SOME PROPERTIES OF THE DIVERGENCE MEASURE OF INFORMATION CONTENT AS RELATED TO QUANTITATIVE ANALYSES*

Karel ECKSCHLAGER and Vladimír ŠTĚPÁNEK

Institute of Inorganic Chemistry, and
Institute of Physics,
Czechoslovak Academy of Sciences, 250 68 Řež

Received January 8th, 1981

Dedicated to Professor RNDr V. Suk on the occasion of his 60th birthday.

Some properties of the divergence measure of the information content of quantitative analysis results are introduced and it is shown how this measure describes the effect of the difference of a result obtained by the analysis from a preliminary one or from an anticipated value. In the same time the measure is compared to the decrease of uncertainty as expressed in terms of Shannon's entropy.

First applications of information theory in analytical chemistry occurred at the end of the sixties[1]. In 1974 Eckschlager and Vajda[2] abandoned the Shannon's approach adopted by that time and started from a model of higher-precision analysis as a process of obtaining information by replacing an apriori normal distribution by another aposteriori[3] normal one and found an adequate expression of the information content for this model in the divergence measure[4] of the information gain. In 1975 the divergence measure was extended to other probability distributions; later on, its use for the needs of analytical practice was compared to the use of other information measures and Eckschlager and Štěpánek[5,6] have recently treated it as a difference of the Kerridge–Bongard measure of inaccuracy[7,8] of a statement made prior to the analysis and the Shannon's entropy[9] of the aposteriori distribution. Some properties of the divergence measure have been investigated in general paper[10].

The divergence measure has been adopted for evaluating the information content of quantitative analysis results mainly by Eckschlager and Štěpánek[11]. Cleij and Dijkstra[12] have paid their attention to the evaluation of results and methods of qualitative and identification analyses and they believe that the use of information theory in quantitative analysis leads only to characteristics not differing essentially from

* Part XVI in the series Theory of Information as Applied to Analytical Chemistry; Part XV: This Journal *46*, 478 (1981).

those of precision. This opinion is correct only as far as we consider other measures of the information content than the divergence measure, *i.e.*, as we adopt the Shannon's entropy and do not take into account inaccuracy in its general sense. The use of information theory principles in qualitative and identification analyses or in structural analysis is certainly of greater practical importance than their use in quatitative analysis, yet it can always yield adequate descriptions of analytical problems and procedures suitable for their solution.

In this paper we are going to show those properties of the divergence measure of the information content which are important for the formulation of problems and for the evaluation of results or methods in quantitative analysis; the practical impact of these properties will be demonstrated in examples.

THEORETICAL

The uncertainty associated with a continuous random variable having a probability density $q(x)$ can be expressed by Shannon's measure (entropy) satisfying a set of reasonably stated assumptions[9]. It is derived as

$$H(q) = -\int_{x_1}^{x_2} q(x) \log q(x)\, \mathrm{d}x \,, \tag{1a}$$

where $\int_{x_1}^{x_2} q(x)\, \mathrm{d}x = 1$ and we put $\lim \left[-q(x) \log q(x)\right] = 0$ for $q(x) \to 0+$.

Two probability distributions having equal Shannon's entropies will be called isentropic. The inaccuracy of an assertion that a continuous random variable has a probability density $p(x)$ while the true distribution is $q(x)$ is characterized by the Kerridge–Bongard measure[7,8]

$$H(q, p) = \int_{x_1}^{x_2} q(x) \log \frac{1}{p(x)}\, \mathrm{d}x \,, \tag{1b}$$

where $p(x) > 0$ for $x \in \langle x_1, x_2 \rangle$ and $\int_{x_1}^{x_2} p(x)\, \mathrm{d}x = \int_{x_1}^{x_2} q(x)\, \mathrm{d}x = 1$.

This measure satisfies a set of assumptions similar to those used in the derivation of the entropy.

The information content of a result of quantitative analysis can be understood as decrease of uncertainty after the analysis was carried out[2,3,11]. If $p(x)$ is the probability density of the apriori distribution describing our assumption about the content of the component to be determined before the analysis and $q(x)$ is the probability density of the aposteriori distribution of the results of the analysis, the information content can be evaluated in terms of the Shannon's entropy as $H(p) - H(q)$; or we can express the information content as the information gain by the means of the divergence measure

$$I(q, p) = H(q, p) - H(q) = \int_{x_1}^{x_2} q(x) \log \frac{q(x)}{p(x)} \, dx \tag{2}$$

with $p(x)$ and $q(x)$ satisfying the above stated conditions. The exact derivation of this measure can be found in chapters 4·3 and 4·4 of the monograph[11].

For arbitrary $p(x)$ and $q(x)$ we have $H(q, p) \geqq H(q)$; this follows from (*2*) and from the non-negativeness of $I(q, p)$ – see property *a*) below. If $p \equiv q$ the sign of equality is valid.

