

# On the Amount of Information Resulting from Empirical and Theoretical Knowledge

Igor VAJDA, Arnošt VESELÝ,

and Jana ZVÁROVÁ

EuroMISE Center

Institute of Computer Science

Academy of Sciences of the Czech Republic

CZ-182 07, Prague — Czech Republic

vajda@utia.cas.cz    vesely@euromise.cz

zvarova@euromise.cz

Recibido: 26 de agosto de 2004

Aceptado: 11 de noviembre de 2004

## ABSTRACT

We present a mathematical model allowing formally define the concepts of empirical and theoretical knowledge. The model consists of a finite set  $\mathcal{P}$  of predicates and a probability space  $(\Omega, \mathcal{S}, P)$  over a finite set  $\Omega$  called ontology which consists of objects  $\omega$  for which the predicates  $\pi \in \mathcal{P}$  are either valid ( $\pi(\omega) = 1$ ) or not valid ( $\pi(\omega) = 0$ ). Since this is a first step in this area, our approach is as simple as possible, but still nontrivial, as it is demonstrated by examples. More realistic approach would be more complicated, based on a fuzzy logic where the predicates  $\pi \in \mathcal{P}$  are valid on the objects  $\omega \in \Omega$  to some degree ( $0 \leq \pi(\omega) \leq 1$ ). We use the classical information divergence to introduce the amount of information in empirical and theoretical knowledge. By an example is demonstrated that information in theoretical knowledge is an extension of the “sematic information” introduced formerly by Bar Hillel and Carnap as an alternative to the information of Shannon.

*Key words:* probability space, ontology, predicate, knowledge area, state of the knowledge area, empirical knowledge, theoretical knowledge, information in empirical knowledge, information in theoretical knowledge.

*2000 Mathematics Subject Classification:* G2B10, 94A17, 94D05.

---

The authors were partially supported by the grant LN00B107 of the Ministry of Education, Czech Republic.

## Introduction

Already Whitehead and Russell [9] distinguished the knowledge based on empirical evidence called “knowledge by acquaintance”, and the alternative knowledge called “knowledge by description”. The empirical evidence usually means statistical data or statistical estimates obtained from such data. The alternative to the empirical knowledge is a theoretical knowledge which is based on procedures of logical induction and deduction.

To illustrate the difference between the empirical and theoretical knowledge we can refer to Hobza and Vajda [5]. These authors studied a “law” formulated in Benford [3], namely that the relative frequencies  $\hat{p}_i$  of the first nonzero digits  $1 \leq i \leq 9$  in a statistical ensemble of decimal measurements taken on real objects and varying in a large interval are surprisingly often close to

$$p_i = \log_{10} \frac{i+1}{i}.$$

Here  $(p_1 \doteq 0.301, p_2 \doteq 0.176, \dots, p_9 = 0.046)$  is a Benford distribution from which we see that the leading significant digit 1 is almost twice as frequent as the digit 2 in the samples satisfying the Benford law, and almost seven times more frequent than the digit 9. Hobza and Vajda mentioned on one hand the empirical distributions  $(\hat{p}_1, \hat{p}_2, \dots, \hat{p}_9)$  presented in Benford [3] and related to as diverse data sources the areas of rivers, specific heats and molecular weights of various chemical compounds, American League sport statistics, numbers appearing on front pages of newspapers, etc. All these distributions close to the theoretical distribution  $(p_1, p_2, \dots, p_9)$  of Benford can serve as examples of empirical knowledge. The Benford theoretical distribution which is obtained as a limit from the Theorem of Hobza and Vajda [5], or this Theorem itself, can serve as examples of theoretical knowledge.

These examples indicate that the formalization of empirical knowledge is relatively simple. Formalization of the theoretical knowledge procedures can be based on the mathematical model of scientific theories developed by Carnap [4] and others, see, e.g., Chapter 4 in Watanabe [12]. We present a version of this model in the frame of which we can simply define the empirical and theoretical knowledge.

