COMPARISON OF DIFFERENT MEASURES FOR EXPRESSING THE INFORMATION PROPERTIES OF QUANTITATIVE ANALYSIS RESULTS*

K. ECKSCHLAGER

Institute of Inorganic Chemistry,
Czechoslovak Academy of Sciences, 250 68 Prague - Řež

Received April 6th, 1976

The use of the Brillouin, Shannon, and Kullback divergence measures is compared in expressing the information content of quantitative analysis results for one or more components of the analysed sample simultaneously; the mutual dependence of these measures is shown. The role of the units used in expressing the information content with the different measures is discussed. For the case of a simultaneous determination of more components, the actual and potential information content is distinguished. On the basis of the divergence measure, the information performance and rentability of an analytical method are defined.

Quantitative analysis as a process of obtaining informations about the composition of a material proceeds in two consecutive systems: in the first one (*e.g.*, an analytical apparatus) the information is formed, comes out in a coded form of analytical signals, and is fed to another system which decodes the information (Fig. 1). The coded analytical information, which comes out from the first system, is formed by a digital or analogue signal intensity, y, which has the character of a random variable, or in the case of a simultaneous determination of k components $(k > 1)$ usually by a sequence of signals characterized by their position, z_i, and intensity, y_i $(i = 1, 2, ..., k)$. It is not important whether the signals have the form of a step-like curve or peaks. The output of the second system yields the result, x, or the results for the individual components, x_i $(i = 1, 2, ..., k)$. The result of the quantitative analysis has the

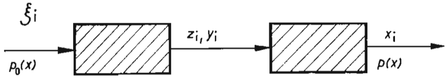

FIG. 1

Analysis as a Process for Obtaining Information

* Part X in the series Theory of Information as Applied to Analytical Chemistry; Part IX: This Journal *41*, 2527 (1976).

character of random, continuously distributed variable with a probability density $p(x)$ or with a simultaneous probability density $p(x_1, x_2, \ldots, x_k)$. Since besides the sample material with a content of the considered component ξ (or several components ξ_i to be determined simultaneously) also a preliminary information about the sample composition characterized by the probability density $p_0(x)$ or densities $p_0(x_i)$, $i = 1, 2, \ldots, k$, is fed to the first system, the information content depends not only on $p(x)$ but also on $p_0(x)$ or $p(x_i)$ and $p_0(x_i)$.

In common case of quantitative analysis, we consider a continuous rectangular distribution as one which characterizes the preliminary information about the composition of the analysed sample in the case where we do not know nothing more about it than that $\xi \in \langle x_0, x_1 \rangle$. The probability density of such a rectangular distribution is

$$p_0(x) = \begin{cases} 1/(x_1 - x_0) & \text{for} \quad x \in \langle x_0, x_1 \rangle \\ 0 & \text{otherwise}\,. \end{cases} \tag{1}$$

In other cases, the preliminary information is a result of an analysis whose repeated determinations are distributed normally; then

$$p_0(x) = \frac{1}{\sigma_0 \sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x - \mu_0}{\sigma_0}\right)^2\right]. \tag{2a}$$

The results of a quantitative analysis are usually distributed normally so that the probability density is

$$p(x) = \frac{1}{\sigma \sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x - \mu}{\sigma}\right)^2\right]. \tag{2b}$$

In a case of the repeated higher precision analysis, *i.e.*, if $p_0(x)$ is given by Eq. (*2a*) and $p(x)$ by (*2b*), the relation $\sigma_0 \geqq \sigma$ must be valid and μ can but need not be different from μ_0.

We shall compare the relations resulting from different measures of the information content for the results of quantitative and repeated higher precision quantitative analysis, and this for the cases of the determination of one or more components at the same time. We shall consider only these measures of the information content which have already been used in evaluating the analytical results, and from the properties of these measures we shall deduce a definition of the information performance and rentability of analytical methods.