We can easily show that the divergence measure of the information content in (*2*) has following properties:

a) $I(q, p) \geqq 0$, $I(q, q) = 0$, *i.e.*, it is non-negative and equal to zero if $p(x) \equiv q(x)$.

b) $I(q_1, p) > I(q_2, p)$ if the expectations of both aposteriori distributions are equal, *i.e.*, $\mu_1(q) = \mu_2(q)$ but the standard deviations are in relation $\sigma_1(q) < \sigma_2(q)$; the pre-distribution $p(x)$ is the same in either case.

c) $I(q, p_1) > I(q, p_2)$ if the expectations of both apriori distributions are the same but $\sigma_1(p) > \sigma_2(p)$; the aposteriori distribution $q(x)$ does not vary.

d) If $\sigma_1(p) = \sigma_2(p)$, $\sigma_1(q) = \sigma_2(q)$ then $I(q_1, p_1) > I(q_2, p_2)$ provided that $|\mu_1(p) - \mu_1(q)| > |\mu_2(p) - \mu_2(q)|$. The difference between $I(q_1, p_1)$ and $I(q_2, p_2)$ can be very small if the probability density of the apriori distribution $p(x)$ is constant in the whole interval $\langle x_1, x_2 \rangle$.

These four properties with be further discussed from the practical point of view of the quantitative analysis. Similarly as in[3,4,10] we will compare these properties of the divergence measure $I(q, p) = H(q, p) - H(q)$ to those of the Shannon's measure $H(p) - H(q)$ for the decrease of uncertainty.

DISCUSSION

Property *a*) has been generally proved, *e.g.*, in[11]. Here it will be treated specifically below, along with properties *b*) through *d*), for cases of a uniform or a normal apriori distribution and for a normal aposteriori distribution.

1) If $p(x)$ is normal $N(\mu_0, \sigma_0^2)$ and $q(x)$ is normal $N(\mu, \sigma^2)$ then we get for the information content – in natural information units

$$I(q, p) = \ln \frac{\sigma_0}{\sigma} + \frac{1}{2} \left[\frac{(\mu - \mu_0)^2 + \sigma^2 - \sigma_0^2}{\sigma_0^2} \right] \tag{3}$$

as has been already shown in[2,11] where the information content of higher-precision analyses has been studied. In this case the difference of Shannon's entropies is equal to $\ln \sigma_0/\sigma$. If we lay down $\mu = \mu_0$ the formula in (*3*) reduces to

$$I(q, p) = \ln \frac{\sigma_0}{\sigma} + \frac{1}{2}\left[\left(\frac{\sigma}{\sigma_0}\right)^2 - 1\right]. \tag{3a}$$

Obviously the information content is zero for $\sigma = \sigma_0$, otherwise it is positive. That becomes apparent, besides others, from the following mathematical manipulation: If we put $\sigma_0/\sigma = A$ then the function

$$I = \ln A + \frac{1}{2}\left(\frac{1}{A^2} - 1\right)$$

assumes its minimum $I = 0$ at $A = 1$ and is positive otherwise. The dependences of the value of the function in (*3a*) on the ratio σ_0/σ and on the logarithm of this ratio are illustrated in Figs 1*a* and 1*b* respectively. On the contrary the difference of entropies takes on negative values whenever $\sigma > \sigma_0$. If $\mu \neq \mu_0$ the information content $I(q, p)$ is always positive, also for $\sigma = \sigma_0$.

II) If $p(x)$ is uniform

$$p(x) = \begin{cases} \dfrac{1}{x_2 - x_1} & \text{for} \quad x \in \langle x_1, x_2 \rangle \\ 0 & \text{otherwise} \end{cases}$$

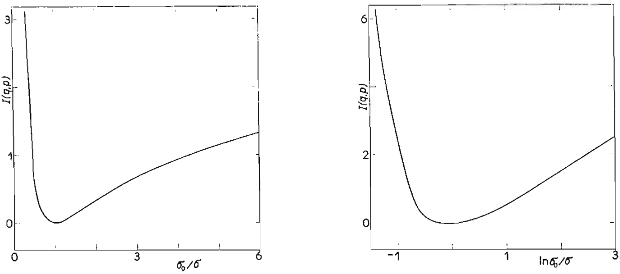

FIG. 1

The dependence of $I(q, p)$ for apriori and aposteriori normal distributions *a*) on σ_0/σ; *b*) on $\ln(\sigma_0/\sigma)$

and $q(x)$ is again normal $N(\mu, \sigma^2)$ and μ lies outside $\langle x_1, x_2 \rangle$ or is close to either limit, *i.e.*, $\mu < x_1 + 3\sigma$ or $\mu > x_2 - 3\sigma$, then

$$I(q, p) = \ln \frac{x_2 - x_1}{\sigma \sqrt{2\pi e}} + \frac{1}{2}\left[\frac{z_2\, \varphi(z_2) - z_1\, \varphi(z_1)}{\phi(z_2) - \phi(z_1)} - 2 \ln \left[\phi(z_2) - \phi(z_1)\right]\right] \quad (4)$$

