KŌDAI MATH. SEM. REP.
19 (1967), 147—155

INTERACTION INFORMATION IN MULTIVARIATE PROBABILITY DISTRIBUTIONS

By Minoru Sakaguchi

We show in this note that the entropy of a multivariate distribution can be expressed in terms of the sum of one-dimensional marginal entropies, the sum of transmitted information between each pair of component variables, the sum of interaction information in trivariate component distributions, and so on (Section 1). Using this result we give, in Section 2, a class of multivariate distributions having specified component densities and some preassigned association measure between some component variables. Proofs of equations and statements which are not so evident are given in Section 3.

1. Multivariate information transmission.

The transmission of information requires the presence of a source of information coupled with an appropriate channel; the two together form what is called an information system. Here an information system is described in terms of joint probabilities of inputs and outputs, and a channel is defined by its transition probabilities. The formulae are written as if x, y, etc., were continuous real variables; the obvious modifications must be made if they are discrete.

Let us consider a communication channel and its input and output. Transmitted information measures the amount of association between the input and output of the channel. If input and output are independent, no information is transmitted. On the other hand, if both are perfectly correlated, all the input information is transmitted through the channel. In most cases, naturally, information transmission is found between these extremes.

We are interested in the amount of information transmitted. Suppose that we have a bivariate probability distribution with the density function $p(x, y)$. This means that if the input variable assumes a value or signal x, then noise of the channel alters it, at the output, to a value between y and $y+dy$ with probability $p(y|x)dy$, where

$$p(y|x)=p(x, y)\Big/\int p(x, y)dy,$$

and that the rules governing the selection of signals at the input must be constructed so that they take on values between x and $x+dx$ with probabilities $p(x)dx=dx\int p(x, y)dy$. To avoid complexity we use the same notation $p(\cdot)$ to

Received December 15, 1965.

represent the various density functions of random variables, without any suggestion that they have the same density.

Under these conditions, and if successive signals are independent the amount of information transmitted per signal is defined by Shannon [4] as

$$T(x;y)=H(x)+H(y)-H(x,y)=\iint p(x,y)\log\frac{p(x,y)}{p(x)p(y)}\,dxdy, \tag{1.1}$$

where $H(x)\equiv-\int p(x)\log p(x)dx$, $H(y)\equiv-\int p(y)\log p(y)dy$ and $H(x,y)=-\iint p(x,y)\log p(x,y)dxdy$. It is well known (Kullback [2]) that $T(x;y)$ is non-negative and equals zero if and only if x and y are independent.

We introduce the conditional entropies

$$H_x(y)\equiv-\iint p(x,y)\log p(y|x)dxdy, \quad \text{etc.}$$

Then we have the additive formula

$$H(x_1,x_2,\cdots,x_k)=H(x_1)+H_{x_1}(x_2)+\cdots+H_{x_1,\cdots,x_{k-1}}(x_k), \qquad k\geqq 2.$$

Thus we have other expressions of $T(x;y)$ as

$$T(x;y)=H(x)-H_y(x)=H(y)-H_x(y). \tag{1.2}$$

Now let us consider the case where we have several sources that transmit to y. Then we take the input variable as multidimensional and we have, for instance,

$$\begin{aligned}T(u,v;y)&=H(u,v)+H(y)-H(u,v,y)\\&=H(u,v)-H_y(u,v)=H(y)-H_{u,v}(y).\end{aligned} \tag{1.3}$$

We can express $T(u,v;y)$ as a combination of the bivariate transmissions between u and y, and v and y. Define $T_{u=u_0}(v;y)$ as transmitted information between v and y for a particular value of u, namely, u_0. If we set

$$T_u(v;y)=\int T_{u=u_0}(v;y)p(u_0)du_0,$$

we easily find that

$$\begin{aligned}T_u(v;y)&=H_u(v)+H_u(y)-H_u(v,y)\\&=H_u(v)-H_{u,y}(v)=H_u(y)-H_{u,v}(y).\end{aligned} \tag{1.4}$$

