CHEMICAL ANALYSIS: INFORMATION CONTRIBUTION OF RESULTS

Karel ECKSCHLAGER

*Institute of Inorganic Chemistry, Czechoslovak Academy of Sciences, 250 68 Řež**

Received April 27, 1990
Accepted August 31, 1990

This paper I dedicate to the memory of my teacher Václav Hovorka,
Corresponding Member of the Czechoslovak Academy of Sciences,
professor of analytical chemistry at the Czech Technical University in Prague.

In this review, analysis is treated as a process of gaining information on chemical composition, taking place in a stochastic system. A model of this system is outlined, and a survey of measures and methods of information theory is presented to an extent as useful for qualitative or identification, quantitative and trace analysis and multicomponent analysis. It is differentiated between information content of an analytical signal and information gain, or amount of information, obtained by the analysis, and their interrelation is demonstrated. Some notions of analytical chemistry are quantified from the information theory and system theory point of view; it is also demonstrated that the use of fuzzy set theory can be suitable. The review sums up the principal results of the series of 25 papers which have been published in this journal since 1971.

INTRODUCTION

The volume of required information on chemical composition, its time changes and the spatial distribution of particular components, as provided by chemical

---

* Present address: Department of Inorganic Chemistry, Charles University, 128 40 Prague 2.

analysis, is growing steadily. Simultaneously, demands on the rapidity and economy of the process of acquiring this information, on a better reliability and selectivity of analytical procedures increase as well; moreover, science and technology are faced with the problem of quantitating analytes in lower and lower amounts. These demands are met, partially at least, by new analytical methods being worked out based on the most diverse principles, developing the corresponding instrumentation and computer facilities, and by advancing in theoretical fundamentals. In this manner, analytical chemistry has developed into a methodologically differentiated exact branch of science, making use of knowledge gained in other branches as well[1].

A significant feature of exact sciences and their conceptualization is the use of notions that are mathematically defined on the one side and empirically grounded on the other side; such notions become indispensable tools in the basic theoretical treatment. Since the 1970's, notions, quantities and definitions taken from information and system theory[2] have been employed in the analytical process description. Such notions have a marked integrating effect upon the entire methologically diversified analytical chemistry owing to its unified approach based on information quantities[3], and, of course, also owing to its universal validity independent of the principle underlying the particular method; the fact that their interrelation can be demonstrated is of importance as well[4].

Information theory in conjunction with the system approach to the analytical process has shown that among factors determining the information gain from an analytical result is how suitably and from which partial operations the analytical system has been set up, how scrupulously the partial operations have been performed and how these proceeded (sampling, sample handling, separation, measurement, calibration, data handling, etc.). Since the chemical and physical nature of the partial processes and the errors involved are rather well understood now[5], an analytical system is never looked upon as a black box. Thus, the description of the analytical process from the information theory and system points of view can form a basis for a further development of universal rules of good laboratory practice[6] and for the theory and practice of quality data assurance[7,8]. Along with the physico-chemical laws and rules that also apply to analytical methods, the information-theory and system approach facilitates seeking for the optimal strategy of addressing particular analytical problems[9].

In the first stage of applying information theory in analytical chemistry, analogies were sought between the analytical system model and the information transmission system model; information properties of the analytical results and methods were characterized by quantities (measures) initially introduced for the purposes of communication theory. These measures can be well applied in qualitative and identification analysis; results of quantitative analysis (multicomponent analysis in particular), however, have some properties that are not regarded in communication theory. A model has been therefore created for a system in which the process of acquisition of information about chemical composition takes place, and the concept of the divergence measure[10] – which was later extended[11] – has been used in assessing this system.

In this paper, which concludes the series *Theory of Information as Applied in Analytical Chemistry* comprising 25 papers published in this journal[12–36] since 1971, an outline is given of a model of the analytical system and of the process of information acquisition concerning chemical composition; a brief overview of measures and methods of information theory and the basic concepts of system theory is also presented to an extent as useful for analytical practice. For a more detailed explanation of information-theory foundations of analytical chemistry, the interested reader is referred to monographs[10,11], papers[37–39] and, in particular, to the review[40]. A survey of quantifications of analytical notions and concepts from the information theory point of view is given in paper[4].

1. BASIC CONCEPTS

Any experiment, also chemical analysis, gives a result which can be

1) a nominal quantity, such as the kind (species), identity, etc. It can be denoted by a name, symbol, formula, etc., but it cannot be arranged in an order or assigned a numerical value;

2) an ordinal quantity, which can be fitted into a seuqence or classed in a level; the levels, however, cannot be assigned numerical values;

3) a cardinal quantity, which can be expressed as a number with a suitable unit; it can be a fixed or a random quantity.

For instance, the result of qualitative or identification analysis is a nominal quantity, the result of quantitative analysis is a cardinal quantity expressed as an amount, concentration or content. The value of any quantitative – even fixed – parameter obtained by measurement (chemical analysis) is always a random quantity because of the occurrence of random errors of measurement.

The aim of any experiment – be its result a nominal, ordinal or cardinal quantity – is to reduce the uncertainty of knowledge of the given object or process, and thus, to bring some information.

1.1. *Information on Chemical Composition*

Chemical analysis proceeds in a system whose input is a material sample and whose output is information. Sample is carrier of "latent" information on the chemical composition as a representative of the material analyzed; actual information is only extracted from it by a suitable reaction with a reagent or interaction with energy.

We distinguish between the

(i) analytical result, i.e. information on the chemical composition of sample analyzed, and the

(ii) analytical information, i.e. information on the composition, its time changes and spatial distribution of the chemical composition within a certain part of material reality.

Analytical information is obtained by chemometric processing and interpretation of a series of results of analyses of different samples, taken from the examined region in a suitable manner[40,41]. It is clear that the information gain of the analysis is considerably higher than that of the unprocessed series of the individual results.

In gaining information on chemical composition, two major stages of the entire analytical process are involved. In the first, information is created and encoded in the signal; with a simplification, this stage can be said to be the object of analytical chemistry[42]. In the second stage, the signal is decoded into the result; this is the object of chemometry[41]. The whole analytical system consists of subsystems in which the operations of the analytical process take place (sampling, decomposition, separation, measurement, calibration, data handling, etc.). All of the subsystems, their suitable sequence, relations and feedbacks are necessary for the whole analytical system to function; on the other hand, errors from operations occurring in them[5,11] contribute to the overall uncertainty of the analytical result[35].

Analytical information is not directly measurable; it is obtained by determining the level of some analytical property that is closely related with the kind and amount of analyte present in the sample. Information on the analytical property is usually encoded into the signal within the analytical process. Analytical signal carrying information on chemical composition is always tied to a change in the chemical or physical state (colour reactions, emission or absorption of radiation or particles, consumption of titrant solution, etc.). The signal can be converted into analytical information if it fulfils a syntactic, semantic and pragmatic function, as described in refs[37,40]. Signal which involves information on chemical composition has a position z, which carries information on qualitative composition, i.e. identity of i-th analyte A_i $(i = 1, 2, \ldots, n)$, and can take continuous or discrete values z_j $(j = 1,2, \ldots \ldots, m)$. Furthermore, the signal has an intensity, y, which contains information on the actual amount or content of i-th analyte, X_i. (In this treatment, the true analyte content will be denoted X and its estimate, i.e. result of quantitative analysis or the a priori assumption, will be denoted x.) The signal position is either a fixed quantity (wavelength of a spectral line), or a random quantity (elution parameters in chromatography) whose expected value can be dependent on the analytical procedure. The result of signal position measurement, though, is always a continuous random quantity associated with a probability density $p(z)$, and the signal intensity is always a continuous random quantity associated with a probability density $p(y)$. Often – in multicomponent analysis in particular – the signal intensity in a certain position,

$y(z_j)$, is considered. The properties of the signal as the analytical information carrier have been dealt with[10,11,37,40]; the signal is of fundamental importance in Belyaev's considerations[42] of the nature of chemical analysis; in Czechoslovakia this concept has been elaborated by Musil[43].

The analytical system must be regarded as stochastic; this means that on repetition with the same input, a whole probability distribution $p(x)$ with the expected value μ_i and variance σ_i^2 is obtained at the output. The expected value is said to be correct if it does not differ statistically significantly from the true analyte content X_i. In identification analysis, also, the established identities of analytes A_i must be regarded with different probabilities $P(A_i)$; for $P(A_i) = 1$ we say that the identification is unambiguous. Thus, analytical results invariably involve some uncertainty.

