Kybernetika

Harish C. Taneja; R. K. Tuteja

Characterization of a quantitative-qualitative measure of inaccuracy

Kybernetika, Vol. 22 (1986), No. 5, 393--402

Persistent URL: `http://dml.cz/dmlcz/124578`

KYBERNETIKA — VOLUME 22 (1986), NUMBER 5

CHARACTERIZATION OF A QUANTITATIVE-QUALITATIVE MEASURE OF INACCURACY

H. C. TANEJA, R. K. TUTEJA

A quantitative-qualitative measure of inaccuracy is suggested and is characterized under a set of assumptions. Some properties of the new measure are also discussed.

1. INTRODUCTION

Let $P = (p_1, p_2, \ldots, p_N)$, $0 \leqq p_i \leqq 1$, $\sum_{i=1}^{N} p_i = 1$ be the probability distribution associated with a finite system of events $E = (E_1, E_2, \ldots, E_N)$ representing the realization of some experiment. The different events E_i depend upon the experimenter's goal or upon some qualitative characteristic of the physical system taken into consideration, that is, they have different weights, or utilities. In order to distinguish the events $E_1, E_2, \ldots, E_N$ with respect to a given qualitative characteristic of the physical system taken into account, ascribe to each event E_i a non-negative number u_i $(\geqq 0)$ directly proportional to its importance and call u_i the utility of the event E_i. Then the weighted entropy [1] of the experiment E is defined as

$$I(P; U) = -\sum_{i=1}^{N} u_i p_i \log p_i \,. \tag{1.1}$$

Now let us suppose that the experimenter asserts that the probability of the ith outcome E_i is q_i, whereas the true probability is p_i, with $\sum_{i=1}^{N} q_i = \sum_{i=1}^{N} p_i = 1$. Thus, we have two utility information schemes:

$$S = \begin{bmatrix} E_1 E_2 \ldots E_N \\ p_1 p_2 \ldots p_N \\ u_1 u_2 \ldots u_N \end{bmatrix}; \quad 0 \leqq p_i \leqq 1 \,, \quad u_i \geqq 0 \,, \quad \sum_{i=1}^{N} p_i = 1 \,, \tag{1.2}$$

of a set of N events after an experiment, and

$$(1.3)\qquad S^0 = \begin{bmatrix} E_1E_2 \dots E_N \\ q_1q_2 \dots q_N \\ u_1u_2 \dots u_N \end{bmatrix}; \quad 0 \leqq q_i \leqq 1\,, \quad u_i \geqq 0\,, \quad \sum_{i=1}^{N} q_i = 1$$

of the same set of N events before the experiment.

In both the schemes (1.2) and (1.3) the utility distribution is the same because we assume that the utility u_i of an outcome E_i is independent of its probability of occurrence p_i, or predicted probability q_i; u_i is only a "utility" or value of the outcome E_i for an observer relative to some specified goal (refer to [5]).

The quantitative-qualitative measure of relative information [8, 9], that the scheme (1.2) provides about the scheme (1.3), is

$$(1.4)\qquad I(P \mid Q; U) = \sum_{i=1}^{N} u_i p_i \log (p_i/q_i)\,.$$

The measure (1.4), in some sense, can be taken as a measure of the extent to which the forecasts $q_1, q_2, \dots, q_N$ differ from the corresponding realizations $p_1, p_2, \dots, p_N$ in a goal oriented experiment $E = (E_1, E_2, \dots, E_N)$. When the utilities are ignored, that is, $u_i = 1$ for each i, the measure (1.4) reduces to the Kullback's measure of relative information [4]. Consider

$$I(P; U) + I(P \mid Q; U) = - \sum_{i=1}^{N} u_i p_i \log p_i + \sum_{i=1}^{N} u_i p_i \log (p_i/q_i) =$$
$$= - \sum_{i=1}^{N} u_i p_i \log q_i\,;$$

and let it be denoted by $I(P; Q; U)$. Thus

$$(1.5)\qquad I(P; Q; U) = - \sum_{i=1}^{N} u_i p_i \log q_i\,.$$

When the utilities are ignored, then (1.5) reduces to Kerridge's inaccuracy [3]. Therefore (1.5) can be viewed as a measure of the inaccuracy associated with the statement of an experimenter made in context with a goal oriented experiment. We can consider (1.5) as a quantitative-qualitative measure of inaccuracy associated with the statement of an experimenter. When $p_i = q_i$, for each i, then (1.5) reduces to (1.1), the weighted entropy [1].