Hence we have, from (1.2), (1.3) and (1.4),

$$T(u,v;y)=T(u;y)+T_u(v;y)=T(v;y)+T_v(u;y), \tag{1.5}$$

which means that the additive formula for information transmission also holds true. We have from (1.2) and (1.4)

$$T(v;y)-T_u(v;y)=(H(y)-H_v(y))-(H_u(y)-H_{u,v}(y))$$

$$=(H(v)-H_y(v))-(H_u(v)-H_{u,y}(v)),$$

which may be positive, zero or negative. These identities show the symmetry of the left-hand side expression in the arguments u and v, and u and y. Since the symmetry v and y is evident from (1. 1) and (1. 3), we get

$$A(uvy)\equiv T(v;y)-T_u(v;y)=T(u;y)-T_v(u;y)=T(u;v)-T_y(u;v). \tag{1. 6}$$

We call this quantity $A(uvy)$, following McGill [3], the *interaction information* between the three variables. It is the gain or loss in transmitted information between any two of the variables, due to additional knowledge of the third variable.

We can derive another expression for $A(uvy)$ as follows: subtracting $T(v;y)$ from both sides of the first identity of (1. 5) we have

$$T(u,v;y)=T(u;y)+T(v;y)-A(uvy). \tag{1. 7}$$

By (1. 3) we have $T(u,v;y)=H(u)+H_u(v)+H(y)-H(u,v,y)$. Then from these two identities and the fact that $T(u;v)=H(v)-H_u(v)$, we finally obtain

$$H(u,v,y)=H(u)+H(v)+H(y)-(T(u;v)+T(u;y)+T(v;y))+A(uvy). \tag{1. 8}$$

According to the definition (1. 6) the interaction information is positive (negative) when the effect of holding one of the interacting variables constant is to decrease (increase) the amount of association between the other two. We easily observe from (1. 8)

THEOREM 1′. *If the random variables in a trivariate distribution are pairwise independent, then* $A(uvy)\leqq 0$, *with equality if and only if the three variables are mutually independent.*

It is well-known that pairwise independence of three random variables does not imply mutual independence. Let $\mathfrak{X}=\{a_1,\cdots,a_4\}$ be a probability space with probabilities 1/4 for each elementary event a_i $(i=1,\cdots,4)$. Let $A=\{a_1,a_2\}$, $B=\{a_1,a_3\}$ and $C=\{a_1,a_4\}$. Let u, v and y be indicator functions of the events A, B, and C, respectively. Then we find that the three random variables are pairwise independent but are not mutually independent. We have

$$A(uvy)=H(u,v,y)-(H(u)+H(v)+H(y))=2\log 2-3\log 2=-\log 2.$$

A class of continuous trivariate distribution having this nature is given in the next section.

In a similar way as the above discussion we can decompose four-variate distributions. Let (x_1,x_2,x_3,y) be a four-variate distribution. Define $A_{x_1=x_1^0}(x_2x_3y)$ as the conditional interaction information between the variables x_2, x_3 and y given that $x_1=x_1^0$. Define

$$A_{x_1}(x_2x_3y)=\int A_{x_1=x_1^0}(x_2x_3y)p(x_1^0)dx_1^0. \tag{1. 9}$$

Then we can prove the following relations:

$$A_{x_1}(x_2x_3y)=T_{x_1}(x_2;\,y)-T_{x_1,\,x_3}(x_2;\,y)=\cdots \tag{1.10}$$

is invariant under any permutations of the variables x_2, x_3 and y;

$$A(x_2x_3y)-A_{x_1}(x_2x_3y)(\equiv A(x_1x_2x_3y),\text{ say}) \tag{1.11}$$

is invariant under any permutation of the variables x_1, x_2, x_3 and y;

$$A((x_1, x_2)x_3y)=A(x_1x_3y)+A(x_2x_3y)-A(x_1x_2x_3y); \tag{1.12}$$

and

$$\begin{aligned} T(x_1, x_2, x_3, y)=&T(x_1;\,y)+T(x_2;\,y)+T(x_3;\,y)\\ &-A(x_1, x_2y)-A(x_2x_3y)-A(x_1x_3y)+A(x_1x_2x_3y). \end{aligned} \tag{1.13}$$

