ON A MEASURE OF ASSOCIATION

By S. D. Silvey

University of Manchester

1. Introduction and summary. The problem of obtaining a satisfactory measure of association between two random variables is closely allied to that of obtaining a measure of the amount of information about one contained in the other. For the more closely associated are the random variables the more information about one ought to be given by an observation on the other and vice versa. It is not, therefore, surprising to find that there have been several suggestions for basing coefficients of association on the now celebrated measure of information introduced by Shannon [9] in the context of communication theory. (See Bell [1] for certain of these and for references to others). Now Shannon's measure of information was based on the notion of entropy which seems to be much more meaningful for finite probability spaces than it is for infinite spaces, and while Gel'fand and Yaglom [2] have suggested a generalisation of Shannon's measure for infinite spaces, there remain difficulties, as indicated by Bell [1], about deriving from it coefficients of association or dependence between random variables taking infinite sets of values.

In the present paper, by adopting a slightly different attitude to information from that of communication theory, we shall obtain a general measure of information which yields a fairly natural coefficient of dependence between two continuous random variables or, more generally, between two non-atomic measures. The next section provides the motivation for the introduction of this measure of information and a general definition is given in Section 3. In Section 4 we discuss some of the properties of this measure regarded as a coefficient of association along the lines suggested by Rényi [8]. Finally, in Section 5, we indicate the relevance of this measure to estimation theory.

2. Motivation. A naive geometric picture of close association between two random variables $\tilde{x}$ and $\tilde{y}$ is that of a distribution concentrated in some sort of way about some sort of curve in the plane. Before we can proceed from this point we require an analytic interpretation of this somewhat vague picture. In view of the affinity between the notions of association and information, it is not surprising to find that in the literature on communication theory initiated by by Shannon [9], and continued by McMillan [7], Khinchin [5], Joshi [4] and many others is implicit the following analytic interpretation. The nature and extent of association between two random variables $\tilde{x}$ and $\tilde{y}$ is described by the Radon-Nikodym derivative $\phi(\tilde{x}, \tilde{y})$ of their joint distribution with respect to the product of their marginal distributions. (We assume for the moment that ϕ exists: the possibility of singularity or part-singularity of one distribution with respect to the other will cause only a slight difficulty with which we shall deal later.)

Received 26 November 1963; revised 22 January 1964.

If $\phi \equiv 1$, then $\tilde{x}$ and $\tilde{y}$ are independent. If ϕ takes large values with probability near 1, $\tilde{x}$ and $\tilde{y}$ are "closely associated" over most of their ranges.

If it is accepted that the random variable ϕ describes association, then measuring association by means of a single number may be interpreted as focusing attention on some particular aspect of its distribution. Which aspect we wish to concentrate on and so which measure of association we use, depends on the problem in hand. This point was made emphatically by Goodman and Kruskal [3] in their comprehensive papers on measures of association for cross classifications. The measure of association which we shall propose is motivated by the kind of considerations underlying communication theory, or rather by their statistical interpretation so clearly stated by McMillan [7].

The information theory outlook on association between random variables $\tilde{x}$ and $\tilde{y}$ is as follows. If $\tilde{x}$ and $\tilde{y}$ are associated then most observations x on $\tilde{x}$ give information about $\tilde{y}$ in the sense that the conditional distribution of $\tilde{y}$, given x, is more concentrated than its unconditional distribution. There are various ways of measuring this difference in concentration which will lead to reasonable measures of the information about $\tilde{y}$ provided by x. The most obvious possibility is to use standard deviation as a measure of concentration. But this is lacking in generality since we may obviously have association and consequent increase in concentration of a conditional distribution without standard deviations being defined. Shannon [9] in the particular case of random vectors taking only a finite set of values introduces entropy and substitutes decrease in entropy for increase in concentration. However, as we have said, the notion of entropy loses some of its significance when we move outside the finite case and this is reflected in difficulties of deriving general measures of association from it.