In the next section, we derive afresh the measure (1.5) under a set of intuitively reasonable assumptions.

2. THE QUANTITATIVE-QUALITATIVE MEASURE OF INACCURACY

Let $I(p_1, p_2, \dots; q_1, q_2, \dots; u_1, u_2, \dots)$ be the measure of inaccuracy associated with the goal oriented experiment $E = (E_1, E_2, \dots)$. In order to characterize the I function we consider the following

A_1. The function I is continuous with respect to its arguments p_i's, q_i's and u_i's. A_2. When N equally likely alternatives, each having the same utility u, are stated to be equally likely the inaccuracy is a monotonic increasing function of N.
A_3. If a statement is broken down into a number of subsidiary statements, then inaccuracy of the original statement is the weighted sum of the inaccuracies of the subsidiary statements.

For example, we must have

$$I(p_1, p_2, p_3; q_1, q_2, q_3; u_1, u_2, u_3) =$$
$$= I\left(p_1, p_2 + p_3; q_1, q_2 + q_3; u_1, \frac{p_2u_2 + p_3u_3}{p_2 + p_3}\right) +$$
$$+ (p_2 + p_3)\, I\left(\frac{p_2}{p_2 + p_3}, \frac{p_3}{p_2 + p_3}; \frac{q_2}{q_2 + q_3}, \frac{q_3}{q_2 + q_3}; u_2, u_3\right).$$

A_4. The inaccuracy of a statement is unchanged if two alternatives about which the same assertion is made are combined.

For example,

$$I(p_1, p_2, p_3; q_1, q_2, q_2; u_1, u_2, u_3) =$$
$$= I\left(p_1, p_2 + p_3; q_1, q_2; u_1, \frac{p_2u_2 + p_3u_3}{p_2 + p_3}\right).$$

A_5. The inaccuracy of a statement is directly proportional to the utilities of the outcomes.

For example, for every non-negative λ, we must have

$$I(p_1, p_2, p_3; q_1, q_2, q_3; \lambda u_1, \lambda u_2, \lambda u_3) =$$
$$= \lambda I(p_1, p_2, p_3; q_1, q_2, q_3; u_1, u_2, u_3).$$

All these A_1 to A_4 are just modifications of Kerridge's inaccuracy assumptions and A_5 is the monotonicity law expressed by the utilities.

In the following theorem we characterize the measure of inaccuracy associated with this system. The proof is on the same lines as in the characterization of Kerridge's inaccuracy [3].

Theorem 1. The only function satisfying the axioms A_1 to A_5 is

$$I(P; Q; U) = -K \sum_i u_i p_i \log q_i, \tag{2.1}$$

where K is an arbitrary positive number and the logarithm base is any number greater then one.

Proof. It is not difficult to verify that the function (2.1) satisfy axioms A_1 to A_5. Now we prove that any function satisfying these axioms must be of the form (2.1).

Consider the case when there are s^m alternatives with utilities $u_1, u_2, \ldots$, which are asserted to be equally likely, they may or may not be so. Then by axiom A_4

$$I(p_1, p_2, \ldots; s^{-m}, s^{-m}, \ldots,; u_1, u_2, \ldots) = I(1; s^{-m}; \bar{u})$$

where $\bar{u} = \sum_i u_i p_i$.

By axiom A_5

$$I(1; s^{-m}; \bar{u}) = \bar{u}\, I(1; s^{-m}; 1)\,. \tag{2.2}$$

Let $I(1; s^{-m}; 1) = A(s^m)$. Now $A(s^m)$ is independent of p_i's, the true probabilities, therefore, $A(s^m)$ remains unchanged if we replace the true value of p_i's by s^{-m} for all i.