THEORETICAL

In the first paper of this series[1] the information content of the result of a quantitative determination of a single component of an analysed sample was expressed with the aid of the Brillouin[2] relation

$$I(U, U_0) = \log_b (U_0/U)\,, \tag{3}$$

where the uncertainty prior to the analysis, U_0, and that remaining after the analysis, U, can be determined as the reliability interval of the distribution $p_0(x)$ and $p(x)$ for a predetermined significance level α. The base of the logarithm determines the units in which the information content is expressed[1]. Most frequent is the use of $b = 2$ ($\log_2 a = \mathrm{ld}\, a$) and the "bit" units, less usual is $b = \mathrm{e}$ ($\log_e a = \ln a$) and "nit" units. The recalculation of the quantities in bit to nit units and vice versa is easy[1]. For a rectangular $p_0(x)$ and normal $p(x)$ the information content in bit units is given by

$$I(U, U_0) = \mathrm{ld}\, \frac{(x_1 - x_0)(1 - \alpha)}{2 z_\alpha \sigma}\,, \tag{4a}$$

where z_α is the critical value of the normal distribution at the chosen significance level α. For normal $p_0(x)$ and $p(x)$ is

$$I(U, U_0) = \mathrm{ld}\,(\sigma_0/\sigma) \tag{4b}$$

assuming $\sigma \leqq \sigma_0$. Here, of course, the value of $I(U, U_0)$ does not depend on the chosen significance level since we choose the same value of α for $p_0(x)$ as for $p(x)$. It is a disadvantage of the Brillouin measure (3) that in some cases the information content, a quantity which is naturally expected to be unique, depends on the arbitrarily chosen significance level. The ratio of U_0/U in Eq. (3) represents the number of possible distinguishings; for continuous distributions $p_0(x)$ and mainly $p(x)$ it is necessary to define on the basis of a substantially arbitrary convention a certain "rastering" of the x values in order to enable the determination of the number of possible distinguishings. Danzer[3-8] proposed a consistent system for evaluating the information properties of analytical results and methods based on the fact that the information content according to the Brillouin measure is given in substance by the logarithm of the number of possible distinguishings.

Another method for determining the information content[9-12] originating from communication techniques and adapted to the needs of analytical chemistry is the so-called Shannon measure

$$I(H, H_0) = H_0 - H\,, \tag{5}$$

where H or H_0 denotes mean information[13] or information entropy, given by the Shannon equation[14]

$$H_0 = -\int_{-\infty}^{+\infty} p_0(x) \operatorname{ld} p_0(x) \, dx \,, \tag{6a}$$

$$H = -\int_{-\infty}^{+\infty} p(x) \operatorname{ld} p(x) \, dx \,. \tag{6b}$$

Hence,

$$I(H, H_0) = \int_{-\infty}^{+\infty} p(x) \operatorname{ld} p(x) \, dx - \int_{-\infty}^{+\infty} p_0(x) \operatorname{ld} p_0(x) \, dx \,. \tag{7}$$

After Wiener[15], Eqs (*6a,b*) were proposed already by von Neumann as a "rational" information measure. It can be in analogy to the Boltzmann's definition of entropy[16] denoted as a measure of the information disorder, "ignorance". The properties of the information entropy have been discussed in the literature[17-20]. For a rectangular $p_0(x)$ and normal $p(x)$ we have

$$I(H, H_0) = \operatorname{ld} \frac{(x_1 - x_0)}{(\sigma \sqrt{2\pi e})} \,. \tag{8}$$

It is obvious that if $z_\alpha/(1 - \alpha) = \frac{1}{2}\sqrt{(2\pi e)} = 2{\cdot}066$ (*i.e.*, for the significance level $\alpha = 0{\cdot}05$), then $I(U, U_0)$ after (*4a*) takes the form of (*8*). The value $\alpha = 0{\cdot}039$ considered in our preceding work[21,22], corresponding to $z_\alpha = \frac{1}{2}\sqrt{(2\pi e)}$, represents the significance level on which the binary logarithm of the reliability interval of the normal distribution according to Eq. (*2b*) corresponds to its entropy given by Eq. (*6b*). As to the determination of the information content of the results of quantitative analyses in practice, Eq. (*5*) has the disadvantage that its use has a sense only if the result of the analysis confirms the assumption, since only if $x_0 \leqq \mu \leqq x_1$, any value of μ represents an equal or zero information and the information content is given by the diminution of the entropy caused by performing the analysis. The information content is then expressed in bit units. The information content of more accurate analyses, when $p_0(x)$ and $p(x)$ are normal, is given according to the Shannon measure by the same relation as according to the Brillouin measure, *i.e.*, by (*4b*). The ratio of $(x_1 - x_0)/\sigma \sqrt{(2\pi e)}$ in Eq. (*8*) as well as σ_0/σ in (*4b*) represent the number of possible distinguishings.