Another way of looking at this, one which has something in common with the idea underlying the measures proposed by Goodman and Kruskal [3], is as follows. If $\tilde{x}$ and $\tilde{y}$ are closely associated then there exist sets of values of $\tilde{y}$ whose conditional probabilities, given a value x of $\tilde{x}$, are much larger than their unconditional probabilities. This suggests the following possibility, in the discussion of which we will not be preoccupied with rigour since the object is simply to provide motivation. Suppose that for each given x we set

$$d(x) = \Pr\{Y_x \mid x\} - \Pr\{Y_x\}$$

where Y_x is a set of values of $\tilde{y}$ chosen to maximise the difference between conditional and unconditional probability. Then $d(x)$ is a possible measure of the amount of information about $\tilde{y}$ contained in the observation x on $\tilde{x}$. It takes values between 0 and 1. If $\tilde{x}$ and $\tilde{y}$ are independent it obviously is zero. If $\tilde{x}$ and $\tilde{y}$ are continuous and their distribution is concentrated on a curve, i.e. if they are mathematically related then usually $d(x) = 1$, for every x. In intermediate cases the naive picture previously mentioned suggests that increases in association will be reflected in increases in $d(x)$, for most x. Now if we average over x we obtain an average measure of the information about $\tilde{y}$ provided by $\tilde{x}$. This measure we shall denote by Δ. It always lies between 0 and 1 and might therefore be used directly, without normalizing, as a measure of the dependence of $\tilde{y}$ on $\tilde{x}$. Since, as we shall see in the ensuing non-rigorous argument it turns out to be symmetric in $\tilde{x}$ and $\tilde{y}$, it is a natural measure of association between them. In fact, Y_x, chosen to maximise $\Pr\{Y \mid x\} - \Pr\{Y\}$ is, if we assume existence of appropriate densities denoted by p, clearly given by

$$Y_x = \{y: p(y \mid x) > p(y)\} = \{y: \phi(x, y) > 1\},$$

where $\phi(x, y) = p(x, y)/[p(x)p(y)]$. Then

$$d(x) = \int_{\{y:\phi(x,y)>1\}} [p(y \mid x) - p(y)]\, dy$$

and

$$\Delta = E(d) = \int_{-\infty}^{\infty} d(x)p(x)\, dx$$

$$= \iint_{\{(x,y):\phi(x,y)>1\}} [p(x, y) - p(x)p(y)]\, dx\, dy.$$

This displays the symmetry of the suggested average measure and also points the way to a quite general definition.

3. General definition. Suppose that $(X, \mathfrak{F}, \mu)$ is a probability space and that corresponding to each $x \in X$ is defined a probability measure ν_x on a measurable space $(Y, \mathcal{G})$. Then the measure μ on X and the family ν_x, $x \in X$, on Y define a joint probability measure ω, say, on $(X \times Y, \mathfrak{F} \times \mathcal{G})$ and a marginal probability measure ν, say, on $(Y, \mathcal{G})$. We shall denote by ω^* the product $\mu \times \nu$, a probability measure on $(X \times Y, \mathfrak{F} \times \mathcal{G})$, and by ϕ the Radon-Nikodym derivative of ω with respect to ω^*. Here we allow a generalised derivative which may take the value $+\infty$; i.e., if ω is not absolutely continuous with respect to ω^* so that there exists a set A such that $\omega^*(A) = 0$ and $\omega(A) > 0$, we set $\phi(x, y) = +\infty$ for all $(x, y) \in A$, and then define $\int_A \phi\, d\omega^*$ to be $\omega(A)$.

Now let $W = \{(x, y): \phi(x, y) > 1\}$ and define

$$\Delta = \int_W (d\omega - d\omega^*) = \int_W (\phi - 1)\, d\omega^*.$$

This number Δ is the generalised version of the coefficient introduced in Section 2 as a coefficient of association between two random variables. In this general form it may be regarded as a coefficient of association between random variables which are not necessarily real-valued. We shall now investigate certain of its properties from this viewpoint.

4. Δ as a coefficient of association. Rényi [8] has stated seven properties which should be satisfied by a measure of dependence of two real random variables and whether or not one agrees that a measure providing a very drastic summary of a complex situation should, irrespective of the object of the summary, invariably have specific properties, Rényi's postulates provide a very useful basis for discussion.

4.1. The first demand is that a coefficient δ of association should be defined for any pair of random variables of which neither is constant with probability 1. In fact Δ is even more generally defined, for it is defined for any pair of random variables. If one of the random variables is constant with probability 1, then $\Delta = 0$ as one would hope, for then $\omega = \omega^*$.