By axiom A_3, $A(s^m) = m\, A(s)$. Using the continuity axiom A_1 and the monotonic character of $A(s)$, we get, (refer to [7, p. 82]), $A(s) = K \log s$, where $K\,(>0)$ is an arbitrary constant.

Consider the case when all the q_i's are rational. They can be then expressed in the form $q_i = n_i/N$, where n_i's are integers and $N = \sum_i n_i$.

By axiom A_3

$$I(p_1, p_2, \ldots; n_1/N, n_2/N, \ldots; u_1, u_2, \ldots) + \sum_i p_i\, I(1; 1/n_i; u_i) \equiv$$
$$\equiv I(1; 1/N; \sum_i p_i u_i)\,,$$

or

$$\begin{aligned} I(p_1, p_2, \ldots; n_1/N, n_2/N, \ldots; u_1, u_2, \ldots) &= \\ = I(1; 1/N; \sum_i p_i u_i) - \sum_i p_i I(1; 1/n_i; u_i) &= \\ = \sum_i p_i u_i\, I(1; 1/N; 1) - \sum_i p_i u_i\, I(1; 1/n_i; 1) &= \\ = K(\sum_i p_i u_i) \log N - K \sum_i p_i u_i \log n_i &= \\ = -K \sum_i p_i u_i \log (n_i/N) = -K \sum_i p_i u_i \log q_i\,. \end{aligned}$$

By the continuity assumption A_1, this holds for all q_i, not only for rational values. □

We shall assume $K = 1$ and take logarithm to the base '2'. We define

$$I(P; Q; U) = -\sum_{i=1}^{N} u_i p_i \log q_i\,, \quad 0 \leqq p_i, q_i \leqq 1\,, \quad u_i \geqq 0\,, \tag{2.3}$$
$$\sum_{i=1}^{N} p_i = \sum_{i=1}^{N} q_i = 1\,,$$

as the quantitative-qualitative measure of inaccuracy associated with the statement of an experimenter who asserts the probabilities of the various outcomes $E_1, E_2, \ldots, \ldots, E_N$ with utilities $u_1, u_2, \ldots, u_N$, as $q_1, q_2, \ldots, q_N$ whereas the true probabilities are $p_1, p_2, \ldots, p_N$.

The absence of a goal implies the absence of a utility measure, that is, the various events are no longer different from a qualitative point of view. The utilities $u_1, u_2, \ldots$ $\ldots, u_N$ in (2.3) are equal to each other; in order to completely avoid their influence we put $u_1 = u_2 = \ldots = u_N = 1$. In this case, (2.3) becomes

$$I(P; Q; U) = -\sum_{i=1}^{N} p_i \log q_i = I(P; Q)$$

which is exactly Kerridge's inaccuracy [3].

When in a goal-directed experiment all events have equal utilities $u_1 = u_2 = \ldots$ $\ldots = u_N = u$. Then (2.3) becomes

$$I(P; Q; U) = -u \sum_{i=1}^{N} p_i \log q_i = u\, I(P; Q)\,,$$

which expresses the increase or decrease of the quantitative-qualitative measure of inaccuracy according to the common utility 'u' of the event.

In particular case when all events have zero utilities with regard to the goal pursued we get a total quantitative-qualitative measure of inaccuracy $I(P; Q; U) = 0$, even if Kerridge's inaccuracy is not zero.

The quantitative-qualitative measure of inaccuracy is also zero if, $p_i = q_i = 1$ for one value and consequently zero for all other i, whatever the utilities $u_i \geqq 0$, $(i = 1, 2, \ldots, N)$ may be.

There is an infinite value of $I(P; Q; U)$ if $q_i = 0$, $p_i \neq 0$, $u_i \neq 0$ for any i.