The Kullback[27] divergence measure used earlier[22-26] for expressing the information content of the analytical results,

$$I(p, p_0) = \int_{-\infty}^{+\infty} p(x) \ln \left[p(x)/p_0(x)\right] dx \,, \tag{9}$$

broadens considerably our possibilities as compared to the measures (*3*) and (*5*). In choosing the base of the logarithm, we need not choose $b = 2$, as usual and founded in communication techniques, and we can simplify the derivation of the equations for $p_0(x)$ and $p(x)$ according to (*2a,b*) by setting $b = e$ (*i.e.*, by using natural logarithms).

If we compare Eq. (*9*) in the form

$$I(p, p_0) = \int_{-\infty}^{+\infty} p(x) \ln p(x) \, dx - \int_{-\infty}^{+\infty} p(x) \ln p_0(x) \, dx \tag{10}$$

with (*7*), it is apparent that they differ except for the base of the logarithms (which is, of course, of little importance) only by the probability densities $p(x)$ and $p_0(x)$ in the second integral, which for a rectangular $p_0(x)$ is equal, disregarding the base of the logarithm, to the entropy H_0 according to Eq. (*6a*) (since for any probability distribution we have $\int_{-\infty}^{+\infty} p(x) \, dx = \int_{-\infty}^{+\infty} p_0(x) \, dx = 1$). Besides it was shown[26] that the divergence measure enables to determine the information content even when the result of the analysis does not confirm the assumption, *i.e.*, when μ lies beyond the interval $\langle x_0, x_1 \rangle$. If we perform the derivation by the method described earlier[26], in which we separate from the interval $(x_1 - x_0) \gg \sigma$ the quantity $\Delta x = \sigma$, which we distribute equally in the range from $\mu - 3\sigma$ to x_0 or from x_1 to $\mu + 3\sigma$ for μ lying beyond $\langle x_0, x_1 \rangle$, then we have for $\mu < x_0$

$$\begin{aligned} I(p, p_0) = \ln (1/\sigma \sqrt{2\pi e}) - \{\Phi(a_0 + 1) \,.\, \ln [1/(x_1 - x_0)(a_0 + 4)] + \\ + [\Phi(-a_1) - \Phi(a_0 + 1)] \,.\, \ln [1/(x_1 - x_0)]\} \end{aligned} \tag{11a}$$

and for $\mu > x_1$

$$\begin{aligned} I(p, p_0) = \ln (1/\sigma \sqrt{2\pi e}) + \\ + \{[1 - \Phi(-a_1 - 1)] \ln [1/(x_1 - x_0)(a_1 + 4)] + \\ + [\Phi(-a_1 - 1) - \Phi(a_0)] \ln [1/(x_1 - x_0)]\} \,. \end{aligned} \tag{11b}$$

Here we denote $a_0 = (x_0 - \mu)/\sigma$, $a_1 = (\mu - x_1)/\sigma$, $\Phi(a_0) = \int_{-\infty}^{x_0} p(x) \, dx$, $\Phi(-a_1) = \int_{-\infty}^{x_1} p(x) \, dx$, $\Phi(a_0 + 1) = \int_{-\infty}^{x_0 + \sigma} p(x) \, dx$, and $\Phi(-a_1 - 1) = \int_{-\infty}^{x_1 - \sigma} p(x) \, dx$, where the probability density $p(x)$ is given by Eq. (*2b*). (For the numerical calculation, it is advantageous that the function Φ is tabulated.)

