Universität Bielefeld/IMW

Working Papers
Institute of Mathematical Economics

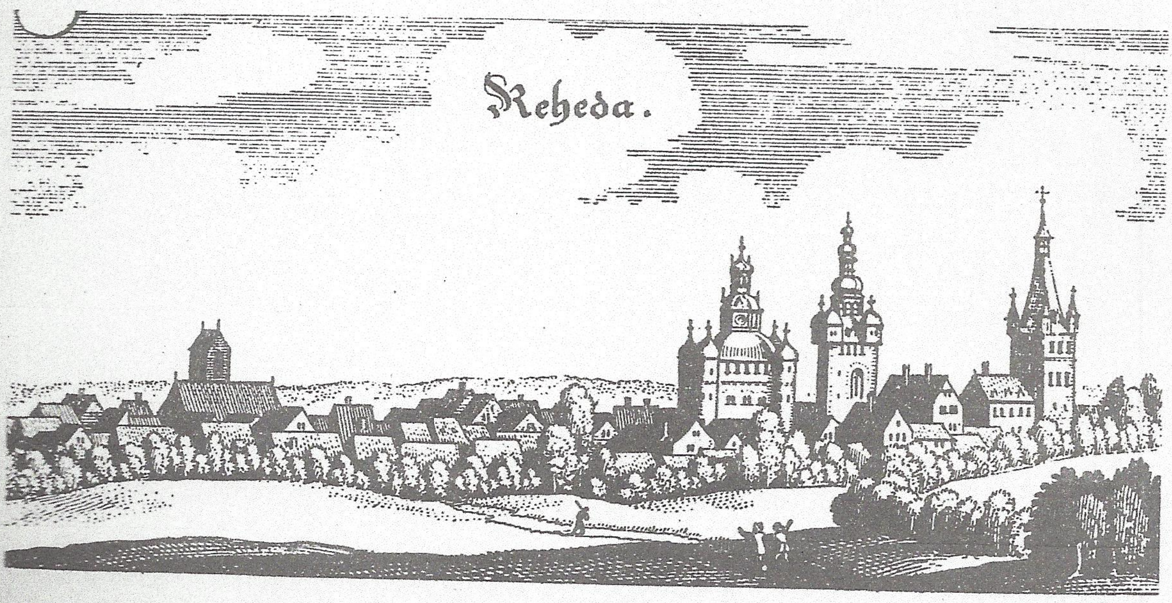
Arbeiten aus dem
Institut für Mathematische Wirtschaftsforschung

Nr. 10

Some Measures of Information Arising in Statistical Games

Hans W. Göttinger

May 1973



Anno 1600 nach Merian

Institut für Mathematische Wirtschaftsforschung
an der
Universität Bielefeld
Adresse/Address:
Schloß Rheda
484 Rheda
Bundesrepublik Deutschland
Federal Republic of Germany

SOME MEASURES OF INFORMATION ARISING IN STATISTICAL GAMES

BY HANS W. GOTTINGER

UNIVERSITY OF BIELEFELD

1. MOTIVATION

A rather general approach toward measuring information has been developed within statistical decision theory. Note that this development is unrelated to R.A. Fisher's theory of statistical estimation where the proposed measure of information (regarding statistical observations) turns out to be a particular case of Shannon's measure.

In modern statistical decision theory we are concerned with information provided by an experiment.

An experiment X is completely described by a random variable associated to some sample space giving rise to a set of conditional probability distributions for every possible value of a parameter (state of nature). Note that X might itself be of fixed sample size or of a sequential type where the experimenter may collect observations finitely many times.

To set up the problem assume you (the experimenter or the statistician) are confronted with an uncertain situation where you wish to know about the true value of a state $\omega \in \Omega$.

Of course you can make some guesses but the only way to gain knowledge about the true value of the state (e.g. to rationalize these guesses) is by performing experiments.

Let μ be some prior probability distribution of the true value ω which indicates the amount of uncertainty or ignorance on your part. (Adopt a Bayesian viewpoint that such μ always exists and is non-null.) Then the information provided by X may be more loosely described as the difference between the amount of uncertainty you attach to the prior distribution and the amount of your expected uncertainty of the posterior distribution (after having performed X), i.e. it reflects the residual value of your uncertainty (reduced).

There is an obvious connection of this situation with the structure of a statistical game in which two players are referred to as 'nature' and 'the statistician'.