Of importance with respect to the reliability of the result is the way in which the signal position and intensity are converted into analytical information (see Chapter 5). In identification analysis, assignment of a signal in position z_j to the identity of analyte A_i can formally be expressed by the operator relation given in monograph[11], p. 10. In practice, this assignment is done by means of tables, collections of spectra, data banks, etc. In quantitative analysis, the relation $x = f(y)$ is given by the stoichiometric equivalent $b = \text{const.}$, i.e. as $x = by$ for chemical methods, or by a calibration function for instrumental methods. The calibration function has to be determined empirically as

$$y(z_j) = f_C(X_a\{X_i\}\{T_j\})\ ; \quad i \neq a\ , \tag{1.1}$$

where X_a is the true analyte concentration, $\{X_i\}$ are concentrations of the other components in the sample, and $\{T_j\}$ are variables (amounts of reagents added, temperature, instrumental parameters, etc.). This approach[43] allows the sensitivity of the analytical method to be defined as

$$S = 1/b \tag{1.2}$$

for chemical analysis, or as

$$S = \mathrm{d}y(z_j)/\mathrm{d}X_a \tag{1.3}$$

or instrumental analysis; the matrix effect can be characterized as the dependence of $y(z_j)$ on $\{X_i\}$ and the ruggedness of the method as the dependence of z or y, or of $y(z_j)$ on $\{T_j\}$.

1.2. *Factual and Quantitative Aspects of Analytical Information*

Any information, and thus also analytical information, has a factual side (its object) and a quantitative side (its detailedness). The factual side and the verity of information cannot be evaluated quantitatively; it is possible, however, to take into account the relevance of the experimentally obtained information for solving a given problem[11,39], to determine how precise and accurate the result must be to really contribute information[7,11,37], and, to a degree, to assess the plausibility of the result obtained, i.e. its agreement with the a priori assumption, i.e. assumption that is, for instance, derived or deduced from a hypothesis[30,40].

Information theory enables us to assess the quantitative aspect of analytical information only, i.e. the contents of the signal, the information gain from the analysis performed, volume of information obtained by multicomponent analysis, or information gain emerging on processing the data set to analytical information, etc. All this is based on Wiener's definition according to which information is reduction in uncertainty, in lack of knowledge concerning some object or process, its causes and consequences. Thus, the content of a signal or a separate measurement or the information gain of the analytical result is expressed as the difference

$$I = H_0 - H\,, \qquad (1.4)$$

where H_0 is the a priori uncertainty, existing before the experiment, and H is the a posteriori uncertainty, remaining after the experiment. The a priori uncertainty is given by the pre-information necessary for the choice of a suitable analytical method, and, of course, also by the question (in analytical practice usually: what, how much, where?) the answer to which is sought by performing the analysis. The various kinds of analysis, the signal property measured, and the question addressed by the analysis are given in Table I. The a posteriori uncertainty, caused by the experiment itself, is given by how perfectly the analysis proceeded. The a posteriori uncertainty of qualitative analysis is affected by how many components can be expected based on test results; the uncertainty of quantitative analysis is affected by the

TABLE I

Kinds of analysis as answers to a priori questions

Question	Signal	Kind of analysis
What	z	identification analysis
What, how much	y	single-component quantitative analysis
	$y(z)$	multicomponent quantitative analysis
What, how much, where	$y(z)$, l_i[a]	local analysis; structure analysis for l_i of molecular dimensions
What, how much, when	$y(z)$, t[b]	process analysis

[a] Spatial coordinates; [b] time.

precision and accuracy of the result; selectivity is of importance in quantitative, qualitative and identification multicomponent analysis; in trace analysis, the important factors are the signal-to-noise ratio, the relation of the actual analyte content to the detection limit (higher/lower), etc. Of importance is also whether a low a posteriori uncertainty is attained by employing a single method or a combination of several methods[10].

The various measures of information content of the signal or of information gain of the result of analysis only differ in the way the a priori and a posteriori uncertainties are expressed. Measures such as Brillouin's, Shannon's and divergence measures and their properties have been compared[11,21,35,40]. In this treatment we will demonstrate those ways of expressing information uncertainty, content and gain, or amount of information, playing a role in analytical chemistry.

The uncertainty of a nominal, ordinal or cardinal quantity which takes various possibilities, levels or discrete values $i = 1, 2, \ldots, n$ with probabilities P_i, can be expressed in terms of Shannon's entropy

$$H(P) = -\sum_{i=1}^{n} P_i \log_b P_i\ ; \quad \sum_{i=1}^{n} P_i = 1\,. \tag{1.5}$$

As early as 1929, Szilárd defined information as negative entropy, "negentropy". The units in which uncertainty is expressed are given by the base b of the logarithm in Eq. (*1.5*). For binary logarithms, lb ($b = 1$), the units are "bits"; for natural logarithms ln ($b = e$), the units are "nits". The use of binary logarithms has been introduced by Hartley so that for an alternative (two-value) phenomenon, where $P_1 + P_2 = 1$, $H(P)$ is unity at $P_1 = P_2 = 1/2$. If $n > 2$, then $0 \leqq H(P) \leqq \log_b n$; therefore, the so-called relative entropy

$$h(P) = H(P)/\log_b n = -(1/\log_b n)\sum_{i=1}^{n} P_i \log_b P_i \tag{1.6}$$

is occasionally used in this case. Since $\log_b a/\log_b n = \log_n a$, the relative entropy is

$$h(P) = -\sum_{i=1}^{n} P_i \log_n P_i\,. \tag{1.7}$$

Then the relation $0 < h(P) < 1$ is always satisfied, and the use of the binary logarithm can be considered as a particular case of $\log_n$ with $n = 2$. Then, it would be necessary to introduce suitable units, of which "bit" and "nit" would be particular cases.

However, we do not consider it proper to terminologically distinguish between thermodynamic entropy and information "intropy" (*in*formation en*tropy*); on the contrary, one should always bear in mind the fact that the quantity is a universally valid measure of uncertainty, disorder, irrespective of the system whose disorder is characterized by it[44].

Entropy characterizing the uncertainty of a continuous random cardinal quantity with a probability density $p(x) > 0$ for $x \in \langle x_1, x_2 \rangle$ is

$$H(p) = - \int_{x_1}^{x_2} p(x) \log p(x) \, \mathrm{d}x \; ; \quad \int_{x_1}^{x_2} p(x) \, \mathrm{d}x = 1 \, . \tag{1.8}$$

This entropy plays the role of a measure of uncertainty of accurate results of direct measurements. Entropies for continuous distributions are given in Table II. The transition from the entropy for discrete distributions to the entropy for continuous distributions has been demonstrated by Peters[44]; theoretically, $H(p)$ can take values of $-\infty \leqq H(p) \leqq +\infty$; in analytical practice, however, these extremes never occur.

The results of quantitative analysis can carry a mean error $\delta = X_i - \mu_i \neq 0$; also, the a priori uncertainty cannot be assumed to be based on an absolutely correct assumption or an absolutely correct pre-information. Then, it is more adequate to express uncertainty in terms of Kerridge's and Bongard's accuracy measure

$$H(r, p) = H(r) + D(r, p) = - \int_{x_1}^{x_2} r(x) \log_b p(x) \, \mathrm{d}x \, , \tag{1.9}$$

where the "error term" $D(r, p)$ expresses the dissimilarity, divergence of the accurate distribution $r(x)$ and the experimental distribution $p(x)$; see also refs[10,11,45,46] and Table III.

The information content of the signal, of the result of observation and of accurate results of direct measurements can be expressed as the difference of Shannon's entropies, i.e. for the discrete case of the a priori (P_0) and a posteriori probabilities or conditional probabilities[11,29] as

$$I = H(P_0) - H(P) \, , \tag{1.10}$$

and for the case of the continuous a priori $p_0(x)$ and the a posteriori $p(x)$ distributions, as

$$I = H(p_0) - H(p) \, . \tag{1.11}$$

For the same symmetric a priori and a posteriori destributions, e.g. for the a priori normal distribution $N(\mu_0, \sigma_0^2)$ and the a posteriori distribution also normal, $N(\mu, \sigma^2)$, $\sigma^2 \leqq \sigma_0^2$, we have $I = \log(\sigma_0/\sigma)$; for a priori uniform distribution $U(x_1, x_2)$ and a posteriori normal distribution we have $I = \ln(x_2 - x_1)/\sigma\sqrt{2\pi e}$.

Kateman[47] expresses information content by means of reduction of variance $R = (\sigma/\sigma_0)^2$, $R \in (0, 1\rangle$, as

$$I = -(1/2) \log R \, . \tag{1.12}$$

Details concerning the concept of the information content can be found in refs[10,11,21,29,35,36,40,47,48].