Hence, the divergence measure (*9*) or (*10*) is a more general case of the Shannon's or Wiener's and Neumann's measure (*7*). The divergence measure takes the form of the latter in the case where *a*) $p_0(x)$ is the probability density of a rectangular distribution, *b*) the result of the analysis, *i.e.*, the expected value of the normal distribution $p(x)$ is $x_0 + 3\sigma \leqq \mu \leqq x_1 - 3\sigma$, and *c*) the width of the interval is $(x_1 - x_0) \gg \sigma$, at least $(x_1 - x_0) \geqq 6\sigma$.

For the case of repeated higher precision analysis[24], *i.e.*, when $p_0(x)$ and $p(x)$ are normal according to (*2a*) and (*2b*), we have the relation

$$I(p, p_0) = \ln \frac{\sigma_0}{\sigma} + \frac{(\mu - \mu_0)^2 + \sigma^2 - \sigma_0^2}{2\sigma_0^2}, \tag{11c}$$

which, however, for $\mu = \mu_0$ does not take the form of (*4b*) and therefore is not a more general case of the latter in the sense in which Eq. (*11a*) or (*11b*) is a more general case of (*8*). Although for $(\mu - \mu_0)^2 = \sigma_0^2 - \sigma^2$ we have $I(p, p_0) = \ln(\sigma_0/\sigma)$, hence Eq. (*11c*) is similar to (*4b*) disregarding the base of the logarithm, this case is from the point of view of a higher precision analysis not one in which we could evaluate the information content only according to the change of the entropy of the probability distribution prior and after the repeated higher precision analysis since the condition of similarity of Eqs (*11c*) and (*4b*) is that the difference of the expected values, $|\mu - \mu_0|$, is not zero. The values of $I(p, p_0)$ expressed with the use of natural logarithms according to (*11a,b*) for $x_1 - x_0 = 10$ and for several values of $a_0 = (x_0 - \mu)/\sigma$ or $a_1 = (\mu - x_1)/\sigma$ are shown in Table I in dependence on σ.

TABLE I

Values of $I(p, p_0)$ and $I(H, H_0)$ for Rectangular $p_0(x)$ and Normal $p(x)$
$x_1 - x_0 = 10$; for $\mu < x_0$ is $a_0 = a$, for $x_1 < \mu$ is $a = a_1$.

a	σ				s	
	0·5	0·1	0·05	0·01	0·005	0·001
+5	3·774	4·588	6·077	7·686	8·379	9·989
+4	3·657	4·471	5·959	7·569	8·262	9·871
+3	3·520	4·334	5·823	7·432	8·125	9·735
+2	3·328	4·142	5·630	7·240	7·933	9·542
+1	2·931	3·746	5·234	6·843	7·536	9·146
0	2·270	3·084	4·572	6·182	6·875	8·485
−1	1·751	2·565	4·054	5·663	6·356	7·966
−2	1·592	2·406	3·895	5·504	6·197	7·807
−3	1·577	2·391	3·879	5·489	6·182	7·792
$I(H, H_0)$	1·577	2·391	3·879	5·489	6·182	7·792

Here we have set $a = a_0 = (x_0 - \mu)/\sigma$ for $\mu < x_0$, and $a = a_1 = (\mu - x_1)/\sigma$ for $x_1 < \mu$. The values of $I(p, p_0)$ according to Eq. (*11c*) are shown in Table II for various values of $|\mu - \mu_0|/\sigma_0$ and σ/σ_0. In the last line of Tables I and II are given the values of $I(H, H_0)$ in natural units with the use of the Shannon measure. It is apparent from Table I that the values of $I(p, p_0)$ for a_0 or $a_1 = -3$ are in accord with $I(H, H_0)$, whereas the values of $I(p, p_0)$ in Table II are always different from $I(H, H_0)$. The values of $|\mu - \mu_0|/\sigma_0$ corresponding to the values of σ/σ_0 used in Table II for which $I(p, p_0) = I(H, H_0)$ are as follows:

σ/σ_0	1	0·950	0·900	0·500	0·100	0·050
$|\mu - \mu_0|/\sigma_0$	0	0·3125	0·4360	0·8667	0·9950	0·9987