In this paper we are interested in the amount of information resulting from empirical and theoretical knowledge. The classical information of mathematical statistics—the Fisher information—characterizes informativity of the statistical models with continuous parameters and it is thus not applicable here. Similarly, the classical information of the Shannon information theory is the information in data about signals statistically correlated with these data. Thus this concept is too narrow to be used for the present purposes. But we use the fact that both the Fisher and Shannon information can be defined by means of the information divergence

$$D(p \parallel q) = \sum_{i=1}^m p_i \log \frac{p_i}{q_i} \quad (1)$$

of probability distributions  $p = (p_1, \dots, p_m)$  and  $q = (q_1, \dots, q_m)$  where  $p \log p/q = 0$  if  $p = 0$ ,  $q \geq 0$  and  $p \log p/q = \infty$  if  $p > 0$  and  $q = 0$ . As it is well known (see, e.g., Menezes et al [7]) this divergence and other similar divergences play a principal role in the statistical data processing. We use the divergence (1) to define the information in empirical and theoretical knowledge.

Our concept of information differs from the extensions of the Shannon information in the literature on generalized uncertainty and fuzzy logic, see, e.g., Morales et al [8] and Klir and Wierman [6]. It is motivated by the need to formalize and estimate informations in empirical data and theoretical statements in some disciplines, e.g., in the medical informatics (see Bemmell [1] and Zvárová [13]).

## 1. Theoretical and empirical knowledge

Every knowledge area (empirical science) is usually based on a certain set  $\mathcal{P} = \{\pi_1, \pi_2, \dots, \pi_r\}$  of predicates describing various properties of the objects of interest, or various relations involving these objects. Moreover, there is given an ontology (reality)  $(\Omega, \mathcal{S}, P)$  where  $\Omega$  is a set of the above considered objects,  $\mathcal{S}$  is a  $\sigma$ -algebra of events  $E \subset \Omega$ , and  $P : \mathcal{S} \rightarrow [0, 1]$  is a probability measure on  $\mathcal{S}$  such that  $P(E) > 0$  if  $E \neq \emptyset$ . The ontology is assumed to verify the predicates  $\pi \in \mathcal{P}$  in the sense of specification the values

$$\pi(\omega) = \begin{cases} 1 & \text{(object } \omega \in \Omega \text{ has the property } \pi), \\ 0 & \text{(object } \omega \in \Omega \text{ has not the property } \pi). \end{cases}$$

By means of the logical operators  $\pi \rightarrow \tilde{\pi}$  interpreted as implication (i.e.,  $\tilde{\pi}$  is valid if  $\pi$  is valid) and  $\neg\pi$  interpreted as negation (i.e.,  $\neg\pi$  is valid if  $\pi$  is not valid) are constructed formulas  $\alpha, \beta$  (formal statements). Examples are, e.g.,

$$\alpha = \pi_1 \rightarrow \pi_2, \quad \beta = \pi_3 \rightarrow \alpha, \quad \gamma = \alpha \rightarrow \beta.$$

Very important examples are the “logical zero”

$$\emptyset = \alpha \rightarrow \neg\alpha$$

which is never valid (i.e.,  $\emptyset(\omega) = 0$  for all  $\omega \in \Omega$ ) and the “logical unit”

$$\square = \alpha \rightarrow \alpha$$

which is always valid (i.e.,  $\square(\omega) = 1$  for all  $\omega \in \Omega$ ). The set  $\mathcal{F}$  of all formulas is infinite and represents possible statements in the knowledge area

$$\langle\langle \Omega, \mathcal{S}, P \rangle, \mathcal{P} \rangle.$$

However, some statements  $\alpha, \beta$  in  $\mathcal{F}$  are equivalent in the sense  $\alpha \equiv \beta$ . This equivalence is defined as follows:

$$\alpha \equiv \beta = (\alpha \rightarrow \beta) \cap (\beta \rightarrow \alpha)$$

where

$$\alpha \cap \beta = \neg(\alpha \rightarrow \neg\beta) \quad (\text{conjunction of } \alpha \text{ and } \beta).$$

Similarly we define

$$\alpha \cup \beta = \neg\alpha \rightarrow \beta \quad (\text{disjunction of } \alpha \text{ and } \beta).$$