4.2. Trivially from the definition Δ is symmetric, that is

$$\Delta(\tilde{x}, \tilde{y}) = \Delta(\tilde{y}, \tilde{x}).$$

4.3. $0 \leqq \Delta \leqq 1$, since $\Delta = \int_{\phi>1} (\phi - 1)\, d\omega^* \geqq 0$, while obviously $\Delta \leqq \int_{X\times Y} d\omega = 1$.

4.4. $\Delta = 0$ iff $\omega = \omega^*$. Obviously $\omega = \omega^*$ implies $\Delta = 0$ while $\Delta = 0$ implies $\omega = \omega^*$ because

$$0 = \int_{X\times Y} (d\omega - d\omega^*) = \int_{\phi \leqq 1} (\phi - 1)\, d\omega^* + \Delta.$$

Now $\Delta = 0 \Rightarrow P^*(\phi > 1) = 0$ directly, and from the above equality, $\Delta = 0 \Rightarrow P^*(\phi < 1) = 0$ also. (Here P and P^* refer to probabilities defined by the measures ω and ω^* respectively.) Hence $P^*(\phi = 1) = 1$ and so $\omega = \omega^*$. Interpreted for the particular case where we are dealing with association between two real random variables, this says that $\Delta = 0$ iff the random variables are independent, which is Rényi's fourth demand.

4.5. The first potential objection to Δ as a coefficient of association is encountered when we consider Rényi's fifth demand which is that if one random variable is a function of another, a coefficient of association between them should take the value 1. In general $\Delta = 1$ iff $P(\phi > 1) = 1$ and $P^*(\phi > 1) = 0$ and this can occur iff $P(\phi = \infty) = 1$, i.e. iff ω is completely singular with respect to ω^*. Now if we particularise to the case of discrete random variables ω is never completely singular with respect to ω^*, not even if one random variable is a function of the other. The most we can say in this connection is that if one random variable $\tilde{y}$, say, is a function of the other, having the property that $\mu\{x : \tilde{y}(x) = k\} = 0$, for every k, then the coefficient Δ of association between them is 1. (This condition imposed on the functional relationship between $\tilde{y}$ and $\tilde{x}$ excludes the possibility that either random variable is discrete.) Thus with respect to Rényi's fifth demand, Δ is well-behaved for continuous random variables—more generally for non-atomic measures μ and ν—but not for discrete random variables.

One of the many measures of association discussed by Goodman and Kruskal [3] was suggested originally by Steffenson [10], [11] and involves essentially the idea of integrating $\phi - 1$ over the region for which $\phi > 1$. However preoccupation with the object of having a coefficient which would take the value 1 whenever the random variables concerned were functionally related resulted in a measure lacking in the operational interpretation advocated by Goodman and Kruskal. It is possible to retain operational interpretation and at the same time to have a coefficient satisfying Rényi's fifth demand simply by considering $G = \int_{\phi>1} d\omega$ as a coefficient of association between two general random variables. It is easily seen that G has the first four properties discussed above. So far as the fifth demand is concerned, G is not subject to the same criticism as Δ in the case of discrete random variables. And since $\Delta = 1$ implies $G = 1$, G is otherwise at least as well-behaved as Δ with respect to Rényi's fifth demand. However we do not suggest that G is to be preferred to Δ as a measure of association on this account. Indeed it is easy to see by looking at a 2×2 contingency table that G can take values slightly greater than $\frac{1}{2}$ with "very little" association between the factors in the table. This may be regarded as an unsatisfactory feature of G in certain contexts. Indeed the outcome of the discussion of Rényi's fifth demand is to lend further support to the argument of Goodman and Kruskal [3] that so far as measures of association are concerned, operational interpretation (either practical or theoretical) should be the main consideration, and not a pre-ordained set of postulates.

4.6. Rényi's next demand again concerns random variables $\tilde{x}$ and $\tilde{y}$ and states that a measure $\delta(\tilde{x}, \tilde{y})$ of association between them should have the following property. If f and g are Borel-measurable, 1-1 mappings of the line into itself then $\delta[f(\tilde{x}), g(\tilde{y})] = \delta(\tilde{x}, \tilde{y})$. This property is obviously possessed by both Δ and G. We can "change the variables" in the integrals defining them to obtain the required results.