The divergence measure enables further to determine the information content even in the case where more quantitative determinations are carried out at the same time as discussed in ref.[25]. In the cited work, however, only the really obtained, actual information content was defined and this as the sum of the information contents of the determinations of individual components. On the other hand, Danzer[4] used the Brillouin measure to distinguish the actual and potential, *i.e.*, maximum attainable (with the use of a given analytical method) information content. Since the distinguishing of the actual and potential content is purposeful, *e.g.*, in deciding to what extent is the use of a given analytical method effective in solving a given problem, we introduce here also the definition of the potential information content

TABLE II

Values of $I(p, p_0)$ and $I(H, H_0)$ for Normal $p_0(x)$ and $p(x)$

$|\mu - \mu_0|$ / σ_0	σ/σ_0					
	1·00	0·95	0·90	0·50	0·10	0·05
0·00	0·000	0·003	0·010	0·318	1·808	2·497
0·20	0·020	0·023	0·030	0·338	1·828	2·517
0·40	0·080	0·083	0·090	0·398	1·888	2·577
0·60	0·160	0·163	0·170	0·478	1·968	2·657
0·80	0·320	0·323	0·330	0·638	2·128	2·817
1·00	0·500	0·503	0·510	0·818	2·308	2·997
1·50	1·125	1·128	1·135	1·443	2·933	3·622
2·00	2·000	2·003	2·010	2·318	3·808	4·497
3·00	4·500	4·503	4·510	4·818	6·308	6·997
$I(H, H_0)$	0·000	0·051	0·105	0·693	2·303	2·996

with the use of the divergence measure:

$$I(p, p_0)_{\text{pot}} = \frac{z_{\max} - z_{\min}}{\Delta z} I(p, p_0) \,. \tag{12}$$

For a rectangular distribution $p_0(x)$ and normal $p(x)$, if $x_0 + 3\sigma \leqq \mu \leqq x_1 - 3\sigma$, we have $I(p, p_0) = \ln [(x_1 - x_0)/\bar{\sigma} \sqrt{2\pi e}]$, where $\bar{\sigma}$ is the mean standard deviation of the determinations of the individual components; for a completely unknown sample $(x_1 - x_0) = 100$. For the case $\mu < x_0$, $I(p, p_0)$ in Eq. (*12*) is given by Eq. (*11a*) and for $x_1 < \mu$ by (*11b*); we substitute $\bar{\sigma}$ for σ. In Eq. (*12*) Δz denotes the least distance of two neighbouring signals necessary for us to consider both signals as isolated in the sense of the Doerffel's definition[28], $z_{\min}$ and $z_{\max}$ the smallest and largest signal positions recorded by the instrument. The more accurate analysis, *i.e.*, the case of normal $p_0(x)$ and $p(x)$, need not be considered for practical reasons since in practice the results of a simultaneous quantitative or semiquantitative determination of more components represent rather preliminary results for a following more accurate determination.

If we define the actual information content $I(p, p_0)_{\text{act}}$ by Eq. (*5*) of ref.[25], then we can define the redundance, the meaning of which was explained earlier[1], as

$$\varrho = \frac{I(p, p_0)_{\text{pot}} - I(p, p_0)_{\text{act}}}{I(p, p_0)_{\text{pot}}} \,. \tag{13}$$

Hence, the redundance refers not only to excessive informations obtained by performing a large number of parallel determinations (as shown in ref.[1]), but also to informations which we could obtain by making use of all possibilities of the analytical method and which we in reality do not obtain. The redundance according to Eq. (*13*) can be even larger than that due to performing several parallel determinations[1,4].

As to the units in which the information content determined on the basis of the divergence measure is expressed, they are generally not identical with the bit or nit units and were defined for special cases[23,26] as multiples of the information content $I(p, p_0) = 1$ characterized always by a set of mutually corresponding values of at least two quantities; this set is valid for a unit isoinform. Only if the divergence measure becomes identical with the Shannon measure the units thus defined are identical with the nit units for $b = e$. As to the base of the logarithms determining the units, there is no reason why to use binary logarithms in expressing the information content as usual and founded in communication techniques. In Eq. (*10*) for the normal distribution the use of natural logarithms is more advantageous. The recalculation of the quantities in bit to nit units and *vice versa* is easy[1,23].