These identities correspond to those, in the trivariate case, (1. 4), (1. 6), (1. 7) and (1. 7), respectively. From (1. 8) and (1. 13) we finally obtain, rewriting y as x_4,

$$H(x_1, x_2, x_3, x_4)=\sum_{i=1}^{4}H(x_i)-\sum_{i<j}{}'\,T(x_i;\,x_j)+\sum_{i<j<l}{}''\,A(x_ix_jx_l)-A(x_1x_2x_3x_4), \tag{1.14}$$

where the sums $\sum'$ and $\sum''$ contain six T-terms and four A-terms, respectively. We call $A(x_1x_2x_3x_4)$ *higher-order interaction information* (with order 2). Note that this is not equal to the interaction information (with order 1) $A((x_1, x_2)x_3x_4)$, as is shown by (1. 12). For the proofs of (1. 10)—(1. 14) see Section 3.

From Theorem 1′ and (1. 14) we have

THEOREM 1″. *If every three random variables in a four-variate distribution are independent, then* $A(x_1x_2x_3x_4)\geqq 0$, *with equality if and only if the four variables are independent.*

Generalization of this theorem to n-dimensional distributions is immediate. Define higher-order interaction information recursively by

$$A(x_1x_2\cdots x_n)=A(x_2x_3\cdots x_n)-A_{x_1}(x_2x_3\cdots x_n), \qquad n\geqq 4,$$

starting from $A(x_1x_2x_3x_4)$, where $A_{x_1}(x_2\cdots x_n)$ is defined by a similar expression used in (1. 9). Then we can prove by mathematical induction the following relations corresponding to (1. 10)—(1. 14):

$$A_{x_1}(x_2x_3\cdots x_n)=A_{x_1}(x_3\cdots x_n)-A_{x_1,x_2}(x_3\cdots x_n)=\cdots \tag{1.15}$$

is invariant under any permutation of the variables $x_2, x_3, \cdots, x_n$;

(1. 16) $A(x_1, x_2\cdots x_n)$ is invariant under any permutation of the variables $x_1, x_2, \cdots, x_n$;

$$A((x_1, x_2)x_3\cdots x_n)=A(x_1x_3\cdots x_n)+A(x_2x_3\cdots x_n)-A(x_1x_2x_3\cdots x_n); \tag{1.17}$$

$$\begin{aligned} T(x_1, \cdots, x_{n-1};\,y)=&\sum_{i=1}^{n-1}T(x_i;\,y)-\sum_{i<j\leqq n-1}{}'\,A(x_ix_jy)\\ &+\sum_{i<j<l\leqq n-1}{}''\,A(x_ix_jx_ly)-\cdots+(-1)^{n-2}A(x_1\cdots x_{n-1}y); \end{aligned} \tag{1.18}$$

and finally

$$H(x_1, \cdots, x_n)=\sum_{i=1}^{n} H(x_i)-\sum_{i<j}{}' T(x_i; x_j) + \sum_{i<j<l}{}'' A(x_i x_j x_l)-\cdots+(-1)^{n-1}A(x_1\cdots x_n). \tag{1.19}$$

From Theorem 1'' and (1. 19) we get

THEOREM 1. *If every* $(n-1)$ *random variables in an n-variate distribution (with* $n\geqq 3$*) are independent, then for the higher-order interaction information (with order* $n-2$*)* $A(x_1\cdots x_n)$ *we have* $(-1)^n A(x_1\cdots x_n)\geqq 0$*, with equality if and only if the n variables are independent.*