When determining the information gain of results of analysis, expressing a priori uncertainty by means of the Kerridge–Bongard measure, we use the divergence measure[10,11,48],

$$I(p, p_0) = H(p, p_0) - H(p) = \int_{x_1}^{x_2} p(x) \log_b [p(x)/p_0(x)] \, dx \,. \qquad (1.13)$$

TABLE II

Entropies of some continuous distributions

Distribution	Probability density[a]	Entropy[b]
Uniform $U(x_1, x_2)$	$\begin{cases} 1/(x_1 - x_2) & (x \in \langle x_1, x_2 \rangle) \\ 0 & (x \notin \langle x_1, x_2 \rangle) \end{cases}$	$\ln (x_2 - x_1) = \ln \sigma \sqrt{12}$
Normal $N(\mu, \sigma^2)$	$[1/\sigma \sqrt{2\pi})] \exp [-(1/2) (c - \mu)^2/\sigma^2]$	$\ln \sigma \sqrt{2\pi e}$
Truncated normal $TN(\mu, \sigma^2; x_0)$	$\begin{cases} \{1/[1 - F(x_0)] \, \sigma \sqrt{2\pi}\} \exp \,. \\ . \, [-(1/2) (x - \mu)^2/\sigma^2] \\ 0 \quad (x \leqq x_0) \end{cases}$	$\ln [1 - F(x_0)] \, \sigma \sqrt{2\pi e} + $ $+ \frac{1}{2} \frac{(x_0 - \mu) f(x_0)}{1 - F(x_0)}$
Lognormal $LN(\ln \mu, \sigma^2; x_0)$	$\begin{cases} \frac{1}{(x - x_0) \, \sigma \sqrt{2\pi}} \,. \\ . \exp \left[-\frac{1}{2} \left(\frac{\ln (x - x_0) - \ln \mu}{\sigma} \right)^2 \right] \\ 0 \quad (x \leqq x_0; \, x_0 \geqq 0) \end{cases}$	$\ln \mu \sigma \sqrt{2\pi e}$

[a] $F(x)$ is the distribution function of normal distribution; [b] $f(x)$ is the frequency function of normal distribution.

TABLE III

Application of various information uncertainty and information content or gain measures

Quantity	Uncertainty	Eq.	Content or gain	Eq.
Nominal Ordinal Cardinal discrete	$H(P) = -\sum P_i \log P_i$	(1.5)	$I = H(P_0) - H(P)$	(1.10)
Cardinal continuous	$H(p) =$ $= -\int p(x) \log p(x) \, dx$	(1.8)	$I(p, p_0) =$ $= H(p, p_0) - H(p)$ unbiased results	(1.13)
	$H(r, p) =$ $= -\int r(x) \log p(x) \, dx$	(1.9)	$I(r, p, p_0) =$ $= H(r, p_0) - H(r, p)$ biased results	(1.14)

This measure assumes that the results of analysis are accurate; therefore the a posteriori uncertainty is expressed by means of Shannon's netropy, and $r(x) = p(x)$ is inserted in the Kerridge–Bongard measure, which expresses the a priori uncertainty. In case the results of quantitative analysis may not be accurate, both the a priori and a posteriori uncertainties are expressed by means of the Kerridge–Bongard measure, so that the information gain is expressed as

$$I(r, p, p_0) = H(r, p_0) - H(r, p) = \int_{x_1}^{x_2} r(x) \log_b [p(x)/p_0(x)] \, dx \, . \qquad (1.14)$$

This measure is the most general expression of the information gain of results of quantitative analysis; we refer to it as the extended divergence measure[11,28,32,35,40]; see Table III. Of importance is the fact that we need not know the actual distribution $r(x)$; the knowledge of its expected value and of its variance[32] (or an estimate of its variance[7]) is sufficient.

The amount of information obtained from the results of multicomponent analysis is usually expressed as

$$M = \sum_{i=1}^{n} I_i \, , \qquad (1.15)$$

where I_i is the information gain of the detection, identification or quantitative determination of components $i = 1, 2, \ldots, n$. This relation, however, is exactly valid only if the results are not mutually correlated.

2. QUALITATIVE AND IDENTIFICATION ANALYSIS

Qualitative analysis answers the question whether a substance is present in the sample, whereas the task of identification analysis is to answer the question which substance is present (see Table I). The system in which chemical or instrumental analysis of this kind occurs has a rather simple first stage: the signal (e.g. change in the colour of the solution or formation of precipitate) either appears or not, or it lies or does not lie in position z_j which corresponds to the identity of the analyte (spectral line or band wavelength, R_F value, etc.). The uncertainty of the result of such an experiment can be expressed in terms of the simplest measure, viz. Shannon's entropy[3,10,11,28,29,49–52], where the choice of binary logarithms is suitable for alternative decision. The use of the relative entropy is warranted for expressing uncertainty with n possibilities (Eq. (1.7)). Conversion of the result of experiment into information, however, is easy in the simplest cases only; in other cases, such as an identification in toxicological or ecological analysis, it may be impracticable without computer techniques. Therefore, assessment of the information gain of results of qualitative and, in particular, identification analysis must take into account the way in which the signal is decoded into information[28].

In the simplest case, where the presence of n components is assumed with equal probabilities $P(A_i) = 1/n$, $i = 1, 2, \ldots, n$, we have $H(P) = -n(1/n) \log(1/n) = + \log n$ (Eq. (*1.5*)). The information content of the result of qualitative analysis then is

$$I = \log n_0 - \log n = \log(n_0/n) , \qquad (2.1)$$

where n_0 is the number of components sought, the presence of which is assumed beforehand with equal probabilities, and n is the number of possible, but yet unidentified analytes after performing the experiment.

The use of relation (*2.1*) for expressing the information content is not limited to qualitative analysis; the relation can be looked upon as a reduction of the number of possible variants n_0 to $n < n_0$. Or it can be regarded as a reduction in a variety, attained, e.g., by classification, by forming clusters, etc. Later, in Paragraph 3.1., we shall express information content in terms of reduction of variance or of the confidence interval width. The reduction of the number of variants frequently concerns nominal and ordinal quantities; reduction of variance or of the width of the interval within which lie values that a quantity can take, concerns rather continuous quantities.

In the case of an unambiguous proof of the presence of a single component we have $n = 1$ and $I = \log n_0$. For expressing the a priori and a posteriori uncertainties in terms of relative entropy, logarithms with the same basis $n_0 \geqq n$ must be used; then, calculating the a posteriori entropy for $P_i = 0$, we put $-0 \log 0 = 0$. For example, if only one out of $N = 6$ cations (Ag^+, Pb^{2+}, Zn^{2+}, Al^{3+}, Ca^{2+} and Na^+) can be present in sample, the presence of all of them being considered identically probable, then in case that addition of dilute HCl gives a white precipitate the information content of this assay is $I = \text{lb}(6/2) = 1.58$ bit, whereas if the precipitate does not form, the information content is $I = \text{lb}(6/4) = 0.58$. Hence, the information content of the result of qualitative detection depends on whether the expected component is detected or not[9,10]. Combinatorial calculus must be sometimes used for various possible combinations of components, particularly for determining n_0 (ref.[10]). For instance, assuming that out of the $N = 6$ cations, $m = 1, 2, \ldots, 6$ can be present, the total number of all the possible combinations will be

$$n_0 = \sum\{N!/[m!(N - m)!]\} = 63 ,$$

neither Ag^+ nor Pb^{2+} being present in 15 of them. The reaction with dilute HCl giving no white precipitate then has an information content of $I = \text{lb}(63/15) = 2.07$ bit; if the reaction is positive, i.e. white precipitate appears indicating that either Ag^+ or Pb^{2+} or both are present, then $I = \text{lb}[63/(63 - 15)] = 0.39$ bit. The numerous simple but illustrative examples given, e.g., in refs[3,9,10,52–54] help to make clear some general viewpoints, document the importance of a suitable order of the assays in qualitative analysis, etc.

The case where, e.g. in identification analysis, all of the components are not expected with the same probability leads to the use of entropy expressed by means of conditional entropies. This approach enables solving the case of instrumental identification analysis where the signals can overlap or where their positions can only be determined with a limited accuracy; it is also possible to solve a case where the uncertainty of the results is aggravated by the fact that the analytes to be detected are present in concentrations close to the detection limit[29]. Denote $P(z_j \mid A_i)$ the probability that signal appears in position z_j, $j = 1, 2, \ldots, m$, if analyte A_i is present, and denote $P(z_j \mid X_i)$ the probability that the signal appears if analyte A_i is present in concentration X_i. The two probabilities – of which $P(z_j \mid A_i)$ is actually a particular case of $P(z_j \mid X_i)$ for X_i much higher that the detection limit – can be established experimentally, e.g. as the frequency of appearance of signal in position z_j in dependence on X_i, which is the basis of frequentomertric analysis[3,9,51]. The probability that component A_i is present if signal appears in position z_j, that is, $P(A_i \mid z_j) = 1$ for case where this signal is not disturbed by a signal corresponding to another component. Then the detection is unambiguous, the a posteriori uncertainty $H(P(A \mid z)) = 0$, and the information content of the detection is $I = H_0$, which is the maximum. If the condition of "undisturbed" signal is not fulfiled, the conditional probability must be determined by using Bayes' rule. This rule, which is sometimes referred to as the theorem of probability of causes, is given by the relation

$$P(A_1 \mid z_j) = \frac{P(A_1)\, P(z_j \mid A_1)}{\sum_{i=1}^{n} P(A_i)\, P(z_j \mid A_i)}\,; \quad j = \text{const.} \tag{2.2}$$

for $i = 1, 2, \ldots, n$. In analytical experiment conditions, we have $P(A_i) = 1$ if the analyte is added, and $P(A_i) = 0$ if the blank experiment is performed. To a first approximation, $P(A_i \mid z_j) = 1/k$ for interferents A_i, $i = 2, 3, \ldots, k$ and for $P(z_j \mid A_i) = 1$. The uncertainty of a single detection can be expressed in terms of the so-called component entropy for conditional probabilities,

$$H(P(A \mid z_j)) = -\sum_{i=1}^{n} P(A_i \mid z_j)\, \mathrm{lb}\, P(A_i \mid z_j)\,; \quad j = \text{const.} \tag{2.3}$$