With the use of the Brillouin or Shannon measure, also certain other quantities were defined, characterizing practically important properties of analytical methods.

We shall give their expressions on the basis of the divergence measure. For example, the quantity of informations obtained in a time unit is usually expressed[8] as the so-called information flow, which is with the use of the divergence measure given by

$$J = \mathrm{d}I(p, p_0)/\mathrm{d}t\,. \tag{14}$$

To evaluate analytical methods, the so-called information performance[8,29] is more advantageous:

$$L = \frac{1}{t_A}\int_0^{t_A} J\,\mathrm{d}t\,, \tag{15}$$

where t_A denotes time necessary to carry out the analysis. The value of L can be expressed in natural information units. s^{-1}. An analoguous quantity is the information rentability[8,29,30], which is with the use of the divergence measure defined as

$$R(p, p_0) = \frac{1}{\tau}\,I(p, p_0)\,. \tag{16}$$

This quantity gives the amount of information obtained by the analysis with the expenses of τ and is expressed in natural information units per currency unit. The practical significance of the quantities (*15*) and (*16*) is obvious and was already discussed[29]. Since at least a partial financial equivalent of time can be frequently found (salaries, depreciation, economical losses due to waiting for the analytical result, *etc.*), the rentability seems to be of more use for the analytical practice than the information performance.

The author is indebted to Dr. M. Šolc of the same Institute for stimulating discussions concerning various measures of the information content.

REFERENCES

1. Eckschlager K.: This Journal *36*, 3016 (1971).
2. Brillouin L.: *Science and Information Theory*. Academic Press, New York 1963.
3. Danzer K.: Z. Chem. *13*, 20 (1973).
4. Danzer K.: Z. Chem. *13*, 69 (1973).
5. Danzer K.: Z. Chem. *13*, 229 (1973).
6. Danzer K.: Z. Chem. *14*, 73 (1974).
7. Danzer K.: Z. Chem. *15*, 158 (1973).
8. Danzer K.: Z. Chem. *15*, 326 (1975).
9. Kaiser H.: Anal. Chem. *42*, No 2, 24A; No 4, 26A (1970).
10. Doerffel K., Hildebrand W.: Wiss. Z. T. H. Chemie "Carl Schorlemmer" Leuna-Merseburg *11*, 30 (1969).
11. Dupuis F., Dijskstra A.: Anal. Chem. *47*, 379 (1975).

12. Grieping B., Krijgsman W.: Z. Anal. Chem. *265*, 241 (1973).
13. Malissa H., Rendl J.: Z. Anal. Chem. *272*, 1 (1974).
14. Shannon C. E.: The Bell System Techn. J. *27*, 379, 623 (1948).
15. Wiener N.: *Kybernetika*, p. 56. Published by SNTL, Prague 1960.
16. Boltzmann L.: *Vorlesungen über die Gastheorie*. J. A. Barth, Leipzig 1896.
17. Dinur U., Levine R. D.: Chem. Phys. *9*, 17 (1975).
18. Peters J.: *Einführung in die Allgemeine Informationstheorie*. Springer, Berlin—Heidelberg—New York 1967.
19. Rényi A.: *Teorie pravděpodobnosti*. Academia, Prague 1972.
20. Fraser D. A. S.: Ann. Math. Stat. *36*, 890 (1965).
21. Eckschlager K.: This Journal *38*, 1330 (1973).
22. Eckschlager K.: Chem. Listy *69*, 810 (1975).
23. Eckschlager K.: Z. Anal. Chem. *277*, 1 (1975).
24. Eckschlager K., Vajda I.: This Journal *39*, 3076 (1974).
25. Eckschlager K.: This Journal *40*, 3627 (1975).
26. Eckschlager K.: This Journal *41*, 1875 (1976).
27. Kullback S.: *Information Theory and Statistics*. Wiley, New York 1959.
28. Doerffel K.: Chem. Analit. *17*, 615 (1972).
29. Eckschlager K.: Anal. Chem., in press.
30. Eckschlager K.: This Journal *37*, 1486 (1972).

Translated by K. Micka.