The equivalence  $\equiv$  defines a decomposition of  $\mathcal{F}$  into a set  $\mathcal{B}$  of mutually equivalent statements. It is known (see, e.g., pp. 321–324 in Watanabe [12]) that  $\mathcal{B}$  is a finite Boolean algebra with respect to the operators  $\neg, \cup, \cap$  where  $\emptyset$  and  $\square$  are the zero and unit of the algebra and if  $\beta, \beta_1$  and  $\beta_2$  are the equivalence classes from  $\mathcal{B}$  then  $\neg\beta, \beta_1 \cup \beta_2$  and  $\beta_1 \cap \beta_2$  are the equivalence classes of the statements  $\neg\alpha, \alpha_1 \cup \alpha_2$  and  $\alpha_1 \cap \alpha_2$  for  $\alpha \in \beta, \alpha_1 \in \beta_1, \alpha_2 \in \beta_2$ . (These classes are independent of the choice of  $\alpha, \alpha_1$  and  $\alpha_2$ .) Similarly we define  $\beta_1 \equiv \beta_2, \beta_1 \rightarrow \beta_2$  for  $\beta_1, \beta_2 \in \mathcal{B}$ . Moreover, it is known that there exists a set of atoms,

$$\mathcal{A} = \{\alpha_1, \alpha_2, \dots, \alpha_{2^r}\} \subset \mathcal{B} \tag{2}$$

(spectrum of the algebra  $\mathcal{B}$ ) such that

$$\alpha_j \cap \alpha_k = \emptyset \quad \text{for } j \neq k \tag{3}$$

and for every  $\beta \in \mathcal{B}$  there exists unique subset  $\mathcal{A}_\beta \subset \mathcal{A}$  with the property

$$\beta = \bigcup_{\alpha \in \mathcal{A}_\beta} \alpha \tag{4}$$

We suppose that all functions  $\alpha_j : \Omega \rightarrow \{0, 1\}$  are  $\mathcal{S}$ -measurable, i.e., that the sets  $\alpha_j^{-1}(1) = \{\omega \in \Omega : \alpha_j(\omega) = 1\}$  belong to  $\mathcal{S}$ . We also remind that  $r$  denotes the number of predicates in  $\mathcal{P}$ . By (2),  $p = 2^r$  is the number of atoms. By the property (4), the number of statements in  $\mathcal{B}$  is  $2^p = 2^{2^r}$ .

*Example 1.1.* The set  $\mathcal{P} = \{\pi_1, \pi_2\}$  defines the Boolean algebra of figure 1 with the atoms of figure 2. Since  $r = 2$ , the number of atoms is  $2^2 = 4$  and the number of statements in the Boolean algebra is  $2^4 = 16$ .

**Definition 1.2.** The state of the knowledge area  $\langle(\Omega, \mathcal{S}, P), \mathcal{P}\rangle$  is a pair  $\langle\mathcal{X}, q\rangle$  where  $\mathcal{X} = \{x_1, x_2, \dots, x_k\}$  with  $1 < k \leq 2^r$  is a subset of the spectrum  $\mathcal{A}$  such that

$$P(\alpha^{-1}(1)) = 0 \quad \text{for every } \alpha \in \mathcal{A} - \mathcal{X} \tag{5}$$

and  $q = (q_1, q_2, \dots, q_k)$  is an estimate of the probability distribution

$$p = (p_i = P(x_i^{-1}(1)) : 1 \leq i \leq k). \tag{6}$$

|             |                            |                        |              |              |                            |                        |                        |
|-------------|----------------------------|------------------------|--------------|--------------|----------------------------|------------------------|------------------------|
| $\beta_1$   | $\beta_2$                  | $\beta_3$              | $\beta_4$    | $\beta_5$    | $\beta_6$                  | $\beta_7$              | $\beta_8$              |
| $\emptyset$ | $\pi_1 \cap \pi_2$         | $\pi_1 \cap \neg\pi_2$ | $\pi_1$      | $\pi_2$      | $\pi_1 \nabla \pi_2$       | $\pi_1 \cup \pi_2$     | $\pi_1 \cup \neg\pi_2$ |
| $\beta_9$   | $\beta_{10}$               | $\beta_{11}$           | $\beta_{12}$ | $\beta_{13}$ | $\beta_{14}$               | $\beta_{15}$           | $\beta_{16}$           |
| $\square$   | $\neg\pi_1 \cap \neg\pi_2$ | $\neg\pi_1 \cup \pi_2$ | $\neg\pi_1$  | $\neg\pi_2$  | $\neg(\pi_1 \nabla \pi_2)$ | $\neg\pi_1 \cap \pi_2$ | $\neg\pi_1 \cap \pi_2$ |