The use of binary logarithms $\log_2 a = \mathrm{lb}\, a$ is commonplace for expressing the uncertainty of qualitative analysis[49,51,52,54] because an alternative phenomenon is evaluated: the signal does or does not appear, the analyte is or is not present, etc For $n > 2$ possibilities, e.g. in identification analysis, the use of the relative entropy with logarithm base $b = n$ (Eqs (*1.6*) and (*1.7*)) may be convenient. It will be clear that the base and the units in which the information quantities are expressed have no effect on the practical conclusions drawn from the information-theoretical considerations.

The amount of information derived from multicomponent identification analysis in which n components are simultaneously detected by measuring signal in $m \geqq n$ positions can be expressed as

$$M = -\sum_{i=1}^{n_0} P_0(A_i) \operatorname{lb} P_0(A_i) + \sum_{j=1}^{m} \sum_{i=1}^{n} P(A_i \mid z_j) \operatorname{lb} P(A_i \mid z_j) . \qquad (2.4)$$

This amount of information primarily depends on the selectivity of the experiment; later it will be shown how the selectivity of qualitative or quantitative multicomponent analysis can be characterized by means of entropy that characterizes the a posteriori uncertainty from Eq. (*2.4*) and is given by

$$H(P(A \mid z)) = -\sum_{j=1}^{m} \sum_{i=1}^{n} P(A_i \mid z_j) \operatorname{lb} P(A_i \mid z_j) . \qquad (2.5)$$

It can be regarded as the sum of the component entropies (Eq. (*2.3*)) over all signals $j = 1, 2, \ldots, m$. For qualitative or identification analysis it is calculated from the $(n \times m)$ matrix of conditional probabilities $P(A_i \mid z_j)$ (refs[9,11,28,51,52]). If this matrix is a unit diagonal one, i.e. $P(A_i \mid z_i) = 1$ and $P(A_i \mid z_j) = 0$ $(i \neq j)$, the procedure is perfectly selective. Selectivity as a property of multicomponent analysis will be also dealt with in Paragraph 4.1. Here we only mention the fact that related with the selectivity of the process of multicomponent qualitative or identification analysis is the equivocation[49]

$$E = -\sum_{j=1}^{m} \sum_{i=1}^{n} P(z_j)\, P(A_i \mid z_j) \operatorname{lb} P(A_i \mid z_j) = -\sum_{j=1}^{m} P(z_j)\, H(P(A \mid z_j)) , \qquad (2.6)$$

where $H(P(A \mid z_j))$ is the entropy for conditional probabilities according to Eq. (*2.3*). The probability of appearance of signal, $P(z_j)$ – which in Eq. (*2.6*) has the meaning of a "weight" assigned to the entropies – is

$$P(z_j) = P(A_i) \sum_{i=1}^{n} P(z_j \mid A_i) .$$

Since usually $P(A_i) = 1/n$, equivocation is largely calculated as

$$E = -(1/n) \sum_{j=1}^{m} \sum_{i=1}^{n} P(z_j \mid A_i)\, P(A_i \mid z_j) \operatorname{lb} P(A_i \mid z_j) . \qquad (2.7)$$

Equivocation characterizes the selectivity of the procedure, serves during the optimization of the qualitative multicomponent or identification analysis procedure, enables signal superposition to be quantified, etc.

For instance, in qualitative planar chromatography of three components, which are expected with the the same probability, we obtain Gaussian signals whose width is characterized by the value $\sigma = 0{\cdot}02\, R_F$ units and the attained precision of measurement is $q = 0{\cdot}05\, R_F$ units. What is the equivocation of the entire procedure? Is the equivocation affected by the position z_j in which the signal is read (measured)? The positions in which the signals for components A_1, A_2, A_3 reach their maxima are $z_1 = 0{\cdot}20$, $z_2 = 0{\cdot}25$ and $z_3 = 0{\cdot}35$. These data indicate that the separation of the components, A_1 and A_2 in particular, will be poor. We set up the matrix of probabilities $P(z_j \mid A_i)$ based on experimental data, calculate the sums $\sum P(z_j \mid A_i)$ for all positions z_j, set up the matrix of probabilities $P(A_i \mid z_j)$ by calculation using the Bayes rule (*2.2*), and determine

$$H = -(1/3) \sum_{i=1}^{3} P(z_j \mid A_i)\, P(A_i \mid z_j) \,\mathrm{lb}\, P(A_i \mid z_j)\,.$$

Equivocation then is readily calculated as the sum of H values for those positions z_j in which the signal is read. An overview is presented in Table IV. Reading the signals in the positions of the maxima, i.e. at $z_j = 0{\cdot}20$, $0{\cdot}25$ and $0{\cdot}35$, we have $E = 0{\cdot}310$, but if the signal for component A_1 is sought at $z_j = 0{\cdot}15$, where the value is not maximum but the signal is no more distrubed by the other components, then $E = 0{\cdot}155$ and the entire procedure is more selective. Zero equivocation cannot be achieved in this case because signal of component A_2 will always be disturbed by signals from components A_1 and A_3.

These calculations and considerations derived from them are based on the assumption that the concentration of analyte X_i is always higher than the detection limit, so that small changes in it have no effect on the appearance of the signal, be it a colour reaction or spot identification, the presence of a line at a given wavelength

TABLE IV

Matrices $P(z_j|A_i)$ and $P(A_i|z_j)$ and entropy H

z_j	$P(z_j\|A_i)$ A_1	$P(z_j\|A_i)$ A_2	$P(z_j\|A_i)$ A_3	$\sum P(z_j\|A_i)$	$P(A_i\|z_j)$ A_1	$P(A_i\|z_j)$ A_2	$P(A_i\|z_j)$ A_3	H
0·15	0·105	0	0	0·105	1	0	0	0·000
0·20	0·790	0·105	0	0·895	0·883	0·117	0	0·155
0·25	0·105	0·790	0	0·895	0·117	0·883	0	0·155
0·30	0	0·105	0 105	0·210	0	0·500	0·500	0·070
0·35	0	0	0·790	0·790	0	0	1	0·000
0·40	0	0	0·105	0·105	0	0	1	0·000

in the spectrum, etc. However, if analyte is to be detected at a concentration approaching the detection limit, then $P(z_j \mid X_i)$ depends on X_i. This dependence was studied by Liteanu[51,52], who was able to demonstrate that its shape corresponds to that of the normal distribution function. Probability $P(z_j \mid X_i)$, i.e. the conditional probability that signal appears in position z_j at an analyte concentration X_i, however, is also determined by the reaction sensitivity[29] or, in instrumental analysis, detector sensitivity. Since in qualitative analysis only two possibilities exist, viz. that the signal intensity is or is not discernible from background ($Y = 1$ or $Y = 0$, respectively), the conditional probability has been introduced in ref.[29] that the signal at analyte concentration X_i appears, i.e. $P(Y = 1 \mid X_i)$, or it does not, $P(Y = 0 \mid X_i)$; Shannon's entropy is defined as

$$H(P(X \mid Y = 1)) = -\sum_{i=1}^{n} P(X_i \mid Y = 1) \operatorname{lb} P(X_i \mid Y = 1) \qquad (2.8)$$

and equivocations as

$$E = \sum_{Y_j=0}^{1} P(Y_j)\, H(P(X \mid Y_j))\,; \quad j = 1, 2, \ldots, m\,. \qquad (2.9)$$

The use of these quantities is only reasonable for a rather narrow concentration region, for which $0 \leqq P(Y = 1 \mid X_i) \leqq 1$. This concentration region is referred to by Liteanu[51] as the region of uncertainty of qualitative detection.