2. Multivariate distributions with given marginal distributions.

There exist infinitely many bivariate distributions with a given pair of marginal distributions. Let $f_1(\cdot)$ and $f_2(\cdot)$ be two given pdf's. A class of bivariate densities $f(x_1, x_2)$ with given marginal densities $f_1(\cdot)$ and $f_2(\cdot)$ is given by

$$f(x_1, x_2)=f_1(x_1)f_2(x_2)\{1+a(1-2F_1(x_1))(1-2F_2(x_2))\}, \tag{2.1}$$

where $F_i(x_i)$, $i=1, 2$, is the cdf of $f_i(x_i)$ and a is an arbitrary constant satisfying $-1\leqq a\leqq 1$ (Gumbel [1]). It is easy to check that the bivariate cdf is given by

$$F(x_1, x_2)=F_1(x_1)F_2(x_2)\{1+a(1-F_1(x_1))(1-F_2(x_2))\} \tag{2.2}$$

and that x_1 and x_2 are independent if and only if $a=0$. However, the correlation coefficient of this bivariate distribution depends on $f_1(\cdot)$ and $f_2(\cdot)$ and cannot be expressed in terms of a only. We want to show that the constant a actually measures dependency between the two variables, independently of $f_1(\cdot)$ and $f_2(\cdot)$, in the following two senses.

(i) Silver [5] defined a measure of association between two random variables x and y by

$$\Delta\equiv\iint_{p(x,y)/(p(x)p(y))>1}(p(x, y)-p(x)p(y))dxdy,$$

about which he showed some desirable properties in measuring association. For the density (2. 1) this becomes

$$\Delta=a\iint_{\substack{a(2u_1-1)(2u_2-1)>0\\ 0\leqq u_1, u_2\leqq 1}}(2u_1-1)(2u_2-1)du_1du_2=\frac{|a|}{16}.$$

Thus this measure of association is a function of $|a|$ only and is monotonically increasing.

(ii) Elementary calculation yields that information transmitted between x_1 and x_2 is given by

$$T(x_1; x_2)=\iint_{-\infty}^{\infty} f(x_1, x_2)\log\frac{f(x_1, x_2)}{f_1(x_1)f_2(x_2)}dx_1dx_2$$
$$=\sum_{m=1}^{\infty}\frac{a^{2m}}{2m(2m-1)(2m+1)^2}\ (\equiv t(|a|), \text{ say}) \tag{2. 3}$$

independently of $f_1(\cdot)$ and $f_2(\cdot)$. This is again a function of $|a|$ only, and increases monotonically from zero to $t(1)$ in $0\leqq|a|\leqq 1$. Usual calculation gives another expression for (2. 3) as

$$t(|a|)=\frac{(1+a)(3+a)}{8a}\log(1+a)+\frac{(1-a)(3-a)}{-8a}\log(1-a)-\frac{3}{4}+\frac{1}{2}\sum_{m=1}^{\infty}\frac{a^{2m}}{(2m+1)^2}$$

and

$$t(1)=\log 2-\frac{5}{4}+\frac{\pi^2}{16},\quad \text{since}\quad \sum_{m=1}^{\infty}m^{-2}=\frac{\pi^2}{6}.$$

The class (2. 1) can easily be extended to trivariate distributions. Let $f_i(\cdot)$, $i=1, 2, 3$, be three given pdf's. A class of trivariate densities $f(x_1, x_2, x_3)$ with given marginal densities is given by

$$f(x_1, x_2, x_3)=f_1f_2f_3\{1+a(1-2F_1)(1-2F_2)+b(1-2F_2)(1-2F_3)$$
$$+c(1-2F_1)(1-2F_3)+d(1-2F_1)(1-2F_2)(1-2F_3)\}, \tag{2. 4}$$

where $F_i(\cdot)$ is the cdf of $f_i(\cdot)$ and the obvious arguments x_i, $i=1, 2, 3$, are omitted. The corresponding trivariate cdf is

$$F(x_1, x_2, x_3)=F_1F_2F_3\{1+a(1-F_1)(1-F_2)+b(1-F_2)(1-F_3)$$
$$+c(1-F_1)(1-F_3)+d(1-F_1)(1-F_2)(1-F_3)\}. \tag{2. 5}$$