4.8. The last of Rényi's seven demands is that, in the case where $\tilde{x}$ and $\tilde{y}$ are jointly normally distributed with correlation coefficient ρ, $\delta(\tilde{x}, \tilde{y})$ should equal $|\rho|$. Bell [1] has criticised this as being too restrictive and replaces it by the demand that $\delta(\tilde{x}, \tilde{y})$ be a strictly monotonic function of $|\rho|$. Since we are using these suggestions as a basis for investigating properties of Δ rather than for justifying it in any way, we content ourselves with the remark that both Δ and G satisfy Bell's demand, but not Rényi's. While it is very easy to convince oneself that this is so, a formal proof at least in the case of Δ is surprisingly difficult. Part of the reason for this difficulty emerges from the following discussion.

It has already been suggested on the basis of a naive picture that increases in association between two real random variables will be reflected in increases in $d(x)$, for most x. To be more explicit about this we refer to Figure 1. In this figure the curve (i) represents $p(y)$. The curve (ii) represents $p(y \mid x)$ for a typical x in the case of random variables $\tilde{x}$ and $\tilde{y}$ which are associated. The curve (iii) represents $p(y \mid x)$ for the same x in the case of random variables $\tilde{x}'$ and $\tilde{y}'$ having the same marginal distributions as $\tilde{x}$ and $\tilde{y}$ respectively, the same "shape" of association, but an increased degree of association. By the same shape of association we mean concentration of the bivariate distribution about the same curve in the plane and this is reflected in the fact that (ii) and

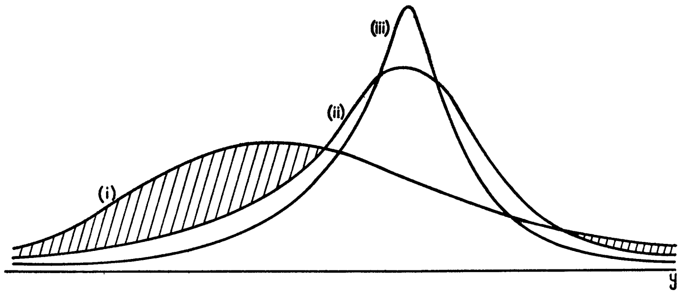

FIG. 1

(iii) have the same mode. Now if $d(x)$ is the measure of the amount of information about $\tilde{y}$ contained in x introduced in Section 2, and $d'(x)$ is the same quantity for $\tilde{y}'$, it is clear that in the situation described by Figure 1, $d'(x) > d(x)$. For $d(x)$ is equal to the shaded area in the diagram since

$$\int_{-\infty}^{\infty} [p(y \mid x) - p(y)]\, dy = 0$$

and so

$$\int_{\phi>1} [p(y \mid x) - p(y)]\, dy = \int_{\phi \leqq 1} [p(y) - p(y \mid x)]\, dy,$$

and similarly $d'(x)$ equals a larger area.

However, when we are dealing with bivariate normal distributions and considering the result of increasing the correlation coefficient ρ, Figure 1 is not appropriate because, as we increase ρ while keeping the marginal distributions fixed the mode of $p(y \mid x)$ does not remain fixed. Nevertheless $d(x)$ does increase for each x with $|\rho|$. For the following proof of this result, which replaces a much longer proof in an earlier version of the paper, the author is indebted to a referee.

LEMMA 1. *If $\Delta(\rho)$ is the above-defined coefficient of association between two jointly normally distributed random variables with correlation coefficient ρ, Δ is a monotonic increasing function of $|\rho|$.*

PROOF. Without loss of generality we may assume that each of the normal random variables $\tilde{x}$ and $\tilde{y}$ has zero mean and unit variance. Also $\Delta(-\rho) = \Delta(\rho)$, so that it is sufficient to consider $0 \leqq \rho < 1$. Let

$$d(x, \rho) = \int_{p(y|x)>p(y)} [p(y \mid x) - p(y)]\, dy,$$

where $p(y \mid x)$ is the $N(\rho x, 1 - \rho^2)$ density and $p(y)$ the $N(0, 1)$ density, so that

$$\Delta(\rho) = E[d(\tilde{x}, \rho)].$$

Now $p(y \mid x) > p(y)$ in an interval $I(\rho, x)$ represented by the line AC in Figure 2 where, when $x > 0$, $AB > BC$; when $x < 0$, $AB < BC$. Also $p(y \mid x) = p(y)$ at the end-points of $I(\rho, x)$. Hence