3. PROPERTIES OF THE QUANTITATIVE-QUALITATIVE MEASURE OF INACCURACY

Following are some of the important properties satisfied by the measure $I(P; Q; U)$:

(1) The measure $I(P; Q; U)$ is non-negative, i.e. $I(P; Q; U) \geqq 0$.

(2) The measure $I(P; Q; U)$ is a symmetric function of its arguments, that is, $I(P; Q; U)$ remains unchanged if the elements of P, Q and U are arranged in the same way so that one to one correspondence among them is not changed.

(3) The measure $I(P; Q; U)$ is a continuous function of its arguments.

(4) The measure $I(P; Q; U)$ satisfies the generalized weighted additivity; i.e.

$$I(P * P'; Q * Q'; U * U') = \bar{U}'\, I(P; Q; U) + \bar{U}\, I(P'; Q'; U')\,,$$

where

$$\begin{aligned} P * P' &= (p_1p'_1, \ldots, p_1p'_M; \ldots; p_Np'_1, \ldots, p_Np'_M)\,, \\ Q * Q' &= (q_1q'_1, \ldots, q_1q'_M; \ldots; q_Nq'_1, \ldots, q_Nq'_M)\,, \\ U * U' &= (u_1u'_1, \ldots, u_1u'_M; \ldots; u_Nu'_1, \ldots, u_Nu'_M) \end{aligned}$$

and

$$\bar{U} = \sum_{i=1}^{N} u_i p_i\,, \quad \bar{U}' = \sum_{j=1}^{M} u'_j p'_j\,.$$

(5) The measure $I(P; Q; U)$ satisfies the branching property as follows:

$$I(p_1, p_2, \dots, p_N; q_1, q_2, \dots, q_N; u_1, u_2, \dots, u_N) =$$

$$= I\left(p_1 + p_2, \dots, p_N; q_1 + q_2, \dots, q_N; \frac{u_1 p_1 + u_2 p_2}{p_1 + p_2}, \dots, u_N\right) +$$

$$+ (p_1 + p_2) I\left(\frac{p_1}{p_1 + p_2}, \frac{p_2}{p_1 + p_2}; \frac{q_1}{q_1 + q_2}, \frac{q_2}{q_1 + q_2}; u_1, u_2\right).$$

(6) The minima of the quantitative-qualitative measure of inaccuracy $I(P; Q; U)$ exists at $q_i = u_i p_i / \sum_{i=1}^{N} u_i p_i$, the normalized preferences if the events E_i, $i = 1, 2, \dots, N$.

The properties (1) to (5) can be verified very easily, however, to prove the property (6), we give the following theorem:

Theorem 2. For fixed u_i, p_i, the minima of $I(P; Q; U)$ exists at $q_i = u_i p_i / \sum_{i=1}^{N} u_i p_i$, the normalized preferences of the events E_i's, $i = 1, 2, \dots, N$.

Proof. We are to find the extreme points for the function

$$I(P; Q; U) = -\sum_{i=1}^{N} u_i p_i \log q_i, \quad u_i \geqq 0, \quad 0 \leqq p_i, q_i \leqq 1,$$

for fixed u_i, p_i, under the condition $\sum_{i=1}^{N} q_i = 1$.

Using the method of Lagrange's multipliers, set

$$F(q_1, q_2, \dots, q_N; \lambda) = -\sum_{i=1}^{N} u_i p_i \log q_i + \lambda\left(\sum_{i=1}^{N} q_i - 1\right), \tag{3.1}$$

where λ is an arbitrary constant called the Lagrange's constant. Now

$$\frac{\partial F}{\partial q_i} = -\frac{u_i p_i}{q_i} + \lambda, \tag{3.2}$$

for $i = 1, 2, \dots, N$; and

$$\frac{\partial F}{\partial \lambda} = \sum_{i=1}^{N} q_i - 1. \tag{3.3}$$

Equation (3.2) when equated to zero gives

$$q_i = \frac{u_i p_i}{\lambda}, \quad i = 1, 2, \dots, N. \tag{3.4}$$

Equating $\partial F / \partial \lambda = 0$, we get

$$\sum_{i=1}^{N} q_i = 1. \tag{3.5}$$

From (3.4) and (3.5), we get $\lambda = \sum_{i=1}^{N} u_i p_i$. Thus extreme value for $I(P; Q; U)$ exists at

$$q_i = u_i p_i \Big/ \sum_{i=1}^{N} u_i p_i \,, \quad i = 1, 2, \ldots, N \,. \tag{3.6}$$

