-1/9\,.$$

Let $(X_1,Y_1),\cdots,(X_4,Y_4)$ be a random sample from (X,Y) with the d.f. $F(x,y)$ and let

$$S=C(X_4-X_1)C(Y_4-Y_1)C(X_4-X_2)C(Y_4-Y_2)$$
$$-2C(X_4-X_1)C(Y_4-Y_1)C(X_4-X_2)C(Y_4-Y_3)$$
$$+2C(X_4-X_1)C(Y_3-Y_1)C(X_4-X_2)C(Y_3-Y_2)-1/9\,.$$

Then $E_F(S)=M_1(F)+2M_2(F)$. Since S depends only on rank statistic, we get $E_F(S)=E_{F_1F_2}(S)$. Since $E_{F_1F_2}(S)=0$, $M_1(F)+2M_2(F)=0$, which implies $F(x,y)\equiv F_1(x)F_2(y)$.

Case 2. $n=3$. To show this part, we need the following lemma without proof.

LEMMA. *For all* $(t_1,\cdots,t_n)\in\Pi_n$ *we have*
(1) $H_F(u,v)\equiv H_F(v,u)$ *implies* $P_F(T=(t_1,\cdots,t_n))=P_F(T=(t_1,\cdots,t_n)^{-1})$,
(2) $H_F(u,v)\equiv u+v-1+H_F(1-u,1-v)$ *implies* $P_F(T=(t_1,\cdots,t_n))=P_F(T=(t'_1,\cdots,t'_n))$, *where* $t_i+t'_{n-i+1}=n+1$ *for* $i=1,2,\cdots,n$, *and*
(3) $H_F(u,v)\equiv u-H_F(u,1-v)$ *implies* $P_F(T=(t_1,\cdots,t_n))=P_F(T=(t_n,\cdots,t_1))$.

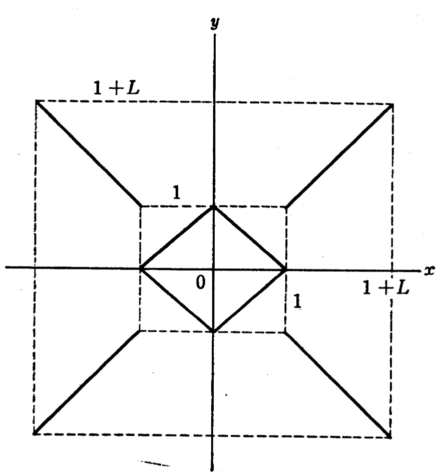

Fig. 2

Let $F_L(x,y)$ be the d.f. which is distributed uniformly on solid lines of Fig. 2 excluding x and y-axis, where L is a nonnegative number.

Let us denote $P_{F_L}(T=(t_1,t_2,t_3))$ by $P_L(t_1,t_2,t_3)$. Since $F_L(x,y)$ satisfies three conditions of the lemma, $P_L(1,2,3)=P_L(3,2,1)$ and $P_L(2,1,3)=P_L(1,3,2)=P_L(3,1,2)=P_L(2,3,1)$. Hence, we have only to show that there exists L such that

$$P_L(T=(1, 2, 3))=3!(2/3L^3+2L^2+2L+1/3)/4^2(1+L)^3 .$$

We denote the right-hand side by $P(L)$. Then $P(0)=1/8<1/6$, $\lim_{L\to\infty} P(L)=1/4>1/6$ and $P(L)$ is continuous on $(-1, \infty)$. This completes the proof.

Acknowledgement

The author wishes to thank the referee for his useful comments.

THE INSTITUTE OF STATISTICAL MATHEMATICS

REFERENCES

[1] Hoeffding,W. (1948). A nonparametric test of independence, *Ann. Math. Statist.*, **19**, 546-557.

[2] Konijn,H. S. (1959). Positive and negative dependence of two random variables, *Sankhya*, **21**, 269-280.

[3] Saks, S. (1937). *Theory of the Integral*, Warszawa,

[4] Yanagimoto, T. and Okamoto, M. (1969). Partial orderings of permutations and monotonicity of a rank correlation statistic," *Ann. Inst. Statist. Math.*, **21**, 489-506.