TABLE I

Values of $\ln (x_2 - x_1)/(\sigma \sqrt{(2\pi e)})$, values of the term

$$A = \frac{1}{2}\left[\frac{z_2\, \varphi(z_2) - z_1\, (\varphi z_1)}{\phi(z_2) - \phi(z_1)} - 2 \ln \left[\phi(z_2) - \phi(z_1)\right]\right]$$

and values of $I(q, p)$ according to (4) for some ratios $(x_2 - x_1)/\sigma$ and some positions of μ with regard to x_1

$(x_2 - x_1)/\sigma$	μ	$\ln (x_2 - x_1)/\sigma \sqrt{(2 \pi e)}$	A	$I(q, p)$
6·0	$x_1 + 3\sigma$	$3{\cdot}7282 \,.\, 10^{-1}$	$1{\cdot}6036 \,.\, 10^{-2}$	$3{\cdot}8885 \,.\, 10^{-1}$
8·0	$x_1 + 4\sigma$	$6{\cdot}6050 \,.\, 10^{-1}$	$6{\cdot}0003 \,.\, 10^{-4}$	$6{\cdot}6110 \,.\, 10^{-1}$
10·0	$x_1 + 5\sigma$	$8{\cdot}8365 \,.\, 10^{-1}$	0·0000	$8{\cdot}8365 \,.\, 10^{-1}$
20·0	$x_1 + 10\sigma$	1·5768	0·0000	1·5768
20·0	$x_1 + 6\sigma$	1·5768	0·0000	1·5768
20·0	$x_1 + 4\sigma$	1·5768	$3{\cdot}0038 \,.\, 10^{-4}$	1·5771
20·0	$x_1 + 3\sigma$	1·5768	$8{\cdot}0253 \,.\, 10^{-3}$	1·5848
20·0	$x_1 + 2\sigma$	1·5768	$7{\cdot}8407 \,.\, 10^{-2}$	1·6552
20·0	$x_1 + \sigma$	1·5768	$3{\cdot}1693 \,.\, 10^{-1}$	1·8937
20·0	x_1	1·5768	$6{\cdot}9510 \,.\, 10^{-1}$	2·2719

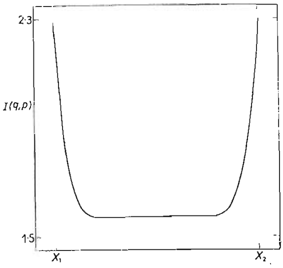

FIG. 2

The dependence of $I(q, p)$ on the position of μ at a given length of interval $\langle x_1, x_2 \rangle$ for an apriori uniform and an aposteriori normal distribution (*cf.* page 115 of the monography[11]) where $z_i = (x_i - \mu)/\sigma$, $i = 1, 2$ and $\varphi(z_i)$ and $\phi(z_i)$ are values of the frequency and the distribution function of the standardized normal variable respectively.

Regarding the basic requirement that the length of the interval $\langle x_1, x_2 \rangle$ be at least 6σ (in order to contract the normal variable into this interval) the first term in (*4*) is positive. Furthermore in respect to that $z_2 > 0$ and $z_1 < 0$ the difference $z_2\ \varphi(z_2) - z_1\ \varphi(z_1)$ is also positive. Finally ln $[\phi(z_2) - \phi(z_1)]$ is always negative so that the whole expression in (*4*) is positive. The dependence of the information content on the ratio $(x_2 - x_1)/\sigma$ for some positions of μ is presented in Table I and the dependence on μ for given values of x_1 and x_2 is illustrated in Figure 2. In this case the decrease of uncertainty is the same as the information content and is therefore positive because we deal with transition from a uniform distribution[11].

We can remark upon the possibility of negative values of the decrease of uncertainty that they are understandable, for instance, when completely incorrect results of a quantitative analysis misinform us in essence. However, a negative value of entropy arises most frequently by underestimating the apriori uncertainty of the probability distribution. Thus we take the non-negativeness of the information content $I(q, p)$ in its use in the evaluation of results of quantitative analyses for an advantage of this measure.

The validity of properties *b*) and *c*) is immediately obvious from formulas (*3*) and (*4*) if we diminish σ in the first case and σ_0 or the length of the interval $\langle x_1, x_2 \rangle$ in the second one – the standard deviation of the uniform distribution depends only on this length. Yet we need to mention that properties *b*) and *c*) remain maintained even then when the information content is understood as the decrease of uncertainty.