The four arbitrary constants a, b, c, and d satisfy

$$|a|\leqq\bar{a},\quad |b|\leqq\bar{b},\quad |c|\leqq\bar{c},\quad |d|\leqq\bar{d}, \tag{2. 6}$$

where $\bar{a}+\bar{b}+\bar{c}+\bar{d}=1$. They measure dependency between various variables in the following sense: $T(x_1; x_2)=t(|a|)$, $T(x_2; x_3)=t(|b|)$, $T(x_1; x_3)=t(|c|)$ and $A(x_1x_2x_3)$ can be expressed by a function of a, b, c and d, independently of f_1, f_2 and f_3. If $a=b=c=0$, then the three variables are independent if and only if $d=0$. The expression (2. 4) gives a class of examples of trivariate distributions with the property that any two variables are independent but are not independent between the three variables. That is, if $f(x_1, x_2, x_3)$ belongs to the class of densities determined by $\bar{a}=\bar{b}=\bar{c}=0$ and $\bar{d}=1$, then $T(x_1; x_2)=T(x_2; x_3)=T(x_1; x_3)=0$, but by (1. 7),

$$-A(x_1, x_2, x_3)=T(x_1, x_2; x_3)$$
$$=\iiint_{-\infty}^{\infty}f(x_1, x_2, x_3)\log\frac{f(x_1, x_2, x_3)}{f(x_1, x_2)f_3(x_3)}dx_1dx_2dx_3$$
$$=\iiint f(x_1, x_2, x_3)\log\{f(x_1, x_2, x_3)/(f_1(x_1)f_2(x_2)f_3(x_3))\}dx_1dx_2dx_3$$
$$=\sum_{m=1}^{\infty}\frac{d^{2m}}{2m(2m-1)(2m+1)^3}$$

by straightforward calculations. The last expression does not involve f_i's and is a function of only $|d|$ increasing monotonically from zero to $\sum_{m=1}^{\infty}\{2m(2m-1)(2m+1)^3\}^{-1}$ in $0\leqq|d|\leqq 1$.

3. Proofs.

(1. 10): From (1. 7) we have

$$H_{x_1}(x_2, x_3, y)=H_{x_1}(x_2)+H_{x_1}(x_3)+H_{x_1}(y)-(T_{x_1}(x_2; x_3)+T_{x_1}(x_3; y)+T_{x_1}(x_2; y))+A_{x_1}(x_2x_3y),$$

which shows invariance of $A_{x_1}(x_2x_3y)$ under any permutations of x_2, x_3 and y.

From (1. 7) we have

$$(*)\qquad A_{x_1}(x_2x_3y)=T_{x_1}(x_2; y)+T_{x_1}(x_3; y)-T_{x_1}(x_2, x_3; y).$$

From (1. 5) we have

$$\begin{aligned}T_{x_1,x_3}(x_2; y)&=T(x_1, x_3, x_2; y)-T(x_1, x_3; y)\\&=(T(x_1; y)+T_{x_1}(x_2, x_3; y))-(T(x_1; y)+T_{x_1}(x_3; y))\\&=T_{x_1}(x_2, x_3; y)-T_{x_1}(x_3; y).\end{aligned}$$

Adding the last identity and (*) side-by-side, we obtain

$$A_{x_1}(x_2x_3y)+T_{x_1,x_3}(x_2; y)=T_{x_1}(x_2; y),$$

which proves (1. 10).

(1. 11): It suffices to show that $A(x_1x_2x_3y)$ is invariant under any permutation of x_1 and any other one variable. From (1. 6) and (1. 10) we have

$$\begin{aligned}A(x_1x_2x_3y)&\equiv A(x_2x_3y)-A_{x_1}(x_2x_3y)\\&=(T(x_2; y)-T_{x_3}(x_2; y))-(T_{x_1}(x_2; y)-T_{x_1,x_3}(x_2; y))\\&=(T(x_2; y)-T_{x_1}(x_2, y))-(T_{x_3}(x_2; y)-T_{x_3,x_1}(x_2; y))\\&=A(x_1x_2y)-A_{x_3}(x_1x_2y).\end{aligned}$$

The last expression is invariant under any permutation of x_1, x_2 and y.