Figure 1: Statements of  $\mathcal{B}$  where  $A \nabla B = (A \cup B) \cap \neg(A \cap B)$  stands for “exclusive or”, usually denoted as “xor”.

|                    |                        |                        |                            |
|--------------------|------------------------|------------------------|----------------------------|
| $\alpha_1$         | $\alpha_2$             | $\alpha_3$             | $\alpha_4$                 |
| $\pi_1 \cap \pi_2$ | $\pi_1 \cap \neg\pi_2$ | $\neg\pi_1 \cap \pi_2$ | $\neg\pi_1 \cap \neg\pi_2$ |

Figure 2: Atoms of the algebra  $\mathcal{B}$  from figure 1.

Notice that the sets  $\alpha^{-1}(1)$  and  $x_i^{-1}(1)$  considered in (5) and (6) belong by assumption to the  $\sigma$ -algebra  $\mathcal{S}$ . Further, by (4),

$$\square = \bigcup_{\alpha \in \mathcal{A}} \alpha$$

so that

$$\bigcup_{\alpha \in \mathcal{A}} \alpha^{-1}(1) = \square^{-1}(1) = \Omega$$

where the sets  $\alpha^{-1}(1)$  and  $\tilde{\alpha}^{-1}(1)$  are by (3) disjoint for  $\alpha \neq \tilde{\alpha}$ . Consequently

$$P\left(\bigcup_{\alpha \in \mathcal{A}} \alpha^{-1}(1)\right) = \sum_{\alpha \in \mathcal{A}} P(\alpha^{-1}(1)) = P(\Omega) = 1.$$

Therefore, by (5),

$$0 = \sum_{\alpha \in \mathcal{A} - \mathcal{X}} P(\alpha^{-1}(1)) = 1 - \sum_{x \in \mathcal{X}} P(x^{-1}(1))$$

so that

$$\sum_{i=1}^k p_i = \sum_{x \in \mathcal{X}} P(x^{-1}(1)) = 1,$$

i.e., (6) is a probability distribution, as asserted in Definition 1.2.

**Definition 1.3.** Empirical knowledge in the knowledge area  $\langle(\Omega, \mathcal{S}, P), \mathcal{P}\rangle$  which is in a state  $\langle\mathcal{X}, q\rangle$  is an estimate  $\hat{q} = (\hat{q}_1, \hat{q}_2, \dots, \hat{q}_k)$  of the distribution (6). This knowledge transforms the knowledge area into a new state  $\langle\mathcal{X}, \hat{q}\rangle$ .

We call the elements  $\beta \in \mathcal{B}$  statements but, as said above,  $\mathcal{B}$  consists of classes of mutually equivalent statements. Therefore two different elements  $\beta_1, \beta_2 \in \mathcal{B}$  are never equivalent (i.e.,  $\beta_1 \equiv \beta_2$  is always false). In what follows we use the relation

$$\beta \equiv \square \pmod{\Omega}$$

for  $\beta \in \mathcal{B}$  different from  $\square \in \mathcal{B}$ . By this relation we mean that  $\beta$  is equivalent to  $\square$  conditionally in the ontology  $(\Omega, \mathcal{S}, P)$  in the sense that

$$\beta(\omega) = 1 \quad \text{for all } \omega \in \Omega.$$

According to (3) and (4), this means that if  $\alpha$  is an atom from  $\mathcal{A} - \mathcal{A}_\beta$  then

$$\alpha^{-1}(1) = \emptyset, \quad \text{i.e., } \alpha(\omega) = 0 \quad \text{for all } \omega \in \Omega.$$

In other words, the atoms from  $\bar{\mathcal{A}}_\beta = \mathcal{A} - \mathcal{A}_\beta$  can be ignored as they represent statements which are always false in the ontology  $(\Omega, \mathcal{S}, P)$ .