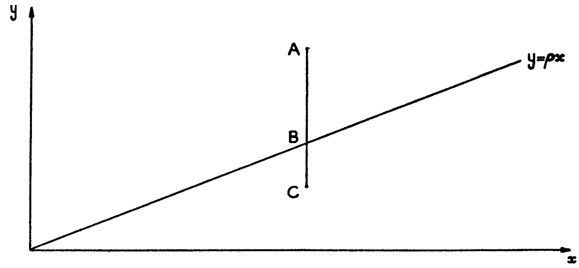

FIG. 2

$$(2\pi)^{\frac{1}{2}} \frac{\partial}{\partial \rho} d(x, \rho) = \int_{I(\rho,x)} \frac{\partial}{\partial \rho} \left[\frac{1}{(1-\rho^2)^{\frac{1}{2}}} \exp\left[-\frac{1}{2} \frac{(y-\rho x)^2}{1-\rho^2} \right] \right] dy$$

$$(1) \qquad = \frac{\rho}{(1-\rho^2)^{\frac{3}{2}}} \int_{I(\rho,x)} \psi(x, y, \rho) \exp\left[-\frac{1}{2} \frac{(y-\rho x)^2}{1-\rho^2} \right] dy$$

$$- x \int_{I(\rho,x)} \frac{\partial}{\partial y} \left(\frac{1}{(1-\rho^2)^{\frac{1}{2}}} \exp\left[-\frac{1}{2} \frac{(y-\rho x)^2}{1-\rho^2} \right] \right) dy,$$

where

$$\psi(x, y, \rho) = 1 - \rho^2 - \int_{I(\rho,x)} (y-\rho x)^2 \exp\left[-\frac{1}{2} \frac{(y-\rho x)^2}{1-\rho^2} \right] dy \Big/ \int_{I(\rho,x)} \exp\left[-\frac{1}{2} \frac{(y-\rho x)^2}{1-\rho^2} \right] dy.$$

The first of the two terms in the right hand side of (1) is positive since

$$\int_{I(\rho,x)} (y-\rho x)^2 \exp\left[-\frac{1}{2} \frac{(y-\rho x)^2}{1-\rho^2} \right] dy \Big/ \int_{I(\rho,x)} \exp\left[-\frac{1}{2} \frac{(y-\rho x)^2}{1-\rho^2} \right] dy$$

$$= E\{(\tilde{y} - \rho\tilde{x})^2 \mid \tilde{x} = x, \tilde{y} \,\varepsilon\, I(\rho, x)\} < E\{(\tilde{y} - \rho\tilde{x})^2 \mid \tilde{x} = x\} = 1 - \rho^2,$$

the inequality holding because $I(\rho, x)$ contains ρx and the conditional distribution of $\tilde{y}$ given x is symmetric about ρx.

The second term in the right hand side of (1) also is positive, since it is

$$-x \left[\frac{1}{(1-p^2)^{\frac{1}{2}}} \exp\left[-\frac{1}{2} \frac{(y-\rho x)^2}{1-\rho^2} \right] \right]_{I(\rho,x)}$$

and when $x > 0$, the square bracket term is negative because $AB > BC$; when $x < 0$ the square bracket term is positive because $AB < BC$.

Hence $(\partial/\partial\rho)d(x, \rho) > 0$ and it follows immediately that $(d/d\rho)\Delta(\rho) > 0$.

This completes the proof of the lemma. It is easier to prove that G shares this property with Δ, and we omit a proof of this statement. However while $\Delta \to 0$ as $\rho \to 0$, this is not true for G. In fact $G \to \frac{1}{2}$ as $\rho \to 0$, confirming a possibly unsatisfactory feature of G noted earlier. We may conclude from the discussion motivating its introduction and from the foregoing investigation of its properties that Δ is a respectable *average* measure of association between two *continuous* random variables. Now we have suggested that the Radon-Nikodym derivative provides an analytic interpretation of a naive picture of association in this case, that in particular the random variables are closely associated if and only if ϕ takes large values with probability near 1. (We emphasize that now we are discussing the continuous case: in the discrete case ϕ may not take very large values even when one random variable is a function of the other.) This suggestion has been advanced without any real justification though some support is lent to it by the fact that Δ, derived from considerations not directly involving ϕ, turns out to depend on ϕ in a very crucial way. More direct support is given by the following result with which we conclude this section.