Next, we verify whether (3.6) is a point of maxima or minima for $I(P; Q; U)$. The border matrix for the system under consideration is

$$\begin{bmatrix} 0 & q_1 & q_2 & \ldots & q_N \\ q_1 & \frac{u_1 p_1}{q_1^2} & 0 & \ldots & 0 \\ q_2 & 0 & \frac{u_2 p_2}{q_2^2} & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots\vdots\vdots & \vdots \\ q_N & 0 & 0 & \ldots & \frac{u_N p_N}{q_N^2} \end{bmatrix} \tag{3.7}$$

It can be very easily verified that the minors of order 3, 4, 5 ... etc. of the determinant of the matrix (3.7) are negative. Thus minima occurs at the point (3.6). □

The minimum of

$$I(P; Q; U) = -\sum_{i=1}^{N} u_i p_i \log \left(u_i p_i \Big/ \sum_{i=1}^{N} u_i p_i\right) =$$
$$= -\sum_{i=1}^{N} u_i p_i \log p_i - \overline{u \log u} + \bar{u} \log \bar{u} \,,$$

where the bar means the mean value with respect to the probability distribution $P = (p_1, p_2, \ldots, p_N)$.

Since $u \log u$ is a convex U function, therefore, $\overline{u \log u} \geqq \bar{u} \log \bar{u}$, and thus, minimum of $I(P; Q; U) \leqq I(P; U)$, the weighted entropy of the experiment $E = (E_1, E_2, \ldots, E_N)$.

4. QUANTITATIVE-QUALITATIVE MEASURE OF INACCURACY AND CODING THEORY

Consider an information source with output symbols $E = (E_1, E_2, \ldots, E_N)$, and let $Q = (q_1, q_2, \ldots, q_N)$ and $P = (p_1, p_2, \ldots, p_N)$ be respectively the asserted and the realized probability distributions for the source alphabet. Let here each letter E_i be characterized by an additional parameter u_i and thus the cost c_i of transmitting E_i through the noiseless channel is proportional to the product $u_i n_i$, where n_i is the length of the codeword associated with E_i. The experimenter constructs code (in fact, personal probability code) keeping in view to minimize the average transmission cost, or equivalently the weighted mean length, (refer to [2]),

$$L(u;q) = \frac{\sum_{i=1}^{N} q_i u_i n_i}{\sum_{j=1}^{N} q_j u_j}, \tag{4.1}$$

while the actual weighted mean length is

$$L(u;p) = \frac{\sum_{i=1}^{N} p_i u_i n_i}{\sum_{j=1}^{N} p_j u_j}. \tag{4.2}$$

Rewriting (4.1) and (4.2) as

$$L(u;q) = \sum_{i=1}^{N} q_i' n_i \tag{4.3}$$

and

$$L(u;p) = \sum_{i=1}^{N} p_i' n_i \tag{4.4}$$

respectively, where

$$q_i' = \frac{q_i u_i}{\sum_{j=1}^{N} q_j u_j}, \tag{4.5}$$

and

$$p_i' = \frac{p_i u_i}{\sum_{j=1}^{N} p_j u_j}. \tag{4.6}$$

The distributions (4.5) and (4.6) represent respectively the auxiliary predicted and actual probability distributions over the source alphabet $E = (E_1, E_2, \ldots, E_N)$.