**Definition 1.4.** Theoretical knowledge in the knowledge area  $\langle(\Omega, \mathcal{S}, P), \mathcal{P}\rangle$  which is in a state  $\langle\mathcal{X}, q\rangle$  is the statement  $\beta \equiv \square \pmod{\Omega}$  for some  $\beta$  from the Boolean algebra  $\mathcal{B}$ . This knowledge transforms the knowledge area into a new state  $\langle\tilde{\mathcal{X}}, \tilde{q}\rangle$ . Here  $\tilde{\mathcal{X}} = \mathcal{X} - \bar{\mathcal{A}}_\beta$  where  $\bar{\mathcal{A}}_\beta = \mathcal{A} - \mathcal{A}_\beta$  is the uniquely defined set of atoms with the property

$$\neg\beta = \bigcup_{\alpha \in \bar{\mathcal{A}}_\beta} \alpha \quad (\text{cf. (4)}) \tag{7}$$

and  $\tilde{q}$  is the corresponding restriction of the distribution  $q$ . If  $\mathcal{X} = \{x_1, x_2, \dots, x_k\}$ ,  $q = (q_1, q_2, \dots, q_k)$ , and  $\tilde{\mathcal{X}} = \{x_1, x_2, \dots, x_m\}$  for  $1 < m \leq k$  then the restriction is  $\tilde{q} = (\tilde{q}_1, \tilde{q}_2, \dots, \tilde{q}_m)$  defined by

$$\tilde{q}_i = \frac{q_i}{q_1 + q_2 + \dots + q_m}, \quad 1 \leq i \leq m.$$

*Example 1.5.* Consider in Example 1.1 the theoretical knowledge  $\beta \equiv \square \pmod{\Omega}$  for  $\beta = \pi_1 \rightarrow \pi_2$ . For example let  $\Omega$  be a collection of persons, let the predicate  $\pi_1$  means that a person is ill and  $\pi_2$  that a person has increased the body temperature. Since  $\pi_1 \rightarrow \pi_2$  is equivalent to  $\neg\pi_1 \cup \pi_2$  which is in turn equivalent to  $\neg(\pi_1 \cap \neg\pi_2)$ , we see that

$$\neg\beta = \pi_1 \cap \neg\pi_2 = \alpha_2 \quad (\text{cf. figure 2}).$$

Hence we obtain from (7) that  $\bar{\mathcal{A}}_\beta = \{\alpha_2\}$ . This means that the theoretical knowledge “ $(\pi_1 \rightarrow \pi_2) \equiv \square \pmod{\Omega}$ ” transforms the initial state  $\langle\mathcal{X} = \{\alpha_1, \alpha_2, \alpha_3, \alpha_4\}, q = (q_1, q_2, q_3, q_4)\rangle$  into the new state

$$\left\langle \tilde{\mathcal{X}} = \{\alpha_1, \alpha_3, \alpha_4\}, \tilde{q} = \left( \frac{q_1}{q_1 + q_3 + q_4}, \frac{q_3}{q_1 + q_3 + q_4}, \frac{q_4}{q_1 + q_3 + q_4} \right) \right\rangle.$$

Since  $p_4 = P(\alpha_4^{-1}(1))$  is the probability of no disease and no increased temperature, in the normal population  $p_4$  will be much larger than  $p_1 + p_3$  while  $p_2 = P(\alpha_2^{-1}(1))$  is zero. Therefore the estimate  $q_4$  may also be expected much larger than  $q_1 + q_2$ , i.e., the first two components of  $\tilde{q}$  will be close to zero while the last component will be close to 1.

## 2. Information in empirical and theoretical knowledge

The amount of information resulting from empirical and theoretical knowledge will be evaluated by comparing distribution  $q$  of the actual state  $\langle \mathcal{X}, q \rangle$  of the knowledge area with the true distribution  $p$  defined by (6).