(1. 12): From (1. 6) and (1. 11) we obtain

$$\begin{aligned}A(x_1x_3y)&=T(x_1; y)-T_{x_3}(x_1; y),\\A(x_2x_3y)&=T(x_2; y)-T_{x_3}(x_2; y),\\-A(x_1x_2x_3y)&=-A(x_1x_2y)+A_{x_3}(x_1x_2y).\end{aligned}$$

Adding together we get

$$\begin{aligned}A(x_1x_3y)+A(x_2x_3y)-A(x_1x_2x_3y)&=T(x_1; y)+T(x_2; y)-A(x_1x_2y)\\&\quad-(T_{x_3}(x_1; y)+T_{x_3}(x_2; y)-A_{x_3}(x_1x_2y))\\&=T(x_1, x_2; y)-T_{x_3}(x_1, x_2; y)\qquad\text{(by (1. 7) and (*))}\\&=A((x_1, x_2)x_3y)\qquad\text{(by 1. 6)),}\end{aligned}$$

which proves (1. 12).

(1. 13): From (1. 7) we have

$$T(x_1, x_2, x_3; y)=T((x_1, x_2), x_3; y)=T(x_1, x_2; y)+T(x_3; y)-A((x_1, x_2)x_3y)$$
$$=(T(x_1; y)+T(x_2; y)-A(x_1x_2y))+T(x_3; y)-A((x_1, x_2)x_3y).$$

(1. 13) follows from (1. 12).

(1. 14): From (1. 1) and (1. 8) we have

$$H(x_1, x_2, x_3, y)=H(x_1, x_2, x_3)+H(y)-T(x_1, x_2, x_3; y)$$
$$=\left(\sum_{i=1}^{3} H(x_i)-\sum_{i<j} T(x_i; x_j)+A(x_1x_2x_3)\right)+H(y)-T(x_1, x_2, x_3; y).$$

(1. 14) follows from (1. 13).

Proofs of the relations (1. 15)—(1. 19) by mathematical induction are similar, so will be omitted.

(2. 3): Termwise integration of the power series

$$\log\frac{f(x_1, x_2)}{f_1(x_1)f_2(x_2)}=\sum_{n=1}^{\infty}\frac{(-1)^{n-1}a^n}{n}(1-2F_1(x_1))^n(1-2F_2(x_2))^n$$

gives (2. 3). Another expression of $t(|a|)$ follows from the decomposition

$$\frac{1}{2m(2m-1)(2m+1)^2}=\frac{-1}{2m}+\frac{1/4}{2m-1}+\frac{3/4}{2m+1}+\frac{1/2}{(2m+1)^2}.$$

(2. 4) and (2. 6): We show that the right-hand side of (2. 4) is non-negative if $\bar{a}+\bar{b}+\bar{c}+\bar{d}=1$. We have

$$\frac{a}{2}\left(\frac{1}{2}-F_1\right)\left(\frac{1}{2}-F_2\right)\geqq-\frac{\bar{a}}{8},\qquad \text{similar two inequalities, and}$$

$$d\left(\frac{1}{2}-F_1\right)\left(\frac{1}{2}-F_2\right)\left(\frac{1}{2}-F_3\right)\geqq-\frac{\bar{d}}{8}.$$

Adding these four inequalities together we get

$$a(1-2F_1)(1-2F_2)+b(1-2F_2)(1-2F_3)+c(1-2F_1)(1-2F_3)+d(1-2F_1)(1-2F_2)(1-2F_3)$$
$$\geqq(\bar{a}+\bar{b}+\bar{c}+\bar{d})=-1.$$

References

[1] Gumbel, E. J., Distributions a plusieurs variables dont les marges sont donnees. C. R. Acad. Sci. Paris **246** (1958), 2717-2720.

[2] Kullback, S., Information Theory and Statistics. John Wiley and Sons, New York (1959).

[3] McGill, W. J., Multivariate information transmission. Trans. I.R.E., PGIT-4 (Sept. 1954), 93-111.

[4] SHANNON, C. E., AND W. WEAVER, The Mathematical Theory of Communication. University of Illinois Press (1949).

[5] SILVEY, S. D., On a measure of association. Ann. Math. Statist. **35** (1964), 1157–1166.

DEPARTMENT OF APPLIED MATHEMATICS,
OSAKA UNIVERSITY.