We have the following theorem:

Theorem 3. If the codeword lengths $n_1, n_2, \ldots, n_N$ satisfy the Kraft's inequality $\sum_{i=1}^{N} D^{-n_i} \leqq 1$, then the weighted mean length is bounded by

$$\frac{I(P;Q;U) - \overline{(u \log u)}_p + \bar{u}_p \log \bar{u}_q}{\bar{u}_p \log D} \leqq L(u;p) < \tag{4.7}$$

$$< \frac{I(P;Q;U) - \overline{(u \log u)}_p + \bar{u}_p \log \bar{u}_q}{\bar{u}_p \log D} + 1,$$

where

$$L(u;p) = \frac{\sum_{i=1}^{N} p_i u_i n_i}{\sum_{j=1}^{N} p_j u_j}, \quad I(P;Q;U) = -\sum_{i=1}^{N} u_i p_i \log q_i,$$

and

$$\overline{(u \log u)}_p = \sum_{i=1}^{N} u_i p_i \log u_i\,, \quad \bar{u}_p = \sum_{i=1}^{N} u_i p_i\,, \quad \bar{u}_q = \sum_{i=1}^{N} u_i q_i\,.$$

Proof. Equation (4.7) immediately follows from the Kerridge's inequality [3],

$$-\sum_{i=1}^{N} p_i' \log q_i' \leqq \sum_{i=1}^{N} p_i' n_i < -\sum_{i=1}^{N} p_i' \log q_i' + 1\,,$$

where $\{p_i'\}_{i=1}^N$ and $\{q_i'\}_{i=1}^N$, as defined by (4.6) and (4.5) respectively, are the auxiliary actual and predicted probability distributions over the source alphabet $E = (E_1, E_2, \ldots, E_N)$. □

Particular cases:

(1) When $P = Q$, (4.7) reduces to

$$\frac{I(P; U) - \overline{u \log u} + \bar{u} \log \bar{u}}{\bar{u} \log D} \leqq L(u; p) < \frac{I(P; U) - \overline{u \log u} + \bar{u} \log \bar{u}}{\bar{u} \log D} + 1\,,$$

where $I(P; U) = -\sum_{i=1}^{N} u_i p_i \log p_i$ is the weighted entropy [1], and the bar means the mean value with respect to the probability distribution $P = (p_1, p_2, \ldots, p_N)$.

These were the bounds obtained by Longo [6].

(2) When the utilities are ignored, that is, $u_i = 1$ for each i, then (4.7) reduces to

$$-\sum_{i=1}^{N} p_i \log q_i \leqq \sum_{i=1}^{N} p_i n_i < -\sum_{i=1}^{N} p_i \log q_i + 1\,;$$

a result obtained by Kerridge [3].

(Received June 4, 1985.)

REFERENCES

[1] M. Belis and S. Guiasu: A quantitative-qualitative measure of information in cybernetic system. IEEE Trans. Inform. Theory *14* (1968), 593–594.

[2] S. Guiasu and C. F. Picard: Borne inferieure de la longueur utile de certains codes. C. R. Acad. Sci. Paris *273* (1971), 248–251.

[3] D. F. Kerridge: Inaccuracy and inference. J. Roy. Statist. Soc. Ser. B *23* (1961), 184–194.

[4] S. Kullback: Information Theory and Statistics. Wiley, New York 1959.

[5] G. Longo: Quantitative-Qualitative Measures of Information. Springer-Verlag, New York 1972.

[6] G. Longo: A noiseless coding theorem for sources having utilities. SIAM J. Appl. Math. *30* (1971), 739–748.

[7] C. Shannon and W. Weaver: The Mathematical Theory of Communication. Univ. of Illinois Press, 1949.

[8] H. C. Taneja: On measures of relative 'useful' information. Kybernetika *21* (1985), 148–156.
[9] H. C. Taneja and R. K. Tuteja: Characterization of a quantitative-qualitative measure of relative information. Inform. Sciences *33* (1984), 217–222.

Dr. H. C. Taneja, Department of Mathematics, Maharshi Dayanand University, Rohtak — 124 001. India.
Dr. R. K. Tuteja, Department of Mathematics, Maharshi Dayanand University, Rohtak — 124 001. India.