LEMMA 2. $\Delta/(1 - k^{-1}) \geqq P\{\phi > k\} \geqq 1 - (1 - \Delta)k$, *for any real* $k > 1$.

The first of these inequalities shows that, if for large k, $P(\phi > k)$ is near 1, then Δ is near 1, while the second shows that if Δ is near 1, say $\Delta = 1 - \delta^2$, then for example $P(\phi > \delta^{-1}) \geqq 1 - \delta$.

PROOF.

(i) $P(\phi > k) = \int_{\phi>k} d\omega$.

Therefore $\int_{k<\phi<\infty} \phi \, d\omega^* \leqq P(\phi > k)$, and so

$$\int_{\phi>k} d\omega^* \leqq k^{-1}P(\phi > k).$$

Hence

$$\Delta = \int_{\phi>1} (d\omega - d\omega^*) \geqq \int_{\phi>k} (d\omega - d\omega^*) \geqq (1 - k^{-1})P(\phi > k),$$

the penultimate inequality holding because

$$\int_{1<\phi\leqq k} (d\omega - d\omega^*) = \int_{1<\phi\leqq k} (\phi - 1)\, d\omega^* \geqq 0.$$

(ii) $\int_{1<\phi\leqq k} d\omega = \int_{1<\phi\leqq k} \phi \, d\omega^* \leqq k \int_{1<\phi\leqq k} d\omega^*$.

Therefore

$$\Delta = \int_{1<\phi\leqq k} (d\omega - d\omega^*) + \int_{\phi>k} (d\omega - d\omega^*) \leqq (1 - k^{-1}) \int_{1<\phi\leqq k} d\omega + \int_{\phi>k} d\omega \leqq (1 - k^{-1})[1 - P(\phi > k)] + P(\phi > k).$$

It follows that $P(\phi > k) \geqq 1 - (1 - \Delta)k$, and this completes the proof of the lemma.

5. Δ as a measure of information. McMillan [7] has drawn the analogy between the communication problem from which arose information theory and the problem of parameter estimation, while Lindley [6], in a Bayesian setting, has used Shannon's mutual information as a measure of the information about a parameter contained in an experiment. In this section we shall discuss the coefficient Δ in the same spirit. We start with one or two general remarks.

We have already stated that Δ, which until now we have been regarding as a coefficient of association, may also be regarded as an average measure of information, subject to a suitable interpretation of information. If, for random variables $\tilde{x}$ and $\tilde{y}$ taking values in spaces X and Y respectively, the coefficient Δ is near 1, then most "observations" x in X should provide a lot of information about a corresponding unknown observation in Y, in the sense that to most x should correspond a "small" set Y_x (small being taken to mean that $\nu(Y_x)$ is near 0), whose conditional probability, given x, is near 1. Thus if Δ is near 1, knowledge of x should enable us relatively to "pinpoint" a corresponding unknown value y, with a small probability of error. What exactly can be said about this? The following lemma indicates what is possible, in general.

LEMMA 3. *If $\Delta = 1 - \epsilon$ then for each x in a set of probability $\geqq 1 - \epsilon/\alpha$, the set Y_x is such that $\nu(Y_x) \leqq \alpha$ and $\nu_x(Y_x) \geqq 1 - \alpha$.*

This result is essentially the same as one proved by Joshi [4] (Lemma 8, p. 126). Since Δ is the expected value of the random variable d, where

$$d(x) = \Pr\{Y_x \mid x\} - \Pr\{Y_x\}$$

and since $0 \leqq d \leqq 1$, all we are really saying is that a random variable, taking values in (0, 1) and having mean near 1, must take values near 1 with high probability. Indeed it is trivial to prove that $\Pr\{d > 1 - \alpha\} \geqq 1 - \epsilon/\alpha$ and this implies the result stated.