Here in order to constitute the statistician's strategy the possibility of 'spying' by performing an experiment plays an important role. Note that in statistical games ω (in some finite set Ω) constitute nature's pure strategies whereas nature's mixed strategies can be identified with your prior distribution over Ω associated to some sample space. More technically, let $\mathcal{M}$ be the set of prior probability distributions μ over Ω (i.e. the space of randomized strategies for nature), define u as a nonnegative, real-valued measurable function on $\mathcal{M}$ which for obvious reasons should be concave, i.e. decreasing with increasing observations. Then u (μ) represents the amount of your uncertainty (before experimentation when your distribution over Ω is μ. Now by performing X and observing values of X you may specify a posterior distribution $\mu(X)$, then your measure of information I is determined by $I(X, \mu, u) = u(\mu) - E\left[u(\mu(X)) | \mu\right]$, where $\mu(X)$ is usually obtained by an appropriate application of Bayes' theorem.

This approach has been consistently pursued by M. de Groot [2], and somewhat earlier, by Blackwell and Girshick [1,Ch.3]. It is usually assumed for reasons of non-trivialty that most experiments provide information and that any experiment being more informative than another is also preferable to the other. Therefore, for any given uncertainty function, I is nonnegative, and also for reasons of convenience continous.
As we clearly recognize this measure of information provided by an experiment relative to the specification of u and μ naturally evolves from a model of statistical decision. Usually the determination of the uncertainty function hinges upon the loss structure of a statistical game. Every experiment X associated to a sample space generates a risk function, defined as the expected value of assigning to every decision act its numerical loss for any given state of nature. From this we learn that information evolving from a statistical decision problem generally takes into account economic type considerations of benefits and costs (via the loss function). From an economic point of view information may be regarded as a particular kind of commodity traded at a certain price yielding benefits for consumers and causing costs for producers. The economic theory of information hence is an appropriate modification of the approach used in statistical decision theory. [1] To put it in other terms, we would be interested

-----

1) I should point out here that there are two main aspects of an economic theory of information to be considered, the first is the micro-aspect which can adequately be dealt within decision theory, the second is the system theoretic aspect (information provided by an economic system) for which other tools (for example, machine, theory) might be more adequate.

in the economic aspect of usefulness of information (based on some kind of utility or loss function) rather than in the (original) physical viewpoint of transmitting and controlling information flows through a large (noisy or noiseless) communication channel. This viewpoint has been consistently advanced by J. Marschak [3] on the basis of earlier results due to Blackwell [1].

Marschak's 'value of information' attached to some experiment w.r.t. any probability distribution μ over Ω (his space of events non-controllable to the decision maker) and his benefit function $b: \Omega \times A \rightarrow R_e$ (A being the set of decision acts) is just the converse value of Blackwell's risk r(X). Note that the risk function is completely specified by a probability distribution over Ω and a loss function on $\Omega \times A$.

One can readily see the strong agreement between information provided by an experiment and the value of information by considering an experiment as a Markov matrix.

In this case null-information corresponds to identical rows in the Markov matrix, i.e. any observation made through an experiment is independent of any state of nature.

Accordingly, the risk function obtained by the less informative experiment is larger in value than the risk function obtained by the more informative experiment. It is obvious that the dual statement holds if we deal with a benefit function instead of a loss function.

In this paper we suggest various measures of information which quite naturally arise in the context of statistical games (also known as games against nature). Although these measures are strictly confined to a situation constituted by these types of games, the remarks above indicate they might be also of interest for an economic theory of information, at least from a methodological point of view.

We here emphasize mathematical aspects of the general game-theoretic situation.

2. THE CONCEPT OF INFORMATION IN A STATISTICAL GAME

We consider a game in which the statistician is able to select a decision strategy on the basis of information available to him.

Hence, let us consider a game (Ω, Y, φ) between nature and the statistician with pay off function φ. Let Ω and Y be compact metric spaces so that φ satisfies some mild continuity condition, e. g. a Lipschitzian condition in $\Omega \times Y$. It is well-known that subsets of metric spaces form a class of Borel sets, hence in defining a probability distribution on a compact metric space it is obvious that this distribution is defined on Borel sets of this space. Since every random variable X associated to a sample space (Z, Ω) induces a probability distribution μ on $Z \times \Omega$, it will be more convenient for our purposes to refer to μ as an experiment whose out comes $z \in Z$ for any $\omega \in \Omega$ are governed by the conditional distribution μ_ω with values $\mu_\omega(z)$. In order to reveal the structure of information in this game we will assume that the person must take a decision $y \in Y$ prior to the experiment μ, and by adopting a Bayesian view, μ should be known to him in choosing a Bayesian strategy which takes into account prior information in a systematic fashion.