All the above considerations apply if a single signal in position z_j corresponds to each analyte A_i; this is commonplace, for instance, in chromatography. In other cases, however – such as emission spectrography – more signals in positions z_j, $j = 1, 2, \ldots, k$, correspond to each analyte. We can introduce, though rather formally, the quantity

$$H(P(A_i \mid z)) = -\sum_{j=1}^{k} P(A_i \mid z_j) \operatorname{lb} P(A_i \mid z_j)\,; \quad i = \text{const.} \qquad (2.10)$$

and refer to it as the signal entropy. Rather than the value of this entropy, it is of importance whether signals for analyte A_i in different positions are disturbed by the same or always a different component of the sample analyzed, whether the component is a majority, minority or trace one, etc. At imperfect selectivity of multicomponent analysis of minority or even trace components, the case where the majority component interferes with the determination is the most unfavourable. The uncertainty of detection of analyte A_i reduces considerably if, based on the absence of signal in another position, that component which interferes with the analyte signal can be ruled out. In principle, it is possible to find a mathematical expression for the information content of the signal taking into account whether and to what extent this signal is disturbed by another signal; this, however, would hardly be of practical value. It is sufficient to take into consideration the fact that in detecting or identifying a certain component, the total composition of the sample and its variability must be known and taken into account. This has been stressed in ref.[9]. All these facts qualitatively agree with experience; information theory, however, enables us to quantitatively express the weight of the effect of the various factors on the uncertainty of results of qualitative or identification analysis.

The main results following from information-theory considerations for the practice of qualitative or identification analysis can be summarized as follows:

1) Results of qualitative and identification analysis are nominal quantities, unamenable to evaluation by statistical methods. Therefore, the application of quantities defined in information theory to the assessment of results and optimization procedures of qualitative and identification analysis is of particular importance.

2) The uncertainty of results of qualitative and identification analysis can be expressed by Shannon's entropy.

3) If the presence of all components is assumed to be equally probable, the entropy reduces to the logarithm of number of these components (Eq. (*2.1*)); else, either entropy according to Eq. (*1.5*) or entropy for conditional probabilities, e.g. according to Eq. (*2.3*) must be used. The a priori uncertainty will frequently be expressed by the conventional probability; the a posteriori uncertainty then – according to circumstances – can also be expressed in conditional probability terms (Eq. (*2.5*)).

4) Conditional probabilities, $P(A_i \mid z_j)$, i.e. probabilities that analyte A_i is present if signal appears in position z_j, enable us:

a) To express the component entropy $H(P(A \mid z_j))$ according to Eq. (*2.3*), which characterizes the information uncertainty of the analytical signal in position z_j and makes it possible to find the most suitable signal positions for the detection of the individual analytes. The conditional probability $P(A_i \mid z_j)$ also serves as a basis for determining the signal entropy $H(P(A_i \mid z))$ according to Eq. (*2.10*) which characterizes the selectivity of the entire procedure.

b) To introduce equivocation (Eqs (*2.6*), (*2.7*) and (*2.9*)) as a characteristics of selectivity of multicomponent qualitative and identification analysis procedure. The optimization of the procedure can be conducted so as to achieve the minimum equivocation.

5) The uncertainty of analyte detection in a concentration approaching the detection limit also depends on the sensitivity of the chemical detection or, in instrumental analysis, on the detector sensitivity. This sensitivity is defined by Eq. (*1.2*) or (*1.3*).

6) Transformation of information obtained by a qualitative experiment into information on the presence or identity of analytes can be differently complicated; invariably, however, the total composition of sample must be taken into account.

7) The most important property of multicomponent qualitative or identification analysis procedures is selectivity, which can be characterized by the value of the entropy $H(P(A \mid z))$ according to Eq. (*2.5*), which can be looked upon as the sum of component entropies over all signals, or of signal entropies over all components. Alternatively, selectivity is expressed by equivocation according to Eqs (*2.6*), (*2.7*) or (*2.9*).

8) If the selectivity of a multicomponent qualitative or identification analysis is imperfect, the overall sample composition and its variability must be taken into account, particularly during the procedure optimization. Most dangerous in multicomponent qualitative analysis of minority or even trace components is the case where the suitable signal is disturbed by a signal corresponding to the majority component.

9) When expressing the information content of results of qualitative or identification analysis according to Eq. (*2.1*), the amount of information according to Eq. (*2.4*), entropy according to Eqs (*2.3*), (*2.5*) or (*2.10*), and equivocation according to Eqs (*2.6*), (*2.7*) and (*2.9*), binary logarithms are frequently employed, because alternative phenomena or two-valued (0,1) quantities are the object of evaluation. The use of logarithms to the base $b = n$ is equally justified when deciding about n possibilities. The logarithm base, however, has no effect upon the conclusions following from the information-theory considerations.

3. QUANTITATIVE SINGLE-COMPONENT ANALYSIS

The results of quantitative analysis answer the question of what the amount of a certain analyte (in single-component analysis) or amounts of more analytes (in multicomponent analysis) in the sample analyzed are.

The system in which the quantitative analysis takes place always consists of a larger number of subsystems. In the first stage, during which information is created, these include, e.g., sampling, sample decomposition and preparation, measurement itself, etc. In the second stage, i.e. the stage of signal processing, the subsystems usually include calibration and calculation of the result and its metrological characteristics. The function of each of the subsystems, their sequence and interrelations, are vital to the process; each, however, also contributes to the random component (sometimes also to the bias) of the a posteriori uncertainty, as demonstrated in refs[11,34,37]. Sometimes the contributions of the operations to the random error or bias can be determined by calculations based on chemical and physico-chemical laws and rules such as the chemical equilibrium law (ref.[5]). Never, however, can an analytical system be regarded as a black-box[2]. The fact that the operations of the analytical procedure affect the final result is the starting point of seeking for the optimal strategy[9], i.e. otimization of the individual operations and seeking for their optimal linking. The way in which information on the signal intensity y or $y(z)$ is transformed to information is of importance as well (Chapter 5).

3.1. *Determination of Majority Components*

The result of quantitative analysis is a continuous random quantity. Its distribution is frequently assumed to be normal, and for higher analyte contents this assumption is usually warranted. In accordance with Eqs (*1.13*) or (*1.14*), the expressing of the information gain is associated with the use of the Kerridge–Bongard measure as a characteristics of the a priori uncertainty; the a posteriori uncertainty is expressed by the entropy (*1.8*) or by the Kerridge–Bongard measure, according to whether accuracy of results is assumed or the occurrence of a bias is admitted. The problems of bias have been dealt with, e.g., in papers[16,17,21,22,24,25,27,28,36] in the series *Theory of Information as Applied in Analytical Chemistry*, in monographs[5,9-11], and elsewhere; however, a purposeful expression of its effect on the information gain of results of quantitative analysis was only made possible by the introduction of the extended divergence measure[7,28,30,32,34-39]. This measure also enables data quality assurance, the laboratory information management system, good laboratory practice, etc., to be put on a rational information-theoretical and system basis[36]. The system approach[11,28,34] made it possible to demonstrate the effect of partial operations of the analytical procedure on the a posteriori uncertainty, i.e. on the precision and accuracy of the results.

Of fundamental importance for determining the information gain of quantitative analysis according to Eq. (*1.4*) is the way the a priori and a posteriori distributions are expressed. If we known no more than that $\mu_x \in \langle x_1, x_2 \rangle$, we use the a priori uniform distribution $U(x_1, x_2)$. The a posteriori distribution is most frequently assumed to be normal (Gaussian), $N(\mu_x, \sigma^2)$; this assumption is sometimes applied to the a priori distribution too. The probability densities of these distributions and their entropies are given in Table II. The a priori uncertainty of quantitative analysis is usually expressed by the Kerridge–Bongard measure because the a priori assumption may not be correct. If it is correct, $H(r, p)$ reduces to $H(p)$: for instance, for $p(x) \rightarrow U(x_1, x_2)$ and $r(x) \rightarrow N(\mu, \sigma^2)$, if $x_1 + 3 \leqq \mu \leqq x_2 - 3\sigma$ we have

$$H(r, p) = \ln\left[1/(x_2 - x_1)\right] = H(p)\,. \qquad (3.1)$$

If the a priori uncertainty or the preinformation is determined by the result of a preliminary analysis, screening for instance, then for $r(x) \rightarrow N(\mu_r, \sigma_r^2)$ and $p(x) \rightarrow N(\mu, \sigma^2)$ the Kerridge–Bongard measure is

$$H(r, p) = \ln \sigma \sqrt{2\pi e^k} + (1/2)\left[(\mu_r - \mu)/\sigma\right]^2\,, \qquad (3.2)$$

where $k = (\sigma_r/\sigma)^2$; again, for $\mu = \mu_r$ and $\sigma^2 = \sigma_r^2$, when $k = 1$, we have $H(r, p) = H(p)$. In Eq. (*3.2*), for the usual values of $0 \leqq k \leqq 1$, the relations are $2{\cdot}507 = \sqrt{2\pi} \leqq \sqrt{2\pi e^k} \leqq \sqrt{2\pi e} = 4{\cdot}133$ and $0{\cdot}919 \leqq \ln \sqrt{2\pi e^k} \leqq 1{\cdot}419$, so that the difference between its maximum and minimum values is 0·5, which is not negligible.