**Definition 2.1.** Let  $\hat{q} = (\hat{q}_1, \hat{q}_2, \dots, \hat{q}_k)$  be the empirical knowledge in the state  $\langle \mathcal{X} = (x_1, x_2, \dots, x_k), q = (q_1, \dots, q_k) \rangle$ , and let  $\hat{q}_i > 0$  if  $p_i > 0$  for the true distribution  $p = (p_1, \dots, p_k)$  defined by (6). Then the information  $\mathcal{I}(\hat{q})$  resulting from this knowledge is the decrement of information divergence

$$\mathcal{I}(\hat{q}) = D(p \parallel q) - D(p \parallel \hat{q}) \quad (\text{cf. (1)}). \quad (8)$$

We see that the information  $\mathcal{I}(\hat{q})$  is infinite if  $D(p \parallel q)$  is infinite which takes place if  $q_i = 0$  when  $p_i > 0$ . If the information is finite then it may be positive, negative or zero, depending in whether the improved estimate  $\hat{q}$  is closer to the true distribution  $p$  than the initial estimate  $q$ . If  $\hat{q}$  is errorless, i.e., if  $\hat{q} = p$  then  $D(p \parallel \hat{q}) = D(p \parallel p) = 0$  so that it follows from (8) that

$$\mathcal{I}(\hat{q}) = D(p \parallel q) \geq 0$$

where  $D(p \parallel q) = 0$  only if  $q = p$ , i.e., only if the initial estimate was errorless. For unbiased estimates  $\hat{q}$ , i.e., for the estimates whose expectations satisfy the condition  $E\hat{q} = p$ , we obtain from the concavity of the logarithmic function the upper bound

$$\begin{aligned} E\mathcal{I}(\hat{q}) &= D(p \parallel q) - ED(p \parallel \hat{q}) \\ &\leq D(p \parallel q) - D(p \parallel E\hat{q}) \\ &= D(p \parallel q) - D(p \parallel p), \end{aligned}$$

i.e., expected information  $E\mathcal{I}(\hat{q})$  at most equals the information divergence  $D(p \parallel q)$ .

We formulate the following property as a theorem.

**Theorem 2.2.** *The information  $\mathcal{I}(\hat{q})$  of Definition 2.1 is given by the formula*

$$\mathcal{I}(\hat{q}) = \sum_{i=1}^k p_i \log \frac{\hat{q}_i}{q_i}. \quad (9)$$

*Proof.* By (9) and (1),

$$\mathcal{I}(\hat{q}) = \sum_{i=1}^k p_i \log \frac{p_i}{q_i} - \sum_{i=1}^k p_i \log \frac{p_i}{\hat{q}_i} \tag{10}$$

where the difference is well defined under the assumption of Definition 2.1. By the convention behind the sum in (1), both sums in (10) extend over all  $i$  such that  $p_i > 0$ . Therefore we may assume without loss of generality that  $p_i > 0$  for all  $1 \leq i \leq k$ . If  $q_i > 0$  for all  $1 \leq i \leq k$  then (9) is a trivial consequence of (10). If  $q_i = 0$  for some  $1 \leq i \leq k$  then (10) implies  $\mathcal{I}(\hat{q}) = \infty$  and it is easy to see that then (9) implies  $\mathcal{I}(\hat{q}) = \infty$  too.  $\square$

**Definition 2.3.** Let  $\beta \equiv \square \pmod{\Omega}$  be the theoretical knowledge in a state  $\langle \mathcal{X} = (x_1, \dots, x_k), q = (q_1, \dots, q_k) \rangle$  which transforms this state into a new state  $\langle \tilde{\mathcal{X}}, \tilde{q} \rangle$  with  $\tilde{\mathcal{X}} \subset \mathcal{X}$  specified in Definition 1.4. Further, let  $q^* = (q_1^*, \dots, q_k^*)$  be the extension of  $\tilde{q}$  defined by  $q_i^* = 0$  for  $x_i \notin \tilde{\mathcal{X}}$ . Then the information  $\mathcal{I}(\beta)$  resulting from this knowledge is the decrement of information divergence

$$\mathcal{I}(\beta) = D(p \parallel q) - D(p \parallel q^*).$$

Now we shall formulate an important consequence of Definition 2.3.