This means that a decision $y \in Y$ is taken that minimizes the average of the pay_off value,

$$E\{\varphi(\omega, y)\} = \int_{\Omega} \varphi(\omega, y)\, d\mu_{\omega}(z).$$

Now let us consider the possibility that the statistician can obtain additional information on his decision problem by performing an <u>auxiliary</u> experiment. Hence, given a modified space of outcomes (compact metric space) Z', consider a corresponding experiment μ' with (conditional) p. d. μ'_{ω} on Z' together with $\mu_{\omega}(z|z') \equiv \mu_{\omega z'}(z)$ which is the conditional p. d. relative to z' on Z. It is in the spirit of the Bayesian approach to assume that a person can perform an auxiliary experiment μ' prior to taking a decision. In this case the average pay-off value will be

$$\int_{Z'} \left\{ \min_{y} \int_{\Omega} \varphi(\omega, y)\, d\mu_{\omega z'}(z) \right\} d\mu'_{\omega}(z')$$

We may call the difference

$$(1) \quad v(\mu, \mu') = \min_{y} \int_{\Omega} \varphi(\omega, y)\, d\mu_{\omega}(z) - \int_{Z'} \left\{ \min_{y} \int_{\Omega} \varphi(\omega, y)\, d\mu_{\omega z'}(z) \right\} d\mu'_{\omega}(z')$$

<u>the value of information</u> in experiment μ generated by experiment μ' This value does not change if we add or subtract some positive amount. By adopting a Bayesian strategy, the statistician would attempt to choose a decision which minimizes his expected loss in terms of the pay-off value, i. e.

$$(2) \quad \psi(\omega) = \min_{y} \varphi(\omega, y),$$

so that the value is given by

$$w(\mu) = \min_{y} \int_{\Omega} \varphi(\omega, y) \, d\mu_{\omega}(z)$$

which represents his expected loss resulting from the uncertainty of the outcome of experiment μ.[2)]

For the next considerations assume that the loss is normalized by the condition $\max_{\omega, y} |\varphi(\omega, y)| = 1$.

In the sequel of this section we put every thing into discrete terms and assume the sets Z, Y and Z' to be finite.

In set-theoretic notation, let

$$\Omega = \{\omega_i\}, Y = \{y_j\}, Z' = \{z_k'\}, \mu = \{\mu_i\}, \quad \varphi = \{\varphi_{ij}\}, \mu' = \{\mu'_k\}.$$

$$(i = 1, \ldots m; \; j = 1, \ldots n; \; k = 1, \ldots \ell)$$

In short-hand notation we replace $\mu_{\omega z'}(Z)$ by μ_i^k omitting ω since no ambiguity will arise.

------

2) An interesting question arises: what could be the relationship between the entropy of an experiment and the value of an experiment in terms of (2)?

For simplicity, let $w = w(\mu)$, $v = v(\mu', \mu)$, define $h = h(\mu)$ as the entropy of the experiment μ corresponding to (1), and let $I = I(\mu', \mu)$ be the amount of information μ generated by experiment μ', corresponding to (1).

Thus,

(3) $h = \sum_i \mu_i \log_n (1/\mu_i)$ and

(4) $I = \sum_i \mu_i \log_n (1/\mu_i) - \sum_k \mu'_k \sum_i \mu_i^k \log_n (1/\mu_i^k)$

<u>THEOREM</u>

The relationships between (1) and (4) and between (2) and (3) respectively are determined by the following simple set of inequalities

(i) $\log_n (1/1-w) \leq h$

(ii) $\frac{1}{2} v^2 \leq \frac{1}{2} \quad (v/\max_{i,i',j} |\varphi_{ij} - \varphi_{i'j}|) \leq I$