The information gain for a priori uniform and a posteriori normal distributions assuming that the results are accurate and that $x_1 + 3\sigma \leqq \mu \leqq x_2 - 3\sigma$, is given by the divergence measure according to Eq. (*1.13*) and takes the value

$$I(p, p_0) = \ln\left[(x_2 - x_1)/(\sigma\sqrt{2\pi e})\right] . \tag{3.3}$$

This relation has been derived and discussed in monograph[10], and its properties as the measurement information are analyzed in the theoretical work[55]. If the condition $x_1 + 3\sigma \leqq \mu \leqq x_2 - 3\sigma$ is not met – or if $(x_2 - x_1) \leqq 6\sigma$ – the information gain is given[10] by the relation

$$I(p, p_0) = \ln \frac{x_2 - x_1}{[F(z_2) - F(z_1)]\,\sigma\sqrt{2\pi e}} + \frac{1}{2}\left[\frac{z_2 f(z_2) - z_1 f(z_1)}{F(z_2) - F(z_1)}\right], \tag{3.4}$$

where F is the distribution function and f the freqency function of the normal distribution and $z_i = (x_i - \mu)/\sigma$, $i = 1, 2$. Information gain $I(p, p_0)$ then is invariably positive and also depends on the difference between μ and the limiting value of x_1 or x_2. This case is a model of an incorrect a priori assumption or of a situation, rather rare in analytical practice, where the a priori uniform distribution is too narrow, so that $(x_2 - x_1) < 6\sigma$.

The true value of the parameter σ is usually unknown and is estimated as

$$\hat{\sigma} = s = \sqrt{\left(\frac{1}{n_s - 1}\sum_{i=1}^{n_s}(x_i - \bar{x})^2\right)} . \tag{3.5}$$

If the arithmetic mean is reported as the result,

$$\hat{\mu} = \bar{x} = (1/n_a)\sum_{i=1}^{n_a} x_i . \tag{3.6}$$

then $\sigma = s/\sqrt{n_a}$ must be inserted in Eq. (*33*); n_a is the number of parallel determinations from which the average has been calculated. Also, $\sqrt{2\pi e}$ must be replaced by $2t(\alpha, f)$ where $t(\alpha, f)$ is the critical value of Student's distribution for $f = n_s - 1$ where n_s is the number of parallel determinations from which the standard deviation estimate is determined according to Eq. (*3.5*). Additional details concerning the notions used in this treatment can be found, e.g., in refs[5,9]. The α value is chosen so that $t(\alpha, f_{\max}) \approx (1/2)\sqrt{3\pi e} = 2{\cdot}066$ (refs[7,10]). Then we have

$$\hat{I}(p, p_0) = \ln\left\{(x_2 - x_1)\sqrt{n_a}/[2st(\alpha, f)]\right\} . \tag{3.7}$$

This relation has been given in the first paper[12] in the present series, where it served for the discussion of the effect of the numbers of parallel determinations n_a and n_s;

for additional details see refs[7,9–11,37]. Eq. (*3.7*) can be regarded not only as an estimate of the information gain in divergence terms, but – within Brillouin's concept – also as a reduction of the $(x_2 - x_1)$ interval to an interval whose width is $2st(\alpha, f)/\sqrt{n_a}$ (ref.[11]). Some reservations against the use of Brillouin's measure in analytical chemistry are given in ref.[21].

Let us determine the information gain from the determination of 0·05 to 6·00% of manganese in steel if $\sigma = 0{\cdot}006\%$. According to Eq. (*3.3*), the information gain is $I(p,p_0) = \ln\left[(6{\cdot}00 - 0{\cdot}05)/(0{\cdot}006 \,.\, 4{\cdot}133)\right] = 5{\cdot}48$ natural units. Actually, however, the σ value is unknown, therefore from $n_s = 8$ we determine the estimate $s = \doteq 0081$; we take $n_s = 25$ as the maximum performable number of parallel determinations, hence $f_{\max} = 24$. Since $t(0{\cdot}05; 24) = 2{\cdot}064 \approx (1/2)\sqrt{2\pi e}$, we use the critical value $t(0{\cdot}05; 7) = 2{\cdot}365$ and carry out $n_a = 2$ parallel determinations. Then $\hat{I} = \ln\left[(6{\cdot}00 - 0{\cdot}05)\sqrt{2}/(2 \,.\, 0{\cdot}0081 \,.\, 2{\cdot}365)\right] = 5{\cdot}39$ natural units. The difference from $I(p, p_0)$ thus is not too high. If the number of parallel determinations were increased to $n_a = 3$, we would have $\hat{I} = \ln\left[(6{\cdot}00 - 0{\cdot}05)\sqrt{3}/(2 \,.\, 0{\cdot}0081 \,.\, 2{\cdot}365)\right] = 5{\cdot}59$ natural units, and for additional increase to $n_a = 4$ the result would be $\hat{I} = 5{\cdot}74$ natural units. Additional increase in the number of parallel determinations would affect the information gain to an even lower degree.

So far we dealt with accurate results of quantitative analysis. If, however, we must admit that the result may involve a nonzero mean error (bias) $\delta \neq 0$, then the information gain is given by the extended divergence measure as

$$I(r, p, p_0) = \log_b\left[(x_2 - x_1)/(\sigma\sqrt{2\pi e^k})\right] - (1/2)(\delta/\sigma)^2 \log_b e \qquad (3.8)$$

or, in natural units, as

$$I(r, p, p_0) = \ln\left[(x_2 - x_1)/(\sigma\sqrt{2\pi e^k})\right] - (1/2)(\delta/\sigma)^2 \,. \qquad (3.9)$$

In chemical analysis procedures, bias can arise, e.g., as a consequence of the establishment of a "nonquantitative" equilibrium[5]; in instrumental techniques, bias rather arises from improper calibration or from the calibration procedure or standards available being inadequate for a perfect elimination of errors appearing during the analysis. Eq. (*3.9*) is derived in ref.[32] and its properties are demonstrated in refs[35,36].

Relation (*3.3*) for the information gain of accurate results of quantitative analysis can be written in a simpler form as

$$I(p,p_0) = -(1/2)\ln R + (1/2)\ln\left[12/(2\pi e)\right], \qquad (3.10)$$

where the variance reduction[47] $R = (\sigma/\sigma_0)^2$, $R \in (0, 1\rangle$, and $(1/2)\ln\left[12/(2\pi e)\right] = -0{\cdot}176$. For instance, for $\sigma = 0{\cdot}006$ and $x_2 - x_1 = 5{\cdot}95$, hence for the standard deviation of the a priori rectangular distribution, we have $\sigma_0 = 5{\cdot}95/\sqrt{12} = 1{\cdot}7176$, so that $R = 1{\cdot}22 \,.\, 10^{-5}$ and $-(1/2)\ln R = 5{\cdot}657$. The information gain then is $I(p, p_0) = 5{\cdot}657 - 0{\cdot}176 = 5{\cdot}481$, i.e. it attains the same value as calculated in the preceding example. Calculation by Eq. (*3.10*) is, of course, simpler than by the definition relation (*3.8*).

Similarly, Eq. (*3.9*) can also be expressed by means of variance reduction as

$$I(r, p, p_0) = -(1/2)\ln R + (1/2)\ln\left[12/(2\pi e^k)\right] - (1/2)(\delta/\sigma)^2 \,. \qquad (3.11)$$

Here, for $0 \leqq k \leqq 1$ we have $(1/2)\ln\left[12/(2\pi e)\right] = -0{\cdot}176 \leqq (1/2)\ln\left[12/(2\pi e^k)\right] \leqq$ $\leqq +0{\cdot}3235 = (1/2)\ln\left[12/(2\pi)\right]$. Expressing $\delta = \sigma z(\alpha)$ where $z(\alpha)$ is the critical value of normal distribution for the $(1 - \alpha)$ level at which the bias δ is statistically significant, the information gain of results involving this bias is

$$I(r, p, p_0) = (1/2)\left\{-\ln R + \ln\left[12/(2\pi e^k)\right] - z(\alpha)^2\right\} . \qquad (3.12)$$