**Theorem 2.4.** Let  $\beta \equiv \square \pmod{\Omega}$  be a theoretical knowledge in a state  $\langle \mathcal{X} = (x_1, \dots, x_k), q = (q_1, \dots, q_k) \rangle$  transformed by this knowledge into a new state

$$\langle \tilde{\mathcal{X}} = (x_1, \dots, x_k), \tilde{q} = (\tilde{q}_1, \dots, \tilde{q}_{\tilde{k}}) \rangle, \quad q \leq \tilde{k} < k. \tag{11}$$

Then

$$I(\beta) = -\log \sum_{i=1}^{\tilde{k}} q_i.$$

*Proof.* By the definition of theoretical knowledge in Definition 1.4 and by (11),

$$p_i = 0 \quad \text{for } \tilde{k} < i \leq k$$

and

$$\tilde{q}_i = \frac{q_i}{c} \quad \text{for } c = \sum_{i=1}^{\tilde{k}} q_i.$$

Therefore (11) together with Definition 2.3 imply

$$I(\beta) = \sum_{i=1}^{\tilde{k}} p_i \log \frac{p_i}{q_i} - \sum_{i=1}^{\tilde{k}} p_i \log \frac{p_i}{q_i/c} = -\log c$$

which completes the proof.  $\square$



This theorem has many interesting applications. One of them is mentioned in the following example.

*Example 2.5.* If under the assumptions of Theorem 2.4 the initial distribution is uniform, i.e., if

$$q = (q_i = 1/|\mathcal{X}| : 1 \leq i \leq k),$$

then

$$I(\beta) = -\log \sum_{i=1}^{\tilde{k}} 1/|\mathcal{X}| = \log \frac{|\mathcal{X}|}{|\tilde{\mathcal{X}}|}.$$

In other words, the information is the reduction

$$I(\beta) = \log|\mathcal{X}| - \log|\tilde{\mathcal{X}}|$$

of logarithms of the number of atoms due to the knowledge. This is the formula for the so-called semantic information introduced by Bar-Hillel and Carnap [2] (see also Tondl [11]).

## References

- [1] J. H. van Bommel, *The structure of medical informatics*, Medical Informatics **9** (1984), 175–180.
- [2] Y. Bar-Hillel and R. Carnap, *Semantic information*, British J. Philos. Sci. **4** (1953), 147–157.
- [3] F. Benford, *The law anomalous numbers*, Proc. Amer. Philosoph. Soc. **78** (1953), 551–572.
- [4] R. Carnap, *Introduction to symbolic logic and its applications*, translated by W. H. Meyer and J. Wilkinson, Dover Publications Inc., New York, 1959.
- [5] T. Hobza and I. Vajda, *On the Newcomb-Benford law in models of statistical data*, Rev. Mat. Complut. **14** (2001), no. 2, 407–420.
- [6] G. J. Klir and M. J. Wierman, *Uncertainty-based information: Elements of generalized information theory*, Studies in Fuzziness and Soft Computing, vol. 15, Physica-Verlag, Heidelberg, 1999.
- [7] M. Menéndez, D. Morales, L. Pardo, and I. Vajda, *Two approaches to grouping of data and related disparity statistics*, Comm. Statist. Theory Methods **27** (1998), no. 3, 609–633.
- [8] D. Morales, L. Pardo, and I. Vajda, *Uncertainty of discrete stochastic Systems: General theory and statistical inference*, IEEE Trans. on Systems, Man and Cybernetics **26** (1996), no. 3, 681–697.
- [9] A. N. Whitehead and B. Russell, *Principia Mathematica*, Cambridge Univ. Press, Cambridge, 1927.
- [10] C. E. Shannon, *A mathematical theory of communication*, Bell System Tech. J. **27** (1948), 379–423, 623–656.
- [11] L. Tondl, *Problems of Semantics*, Reidel, Boston, 1981.
- [12] S. Watanabe, *Knowing and guessing*, John Wiley & Sons Inc., New York, 1969.
- [13] J. Zvárová, *On the medical informatics structure*, International Journal of Medical Informatics **44** (1997), 75–81.