(iii) $v\log_n \quad (v/1-v) \leq I, \quad v \geq 2/3.$

<u>Outline of Proof.</u> 1. We first verify inequality (i). Let z' be a particular outcome in Z' and let $M_{z',m} = \{\mu_i : h(\mu) \leq z'\}$. Denote by $\bar{\Phi}_{m,n}$ the class of $m \times n$ matrices φ satisfying the normalization condition, and by change of notation as introduced above, also the condition $\min_j \varphi_{ij} = 0$ for all i. Accordingly, let $\bar{\Phi}^o_{m,n} \subseteq \bar{\Phi}_{m,n}$ be the subclass of those characteristic matrices all of whose elements are zeros or ones. Set $w(\mu) = w(\mu, \varphi)$ and let $s = s(z')$ be the least upper bound

$$s = \sup_{m,n} \max_{\mu \in M_{z'm}} \max_{\varphi \in \bar{\Phi}_{m,n}} w(\mu, \varphi)$$

Now one can easily verify that w.r.t. $\bar{\Phi}^o_{m,n}$ there exists a l. u. b. s' such that

$$\sup_{m,n} \max_{\mu \in M_{z'm}} \max_{\varphi \in \bar{\Phi}^o_{m,n}} \min_{j} (1 - \sum_{i \in \Gamma_j(\varphi)} \mu_i)$$

$$= \quad s' = s = \sup_{m,n} \max_{\mu \in M_{z,'m}} \max_{\varphi \in \bar{\Phi}^o_{m,n}} w(\mu, \varphi) \quad ,$$

where $\Gamma_j(\varphi) = \{ i : \varphi_{ij} = 0 \}$

Consider a particular $\varphi \in \bar{\Phi}^o_{m,n}$ with $\varphi_{ij} = 0$ for all i, then we may compute

$$(*) \qquad s = \sup_m \max_{\mu \in M_{z',m}} \min_i (1 - \mu_i)$$

Let

$$(**) \qquad s_m = \max_{\mu \in M_{z',m}} \min_i (1 - \mu_i) \text{ and let}$$

$$\hat{\mu}_m = \{ \mu_{mi} \}_{i=1}^m$$

be the maximizing experiment among all experiments $\mu \in M_{z'm}$ so that $\hat{\mu}_m = \max_i \mu_{mi} = 1 - s_m = r_m$

Denote by q and q_m the integral number parts of the numbers $1/(1-s)$ and $1/(1-s_m)$ corresponding to (*), (**) respectively. Then we can easily verify the inequalities

$$\log_n (1/1-s_m) = \log_n (1/r_m) \leq q_m r_m \log_n (1/r_m)$$

$$+ (1-q_m \cdot r_m) \log_n (1/(1-q_m \cdot r_m))$$

$$= \min_{0 \leq \mu_i \leq r_m} \sum \mu_i \log_n (1/\mu_i) \leq h(\mu_m) \leq z', \text{ for } m \geq q.$$

Hence $\log_n (1/(1-s)) \leq z'$.

Now let an experiment $\mu = \{\mu_i\}_{i=1}^{m}$, $\mu \in M_{z',m}$ and a pay-off matrix $\varphi \in \bar{\Phi}_{m,n}$ be given. By definition of w and s we have $w(\mu,\varphi) \leq s(z')$ so that

$$h(\mu) \geq \log_n (1/(1-s(z')) \geq \log_n (1/(1-w(\mu,\varphi))$$

2. Since inequality (iii) is a much stronger statement than (ii) we omit the proof of (ii), and now outline the proof of (iii).

Evaluating two experiments μ and ν is equivalent to evaluating the matrix $\eta = [\nu_k \mu_i^k]$, $i = 1, \ldots, m$; $k = 1, \ldots, \ell$.
Let us specify $v_o = \sup_n \max_{\varphi \in \Phi_{m,n}} (\nu,\mu,\varphi)$, substituting

$v(\nu,\mu,\varphi)$ for $v(\nu,\mu)$. By $\bar{\Phi}'_{m,\ell}$ we denote the class of matrices

$\varphi \in \bar{\Phi}_{m,\ell}$ such that

$$w(\mu,\varphi) = \sum_i \varphi_{ij} \mu_i \quad (j = 1, \ldots, m),$$

$$\min_j \sum_i \varphi_{ij} \mu_i^k = \sum_i \varphi_{ik} \mu_i^k \quad (k = 1 \ldots, \ell).$$

Then we may conclude without great difficulty that

$v_o = \max_{\varphi \in \bar{\Phi}'_{m,\ell}} v(\nu,\mu,\varphi)$. Let $\hat{\varphi} = [\hat{\varphi}_{ij}]$ be the maximizing

matrix of $\bar{\Phi}'_{m,\ell}$.