For the case that the preinformation is a normally distributed result, e.g. of screening analysis[10,17,35,48], the a priori and a posteriori uncertainties are given by the entropy, and the information content according to Eq. (*1.11*) is

$$I = \ln(\sigma/\sigma_0) = -(1/2)\ln R \,, \qquad (3.13)$$

where $R = (\sigma/\sigma_0)^2$ is variance reduction[35,47]. This approach is suitable if good laboratory practice[6] and quality data assurance methods[7,8] prevent bias from appearing (refs[35,36,40]). Information gain for accurate results can be expressed in divergence measure terms[10,17,35] as

$$\begin{aligned} I(p, p_0) &= \ln(\sigma_0/\sigma) + (1/2)\left[(\mu - \mu_0)^2/\sigma_0^2 + (\sigma^2 - \sigma_0^2)/\sigma_0^2\right] = \\ &= -(1/2)\ln R + (1/2)\left[(\mu - \mu_0)^2/\sigma_0^2 + (R - 1)\right] ; \end{aligned} \qquad (3.14)$$

for inaccurate results it can be expressed in extended divergence terms[35] as

$$\begin{aligned} I(r, p, p_0) &= \ln(\sigma_0/\sigma) + (1/2)\left[(\mu - \mu_0)^2/\sigma_0^2 + \right. \\ &\left. + k(\sigma^2 - \sigma_0^2)/\sigma_0^2\right] - (1/2)(\delta/\sigma)^2 \end{aligned} \qquad (3.15)$$

or, on inserting $R = (\sigma/\sigma_0)^2$ and $z(\alpha) = (\delta/\sigma)$, as

$$I(r, p, p_0) = -(1/2)\ln R + (1/2)\left[(\mu - \mu_0)^2/\sigma_0^2 + k(R - 1) - z(\alpha)^2\right] . \qquad (3.16)$$

If k is low enough for the relation $k(R = 1) \ll -\ln R$ to hold true, then

$$I(r, p, p_0) = -(1/2)\ln R + (1/2)\left[(\mu - \mu_0)^2/\sigma_0^2 - z(\alpha)^2\right] , \qquad (3.17)$$

where $z(\alpha)$ is the critical value of normal distribution at the α level at which the bias is statistically significant.

Thus, for accurate results, when $z(\alpha) = 0$, $k = 0{\cdot}0625$, and for $R = 10^{-3}$ we have $-(1/2) \ln R = 3{\cdot}454$; then $(\mu - \mu_0) \leqq \sigma_0^2$ has a negligible effect. For example, for $(\mu - \mu_0) = 0{\cdot}5\sigma_0$ the information gain is $I(r, p, p_0) = 3{\cdot}454 + 0{\cdot}125 = 3{\cdot}579$. However, for a surprising result where, for instance, $(\mu - \mu_0) = 2{\cdot}65\sigma_0$, we have $I(r, p, p_0) = 3{\cdot}454 + 3{\cdot}511 = 6{\cdot}965$ natural units. On the other hand, $(1/2)\, k(R - 1) = 0{\cdot}031$ is virtually negligible against $-(1/2) \ln R = 3{\cdot}454$.

Consider Eqs (*3.3*), (*3.4*) and (*3.7*) through (*3.17*) as mathematical models of various cases of information gain of quantitative analysis; we shall discuss them in some more detail. According to these relations, information content or gain can be expressed by means of three terms. The first, $-(1/2) \ln R$, shows the effect of variance reduction[47], i.e. improvement of the random uncertainty component resulting from the carrying out of the analysis. The second term, $-(1/2)\, z(\alpha)$, characterizes the accuracy of the results, the statistical significance of the bias being more important than its actual magnitude. For accurate results this term vanishes. The third term involves the remaining properties of the result that are included in its uncertainty[35,36]. For instance, information content $I(p, p_0)$ according to Eq. (*3.13*) and information gain $I(r, p, p_0)$ according to Eq. (*3.14*) comprise the $(\mu - \mu_0)/\sigma_0$ term which is no metrological quantity dependent on the properties of the analytical method, which, however, characterizes the "moment of surprise" from the result μ or its "plausibility", i.e. agreement with the a priori assumption μ_0 following, for instance, from the theory[17,30,35]. When expressing the information gain of results which may involve bias $\delta \neq 0$, this third term also depends on the $k = (\sigma_r/\sigma)^2$ ratio, etc.

Relations (*3.8*), (*3.9*), (*3.11*), and (*3.15*) through (*3.17*) enable the $I(r, p, p_0)$ value to be determined in all cases, where the bias δ is known from theory or can be established experimentally. From theory, δ can be determined in titrimetric methods by equilibrium calculations (the so-called titration error[5]) or, in instrumental neutron activation analysis, from the accuracy of adjusting the detector with respect to sample; conclusions concerning the information gain of this method and optimization of its procedure are given in refs[57,56]. In analytical practice, bias is frequently determined by carrying out a blank experiment δ_0. By subtracting it, the bias is eliminated or at least reduced; variance, however, increases, viz. by the blank experiment variance σ_0^2, so that

$$\sigma = \sqrt{(\sigma_i^2 + \sigma_0^2)} \tag{3.18}$$

where σ_i^2 is the initial variance. For $\sigma_i^2 = \sigma_0^2$ we have $\sigma = \sigma_i \sqrt{2} \approx 1{\cdot}414\sigma_i$. In monograph[10], Paragraph 6.5., it is demonstrated that subtraction of the blank $\delta_0 \geqq \sigma_0 \sqrt{\ln 2}$ invariably results in information content increase. For example, in a determination of 0·05–6·0% manganese in steel, $\sigma = 0{\cdot}006\%$. Provided that the results are accurate, $I(p, p_0) = 5{\cdot}48$ natural units. However, verification of the method by using a reference material ($k = 0{\cdot}0625$) reveals that a bias of $\delta = 0{\cdot}008\%$ is present. Since $\sqrt{2\pi e^{k}} = 2{\cdot}586$, we have $I(r, p, p_0) = \ln[(6{\cdot}00 - 0{\cdot}05)/(0{\cdot}006\,.\,2{\cdot}586)] - (1/2)(0{\cdot}008/0{\cdot}006)^2 = 5{\cdot}95 - 0{\cdot}89 = 5{\cdot}06$ natural units. This bias can be eliminated by subtracting the blank; although the standard deviation is thereby increased to $\sigma = 0{\cdot}006\sqrt{2} = 0{\cdot}0085$, we obtain $I(r, p, p_0) = \ln[(6{\cdot}00 - 0{\cdot}05)/(0{\cdot}0085\,.\,2{\cdot}586)] = 5{\cdot}60$ natural units. Thus, subtraction of the blank brough about a substantial increase in the information gain of the result of determination of manganese in steel.

All the above considerations were made assuming that the parameter σ is known; similar conclusions follow from the use of $\hat{I}(p, p_0)$ according to Eq. (*3.7*) or $\hat{I}(r, p, p_0) = \ln[(x_2 - x_1)\sqrt{n_a}/2st(\alpha, f)] - (n_a/2)(d/s)^2$ for the estimate $\hat{\sigma} = s$ and for $\hat{\delta} = d = (X - \bar{x})$. For instance, if we determine 0·1 to 15% Cu by a method of which we do not know whether it is associated with a bias, we analyze a standard sample containing 1·67% Cu. We perform always $n_a = 3$ parallel determinations, so that $t(0{\cdot}0389; 2) = 4{\cdot}9282$ because $z(0{\cdot}0389) = (1/2)\sqrt{2\pi e} = 2{\cdot}066$. The first day we obtained the following results: 1·68, 1·69 and 1·70% Cu, i.e. $\bar{x} = 1{\cdot}69\%$ Cu, $s = 0{\cdot}01$, $d = 0{\cdot}02\%$ Cu. Thus, $\hat{I}(r, p, p_0) = \ln[(15{\cdot}0 - 0{\cdot}01)\sqrt{3}/0{\cdot}01\,.\,2\,.\,4{\cdot}9282] - (3/2)\,.\,(0{\cdot}02/0{\cdot}01)^2 = 5{\cdot}568 - 6{\cdot}000 = -0{\cdot}43$ natural units. The results obtained the next day were: 1·68, 1·67 and 1·66% Cu, $s = 0{\cdot}01\%$ and $d = 0$; thus $I(r, p, p_0) = 5{\cdot}568$ natural units. The blank is small, $d_0 = 0{\cdot}005$, so that the first day's results after its subtraction represent information gain of $\hat{I}(r, p, p_0) = [(15{\cdot}0 - 0{\cdot}01)\,.\,\sqrt{3}/0{\cdot}0141\,.\,2\,.\,4{\cdot}9282] - (3/2)(0{\cdot}02 - 0{\cdot}005/0{\cdot}0141)^2 = 5{\cdot}224 - 1{\cdot}698 = 3{\cdot}526$ natural units. For the second day's results we have $\hat{I}(r, p, p_0) = 5{\cdot}224 - (3/2)\,.\,(-0{\cdot}005/0{\cdot}0141)^2 = 5{\cdot}036$ natural units. This example illustrates the fact that it is appropriate to subtract the blank result even if it only partly compensates the bias, and it also demonstrates that the actual information gain attained by an analysis varies appreciably with the results obtained. Thus, we can speak about the information gain stability in dependence on concrete analytical results. This information gain stability naturally depends on the "ruggedness" of the analytical method, on how its results depend on the conditions in which the analysis occurs.