In the context of Bayesian estimation of an unknown parameter we wish to think of the space X as a parameter space and to emphasize this we shall now denote it by Θ and its typical point by θ. The measure μ on this space may be regarded as a prior distribution on the parameter space. Let us suppose that, for each n, we have a family $\{\nu_\theta^{(n)}, \theta \varepsilon \Theta\}$ of probability measures on a sample space $Y^{(n)}$. This family, together with μ, defines a marginal measure $\nu^{(n)}$ on $Y^{(n)}$ and a measure $\omega^{(n)}$ on $\Theta \times Y^{(n)}$. So we can define the coefficient of association Δ_n, between $\tilde{\theta}$ and $\tilde{y}^{(n)}$. Often in cases of practical interest this will be the formal mathematical description of a situation in which it is intuitively clear that we ought to be getting "more information about θ" as n increases, that in fact, Δ_n should tend to 1 as $n \to \infty$. Let up suppose that this is so. The implications regarding Bayesian estimation which follow from Lemma 3 are fairly obvious. By choosing n large enough (so large that $\Delta_n \geqq 1 - \epsilon^2$) we can ensure that, to each $y^{(n)}$ in a set of $\nu^{(n)}$-measure $\geqq 1 - \epsilon$, there corresponds a set $\Theta_y^{(n)}$ in Θ whose prior probability is less than ϵ and whose conditional probability, given $y^{(n)}$, is greater than $1 - \epsilon$; this being so for arbitrarily small positive ϵ. Thus $\Delta_n \to 1$ as $n \to \infty$ implies a kind of consistency in estimation which involves a measure rather than a metric on the parameter space. It implies that, if n is large, we can, for most "observations" $y^{(n)}$, find a very "small" Bayesian confidence set with confidence coefficient near 1.

In the context of this Bayesian approach to inference, Lindley [6] has suggested using $E(\log \phi)$ as a measure of the information provided by an experiment and he has also put forward the suggestion that experiments may be partially ordered according to how informative, in this sense, they are. It is obvious that we might partially order experiments in the same way, using Δ rather than $E(\log \phi)$ as a measure of information. While we do not wish to suggest that the use of Δ provides in general a better ordering than the use of $E(\log \phi)$, it might be argued that it provides one which is susceptible to more direct interpretation so far as Bayesian estimation is concerned. However we do not propose to press this point since if we are thinking about how informative an experiment is, in the final event we ought to be discussing the question "In what aspect of this complex notion of information are we primarily interested?" The answer to this question will decide which measure of information should be used.

REFERENCES

[1] BELL, C. B. (1962). Mutual information and maximal correlation as measures of dependence. *Ann. Math. Statist.* **33** 587–595.

[2] GEL'FAND, I. M. and YAGLOM, A. M. (1957). Computation of the amount of information about a stochastic function contained in another such function. (Russian) *Upsehi Mat. Nauk.* (N.S.) **12** 3–52 (MR **18** 981).

[3] GOODMAN, L. A. and KRUSKAL, W. H. (1954, 1959). Measures of association for cross classifications. *J. Amer. Statist. Assoc.* **49** 732–764 and **54** 123–163.

[4] JOSHI, D. D. (1959). L'information en statistique mathématique et dans la théorie des communications. *Publ. Inst. Statist. Univ. Paris* **8** 83–158.

[5] KHINCHIN, A. I. (1957). *Mathematical Foundations of Information Theory.* Dover, New York.

[6] LINDLEY, D. V. (1956). On a measure of the information provided by an experiment. *Ann. Math. Statist.* **27** 986–1005.

[7] MCMILLAN, B. (1953). The basic theorems of information theory. *Ann. Math. Statist.* **24** 196–219.

[8] RÉNYI, A. (1959). On measures of dependence. *Acta Math. Acad. Sci. Hungar.* **10** 441–451.

[9] SHANNON, C. E. (1948). A mathematical theory of communication. *Bell System Tech. J.* **27** 379–423 and 623–656.

[10] STEFFENSON, J. F. (1934). On certain measures of dependence between statistical variables. *Biometrika* **26** 250–255.

[11] STEFFENSON, J. F. (1941). On the ω test of dependence between statistical variables. *Skand. Aktuarietidskr.* **24** 13–33.