The following notation is convenient:

$$\varepsilon_i^k = \mu_i^k / \mu_i$$

$$\rho_{ik} = \mu_i \nu_k (1 - \hat{\varphi}_{ik}), \quad \sigma_{ik} = \mu_i \nu_k \hat{\varphi}_{ik})$$

$$\gamma = \sum_{i,k} \rho_{ik} = 1 - \sum_{i,k} \sigma_{ik}$$

Then we have

$$v_o = \sum_{i,k} \rho_{ik} (\varepsilon_i^k - 1) = \sum_{i,k} \sigma_{ik} (1 - \varepsilon_i^k)$$

By straight-forward application of Jenson's inequality we get

$$\sum_{i,k} \rho_{ik} \varepsilon_i^k \log_n \sum_{i,k} (\rho_{ik}/\gamma) \varepsilon_i^k + \sum_{i,k} \sigma_{ik} \varepsilon_i^k \cdot \log_n \sum_{i,k}$$

$$(\sigma_{ik}/1 - \gamma) \varepsilon_i^k \leq \sum_{i,k} \rho_{ik} \varepsilon_i^k \log_n \varepsilon_i^k + \sum_{i,k} \sigma_{ik} \varepsilon_i^k \log_n \varepsilon_i^k = I.$$

The first part of this inequality can be conveniently decomposed into

$$(v_o + \gamma) \log_n((v_o + \gamma) \quad) + (1 - v_o - \gamma) \log_n ((1 - v_o - \gamma)/(1 - \gamma)) =$$

$$[(v_o + \gamma) \log_n(v_o + \gamma) + (1 - v_o - \gamma) \log_n(1 - v_o - \gamma) +$$

$$[- (v_o + \gamma) \log_n \gamma - (1 - v_o - \gamma) \log_n(1 - \gamma)] = A(\gamma) + B(\gamma).$$

There from we derive for $0 \leq \gamma \leq 1 - v_o$ the inequalities

$$A(\gamma) \geq v_o \log_n v_o \; (1 - v_o) \log_n (1 - v_o)$$

$$B(\gamma) \geq -\log_n (1 - v_o),$$

which together yields

$$I \geq v_o \log_n \frac{v_o}{1 - v_o}.$$

3. AN INFORMATIONAL METRIC

It seems natural in view of the approach adopted here to use the metric in a compact metric space for constructing some notion of informational distance in a subjective sense.

From the Bayesian point of view we may assume that a person - before performing an experiment - knows about a particular presentation of nature's pure strategy, given by a point $\omega \in \Omega$.

Now, after having performed the experiment, this person observes the actual state to be $\omega_0 \in \Omega$. Let $\delta : \Omega \times \Omega \rightarrow Re$ be the ordinary metric such that δ associates a real number $\delta(\omega, \omega_0)$ with every pair (ω, ω_0) of elements of Ω. Then $\delta(\omega, \omega_0)$ represents a change of the informational state of a person (change of belief) in terms of a distance, satisfying well-known conditions of a numerical metric. Given a set of experiments $X_1 \ldots, X_n$, sequentially designed, the search problem of a person would consist in observing a sequence of points $\omega_1, \ldots, \omega_n$ approaching the true state $\omega_0 \in \Omega$. Now in the context of a statistical game (Ω, Y, φ) there is an interesting way to reformulate an informational metric in terms of an economic value of information. Let us assume a person by observing ω takes a decision y out of his decision set

$$\mathcal{D}(\omega) = \{y: \varphi(\omega, y) = \psi(\omega)\}$$

with $\psi(\omega) = \min_{y \in Y} \varphi(\omega, y)$.

Hence, the value of information of a person selecting a decision y on the basis of observation ω compared to a true state ω_o can be given as a number

$$m(\omega_o,\omega) = \max_{y \in \mathcal{D}(\omega)} \varphi(\omega_o, y) - \min_{y} (\omega_o, y)$$

Let δ be an ordinary metric such that

$$\delta(\omega_1, \omega_2) = \max_{y} | \varphi(x_1, y) - \varphi(x_2, y)| .$$

Now, if $\delta(\omega_o,\omega_n)$ converges to zero for n sufficiently large, by implication $m(\omega_o, \omega_n)$ converges to zero for n sufficiently large. Hence, the inequality

(*) $\quad m(\omega_o, \omega) \leq 2\delta(\omega_o, \omega)$ holds.