The property *d*), *i.e.*, the dependence of the value of $I(q, p)$ on the difference of the expectations of the apriori and the aposteriori distributions, is very important for the assessment of the information content of quantitative analysis results. In opposition to the decrease of uncertainty which is expressed in terms of precision the divergence information measure characterizes also the effect of the difference of the result obtained by the analysis from the preliminary assumption. What the divergence measure contains in addition to the difference $H(p) - H(q)$ is well observable from formulas (*3*) and (*4*) which are adjusted here in such a way that the first logarithmic term corresponds to the decrease of uncertainty and the second one, *i.e.* $\frac{1}{2}[\ldots]$, depending on $|\mu - \mu_0|$ or on $z_i = (x_i - \mu)/\sigma$, $i = 1, 2$, then expresses that effect of the deviation of the result from the assumption. It is best seen in the case of the formula (*4*) if we expect any result from the interval $\langle x_1, x_2 \rangle$ with the same likelihood; the term $\frac{1}{2}[\ldots]$ increases the more the closer to zero is $|z_i|$. In Table I this term is entered also in some cases when $\mu \in (x_1 + 3\sigma, x_2 - 3\sigma)$ and when we calculate the information content no longer according to (*4*). In these particular cases it is practically zero. The importance of the term $\frac{1}{2}[\ldots]$ in formulas (*3*) and (*4*) excels in the case of isentropic apriori and aposteriori distributions when $H(p) - H(q) = 0$. So, *e.g.*, for both distributions being normal with the same value σ but with different average values, *i.e.*, for two isentropic mutually shifted normal distributions, the quantity in (*3*) becomes

$$I(q, p) = \frac{1}{2}\left(\frac{\mu - \mu_0}{\sigma}\right)^2 \qquad (3b)$$

which enabled to find the information content of quantitative analysis results subject to a systematic error[11,14].

The importance of described properties in analytical practice can be shown on results of a determination of nitrogen by Kjeldahl's method. For a content of about 3% N the value of $\sigma = 0{\cdot}05$ was found. If we assume that the content of N in the analyte is 3·00 ± 0·50% then $(x_2 - x_1)/\sigma = 20$ and the result 3·00 ± 0·20% has the information content $I(q, p) = 1{\cdot}577$ natural units, the result 2·65 or 3·35% N has the information content 1·584 nits and the results on the anticipated limits, *i.e.* 2·50 or 3·50% N, have $I(q, p) = 2{\cdot}272$ nits. More expressive is the effect of the deviation of the results from a preliminary result; if we find the content of nitrogen 3·00% with the photometric method ($\sigma_0 = 0{\cdot}0625$) and the Kjeldahl's method confirms this finding the information content of this result is relatively small, only 0·043 units. Yet if we observe 3·10% N with Kjeldahl's method then $I(q, p) = 1{\cdot}325$ units and for a result 3·20% N we get even $I(q, p) = 5{\cdot}172$ units.

The effect of the difference as it has been shown in quoted examples can be interpreted as follows: for an apriori uniform distribution, in our case, *e.g.*, for $x \in \langle 2{\cdot}50; 3{\cdot}50\rangle$ when we anticipate nothing more than that we will find the result within this interval the effect of the deviation of the result is fairly small as far as it lies within these limits. However, where the apriori distribution is derived from results of a preliminary analysis there the finding of the deviation of the result obtained by a method yielding accurate results[15] represents very relevant information which will become evident also in the value of the divergently measured information content.

REFERENCES

1. Malissa H.: Fresenius' Z. Anal. Chem. *241*, 1 (1969); *256*, 7 (1971).
2. Eckschlager K., Vajda I.: This Journal *39*, 3076 (1974).
3. Eckschlager K.: Fresenius' Z. Anal. Chem. *277*, 1 (1975).
4. Eckschlager K.: This Journal *42*, 225 (1977).
5. Eckschlager K., Štěpánek V.: This Journal *45*, 2516 (1980).
6. Eckschlager K.: Chem. listy *73*, 1188 (1979).
7. Kerridge D. F.: J. Roy. Stat. Soc. B *23*, 184 (1961).
8. Bongard M. M.: Probl. Kibernet. *9*, (1963).
9. Shannon C. E.: The Bell Syst. Techn. J. *27*, 379, 623 (1948).

10. Vajda I., Eckschlager K.: Kybernetika *16*, 120 (1980).
11. Eckschlager K., Štěpánek V.: *Information Theory as Applied to Chemical Analysis.* Wiley--Interscience, New York 1979.
12. Cleij P., Dijkstra A.: Fresenius' Z. Anal. Chem. *298*, 97 (1979).
13. Danzer K., Marx G.: Chem. Anal. (Warsaw) *24*, 33, 43 (1979).
14. Eckschlager K.: This Journal *44*, 2373 (1979).
15. Eckschlager K.: This Journal *25*, 987 (1960).

Translated by V. Štěpánek.