Let $\mathcal{M}$ denote the space of all probability distributions μ over Ω, characterizing randomized strategies for nature. Then the pay-off value in the game $(\mathcal{M}, Y,\lambda)$ is given by $\lambda(\mu, y) = \int_{\Omega} \varphi(\omega, y) \; d\mu$. Accordingly, we may introduce in $\mathcal{M}$ a metric δ in $\mathcal{M}$ which associates with every pair (μ_1, μ_2) a number $\delta(\mu_1,\mu_2)$.

In order to obtain a value of information in this case we may assume accordingly that the a priori probability distribution known to the player is not the true distribution μ_o but some distribution close to it. Suppose he knows the conditional distribution $\eta(z|\mu, z')$, if the prior distribution is μ and the outcome of the experiment V is z'. Assume that the player starts with some distribution μ which is close to the true distribution μ_o. Then by taking a decision a priori the player minimizes

$$\int_{\Omega} \varphi(\omega, y) \, d\mu(z) \quad \text{w.r.t.} \quad y \in Y.$$

If the decision is taken after the experiment ν has been performed then he minimizes (setting $\eta(z|\mu, z') \equiv \eta_{\mu, z'}(z)$)

$$\int_{\Omega} (\omega, y) \, d\eta_{\mu, z'}(z) \quad \text{w.r.t.} \quad y \in Y.$$

Taking into account that the true distribution is μ_o we may compute the value of information in the game $(\mathcal{M}, Y, \lambda)$ by

$$\bar{v} = \bar{v}(\mu, \nu) = \int_{\Omega} \varphi(\omega, y) \, d\mu(z) - \int_{Z'} \int_{\Omega} \varphi(\omega, y_z)$$

$$d\eta_{\mu_o, z'}(z) d\nu(z'); \; \bar{y} \in \mathcal{D}(\mu), \quad \bar{y}_{z'} \in \mathcal{D}(\eta_{\mu_o, z'}).$$

Since the decisions $\bar{y}$, $\bar{y}_{z'}$ need not be uniquely determined, the quantity $\bar{v}(\mu, \nu)$ may also not be unique. Let ξ be some other experiment which does not generate true distribution μ_o, and let $v = v(\mu, \nu) = \int_{\Omega} \varphi(\omega, \bar{y}) \, d\mu(z) - \int_{Z'} \int_{\Omega} \varphi(\omega, \bar{y}_{z'})$
$d\eta_{\mu, z'}(z) \, d\xi(z')$.

By comparing the values of information provided by two experiments ν and ξ we form the relations

$$\max |\bar{v} - v| = m(\mu_o, \mu) - \int_{Z'} m(\eta_{\mu_o, z'}, \eta_{\mu, z'}) \, d\nu(z')$$

$$\leq \max \left[m(\mu_o, \mu), \int_{Z'} m(\eta_{\mu_o z'}, \eta_{\mu, z'}) d\nu(z') \right].$$

Consequently, with the help of inequality (*) we can write

$$|\bar{v}-v| \leq 2\max\left[\delta(\mu_o,\mu), \quad \int_{Z'} \delta(\eta_{\mu_o,z'}\ \eta_{\mu,z'})\ d\gamma(z')\right].$$

The metric in $\mathcal{M}$ is given by

$$\delta(\mu_o, \mu) = \max_y \left| \int_{\Omega} \varphi(\omega, y)\ d\mu_o(z) - \int_{\Omega} \varphi(\omega, y)\ d\mu(z) \right|.$$

$$= \max_y \left| \int_{\Omega} \varphi(\omega, y) \left[d\mu_o(z) - d\mu(z) \right] \right|.$$

By imposing a rather mild continuity condition on φ, i. e. by requiring that φ satisfies a Lipschitzian condition in the first variable with a constant k, we have the relation

$$\delta(\mu_o,\mu) \leq k\,\delta(\mu_o,\mu).$$

4. AN ALTERNATIVE MEASURE OF INFORMATION

Consider the game (Ω, Y, φ). To obtain information about the true state ω_o may result in the specification of a set $\mathcal{H} \subset \Omega$ to which ω_o belongs. Hence, we introduce the class $H = \{\mathcal{H}\}$ of closed subsets of Ω. In this case the game is defined by the triple $(\mathcal{H}, Y, \varphi)$, and the pay-off value will result from the player's decision on the basis of his representation $\mathcal{H}$. Here it might be useful to introduce the concept of diameter value of a set $\mathcal{H}$ which characterizes the player's losses resulting from his ignorance about the true state $\omega_o \in \Omega$. The diameter value $D(\mathcal{H})$ is specified with respect to the strategy of a player with representation $\mathcal{H}$. The true representation is the singleton $\{\omega_o\}$.

If the player takes a decision that minimizes a certain pay-off-value $\varphi(\mathcal{H}, y)$ the diameter value is given by

$$D(\mathcal{H}) = \max_{\omega_o \in \mathcal{H}} m(\{\omega_o\}, \mathcal{H}) = \max_{\omega_o \in \mathcal{H}} \max_{y \in \mathcal{D}(\mathcal{H})} \left[\varphi(\{\omega_o\}, y) - \min_{y} \varphi(\{\omega_o\}, y)\right]$$

where $\mathcal{D}(\mathcal{H}) = \{ y : \varphi(\mathcal{H}, y) = \min \}$

There are three main decision criteria available for the player which he might consider:

(a) $\varphi_1 (\mathcal{H}, y) = \lambda \max_{\omega \in \mathcal{H}} \varphi(\omega, y) + (1-\lambda) \min_{\omega \in \mathcal{H}} \varphi(\omega, y), \quad (0 \leq \lambda \leq 1)$

(Hurwicz strategy or minimax strategy for $\lambda = 1$)

(b) $\varphi_2 (\mathcal{H}, y) = \max_{\omega \in \mathcal{H}} \left[\varphi(\omega, y) - \min \varphi(\omega, y)\right]$

(Minimax strategy of losses)

(c) Let $\mu(z|\mathcal{H})$ be the conditional distribution depending on a parameter $\mathcal{H}$, the payoff-value corresponding to a Bayes strategy is

$$\varphi_3 (\mathcal{H}, y) = \int_{\mathcal{H}} \varphi(\omega, y)\, d\mu(z|\mathcal{H}).$$

Except for (b), $D(\mathcal{H})$ does not increase whenever $\mathcal{H}$ increases in all other cases.

In particular, define

$$D_o(\mathcal{H}) = \max_{\omega_o \in \mathcal{H}} \max_{\omega \in \mathcal{H}} m(\omega_o, \omega).$$

This construction relates to the strategy of a player who, by knowing that $\omega_o \in \mathcal{H}$, selects any point $\omega \in \mathcal{H}$ at random basing his decision on the assumption that this point is true. Also $D_o(\mathcal{H})$ increases with $\mathcal{H}$. Now in case that $D(\mathcal{H})$ increases with $\mathcal{H}$ (where we can only take a restricted class of available decision criteria indicated above), this value can be interpreted as an analogue of the entropy value. This interpretation makes sense in view of the following situation.

Suppose that a player before taking a decision with a representation $\mathcal{H}$ performs an auxiliary experiment ν yielding a representation $\mathcal{H}'$ where $\mathcal{H}'$ is considered to be a closed subset of Ω having a non-empty intersection with $\mathcal{H}$. When receiving a message '$\omega_o \in \mathcal{H}'$' the player forms a representation $\mathcal{H} \cap \mathcal{H}'$ with diameter value $D(\mathcal{H} \cap \mathcal{H}') \leq D(\mathcal{H})$

The difference

$$D(\mathcal{H}) - D(\mathcal{H} \cap \mathcal{H}')$$

is nonnegativ and is a suitable analogue of the information value of an experiment μ generated by an experiment ν.

References

[1] D. Blackwell and M.A. Girshick, Theory of Games and Statistical Decisions, Wiley: New York 1954.

[2] M. de Groot, 'Uncertainty, Information and Sequential Experiments'. Ann. Math. Stat. 33, 1962, 404-418.

[3] J. Marschak, 'Economics of Information Systems', in Frontiers of Quantitative Economics (M.D. Intriligator, ed.), North-Holland: Amsterdam 1970.

Notes added in proof. Note that the existence of the metric is presupposed. Since we consider the metric as some measure of informational distance in a subjective sense it would be interesting to pursue proberties of the underlying qualitative structure. Various structures of this sort have recently been examined by R.D. Luce (still unpublished notes) in terms of proximity structures and extensive measurement.
By choosing a metric as a measure of informational distance in a statistical game one can fully exploit the generality of metric spaces. If necessary one can generalize the metric to a probabilistic metric constituting uncertainty about the true distance (see B. Schweizer and A. Sklar, 'Statistical Metric Spaces', Pacific Jour. Math. 10, 1960, 313 - 333.)