In practice, however, we do not concern ourselves with mean errors insignificant at a chosen level, e.g. $(1 - \alpha) = 0{\cdot}95$, and regard them as a consequence of random errors. In relations for $I(r, p, p_0)$ and $\hat{I}(r, p, p_0)$, however, any mean error value should be inserted irrespective of whether it is statistically significant or not; the $(1/2)\,z(\alpha)^2$ value e.g. for $\alpha = 0{\cdot}05$ is 1·92 if the information gain is expressed in natural units. Furthermore, determining δ is frequently impracticable, only $d = (X - \bar{x})$ can be determined and the relation $\delta = d$ can be adopted for the analysis, which, however, in view of the existence of the matrix effect may not be true. Moreover, the X value is only known with a precision characterized by the standard deviation σ_r. Therefore, when using Eqs (*3.8*) and (*3.15*) we put $r(x) \to N(X, \sigma_r^2)$,

$k = (\sigma_r/\sigma)^2$ and $z(\alpha) = (\delta/\sigma)$ bearing in mind the fact that this is a first approximation only (see also refs[7],[34-37]). The dependence of $I(r, p, p_0)$ according to Eqs (*3.8*), (*3.9*) and (*3.15*) on the precision and accuracy of the results has been discussed in ref.[35] and is shown in Fig. 1.

The quantity $k = (\sigma_r/\sigma)^2$, $k \in (0, 1\rangle$ characterizes the reliability of verification of the accuracy of analytical results. This also differentiates Eq. (*3.9*) for $\delta = 0$ from Eq. (*3.3*), or (*3.15*) for $\delta = 0$ from Eq. (*3.14*). Relations (*3.3*) and (*3.14*) model a case where $\delta = 0$ is assumed, whereas relations (*3.9*) and (*3.15*) model a case where the existence of a nonzero bias is admitted but it is proved experimentally that $\delta = 0$. Therefore, for $k < 1$ the information gain according to Eqs (*3.9*) and (*3.15*) for $\delta = 0$ is higher than that according to Eqs (*3,3*) and (*3.14*); the difference is a contribution to the information gain following from the fact that a verified method is used. This difference is determined by how realiable the employed standard or reference material (RM) is. Regardless of the a priori distribution, it can be as high as 0·5 natural units (see also refs[32,36,37]).

The following will be clear from what has been said: in cases, which are common in analytical practice, namely that calibration proves that the bias at a chosen $(1 - \alpha)$ level is insignificant but we do not know its actual value, we only know that

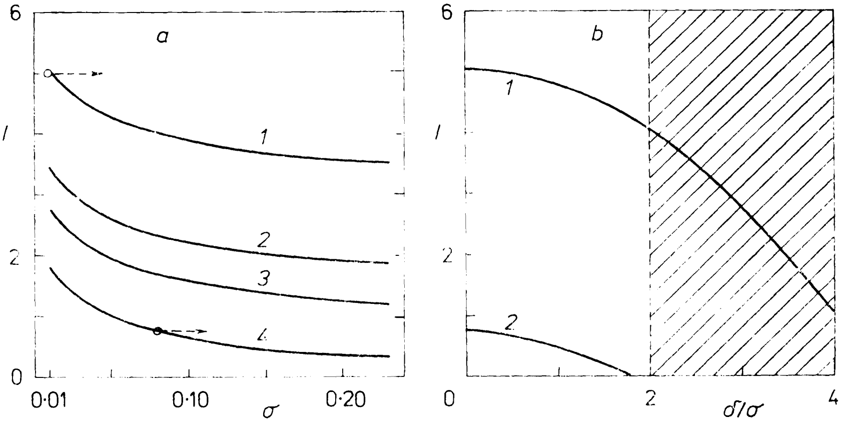


Fig. 1

Dependence of information gain (I) on the precision (σ) and accuracy (δ) of results of quantitative analysis. *a* Dependence of $I(r, p, p_0)$ on σ, $k = 1$, a priori distribution rectangular. Curves: *1* $(x_2 - x_1) = 1\,000$; *3* $(x_2 - x_1) = 10$, a priori distribution normal, $\mu = \mu_0$; *2* $\sigma_0 = 15$; *4* $\sigma_0 = 0{\cdot}6$. *b* Dependence of $I(r, p, p_0)$ on $z = \delta/\sigma$, $k = 1$. Curves: *1* $(x_2 - x_1) = 1\,000$, $\sigma = 0{\cdot}01$; *2* $\sigma_0 = 0{\cdot}6$, $\mu = \mu_0$, $\sigma = 0{\cdot}08$. Hatched region: results involve a mean error statistically significant at the $(1 - \alpha) = 0{\cdot}95$ level

$$\ln \frac{x_2 - x_1}{\sigma_1 \sqrt{2\pi e^k}} - \frac{1}{2} z(\alpha)^2 \leqq I(r, p, p_0) \leqq \ln \frac{x_2 - x_1}{\sigma_2 \sqrt{2\pi e^k}} . \tag{3.19}$$

While the upper limit of this interval is always positive, the lower limit may be negative. For a homoskedastic calibration dependence, $\sigma_1 = \sigma_2$; for a heteroskedastic dependence[32] a value must be inserted for σ_1 such that the information gain is minimal, which may not be the maximum σ value; for a mean error constant over the entire $\langle x_1, x_2 \rangle$ region this is usually the value $\sigma = \delta$. On the other hand, the minimum σ value, valid for the determination of $x \in \langle x_1, x_2 \rangle$, is always inserted for σ_2. Similarly we have

$$\ln \frac{\sigma_0}{\sigma} + \frac{1}{2}\left[\left(\frac{\mu - \mu_0}{\sigma_0}\right)^2 + k\left(\frac{\sigma_1^2 - \sigma_0^2}{\sigma_0^2}\right) - z(\alpha)^2\right] \leqq I(r, p, p_0) \leqq$$

$$\leqq \ln \frac{\sigma_0}{\sigma} + \frac{1}{2}\left[\left(\frac{\mu - \mu_0}{\sigma_0}\right)^2 + k\left(\frac{\sigma_2^2 - \sigma_0^2}{\sigma_0^2}\right)\right] \tag{3.20}$$

if the a priori uncertainty is given by normally distributed results of preinformation obtained by measurement. The upper limit of the interval (3.20) is always positive, the lower may be zero or negative.

Since the mean error δ may be positive or negative, the dependence of $I(r, p, p_0)$ on the result of analysis can be represented by Fig. 2; for details see ref.[7]. This figure also demonstrates the importance of analytical result precision. Results that are low precise against the a posteriori assumption can have a zero or even negative information content if they involve a higher bias; this may be true even if they lie within the confidence interval for the chosen significance level $(1 - \alpha)$. The effect of precision of results against the a priori assumption, e.g. $w = (x_2 - x_1)/\sigma$, is discussed in monograph[11], p. 108.

When using the result of a single-component analysis, e.g. as information that should serve as a basis for a decison, its relevance is given by the information gain solely. Decisive then only is whether this gain is positive, i.e. it really contributes information, or here $I(r, p, p_0) \leqq 0$. Classical information theory does not define negative information, but in view of the pragmatic meaning of analytical result, the case of $I(r, p, p_0) < 0$ can be interpreted as a situation where incorrect results misinform us[37]. To always obtain a positive information gain $I(r, p, p_0)$, the relation

$$\sigma < [(x_2 - x_1)/(\sqrt{2\pi e^k})] \exp [(1/2)\, z(\alpha)^2] \tag{3.21}$$

must be satisfied for the case of a uniform a priori distribution. For the conventional $(1 - \alpha) = 0{\cdot}95$ level, when $z(0{\cdot}05) = 1{\cdot}96$, and for the most frequent case where

$\sigma_{\mathrm{rel}} = 100(\sigma/\mu)$ is highest for $\mu = x_1$, the relation

$$\sigma_{\mathrm{rel}} < (x_2/x_1 - 1) \,.\, 5{\cdot}84/\sqrt{\mathrm{e}^{\mathrm{k}}} \qquad (3.22)$$

must be satisfied, or δ may be significant at a level corresponding to the critical value

$$z(\alpha) < \sqrt{\{\ln [(x_2 - x_1)^2/(2\pi \mathrm{e}^{\mathrm{k}}\sigma^2)]\}} \,. \qquad (3.23)$$

For the a priori normal distribution, the relation

$$z(\alpha) < \sqrt{[(\ln (\sigma_0/\sigma)^2 + (\mu - \mu_0)^2/\sigma_0^2 + k(\sigma^2 - \sigma_0^2/\sigma_0^2)]} \qquad (3.24)$$

or, quite generally,

$$z(\alpha) < \sqrt{2I_0} \qquad (3.25)$$

must be satisfied; here I_0 is $I(r, p, p_0)$ for $\delta = 0$. It is clear that for quantitative determination of higher analyte contents, for $z(0{\cdot}05) = 1{\cdot}96$ the conditions (*3.23*) through (*3.25*) are always fully satisfied. For trace analyses (Paragraph 3.2.), however, attaining a positive information gain may not be easy. For $I_0 \leqq 1{\cdot}92$ a case can occur where the results represent information with zero content or they even

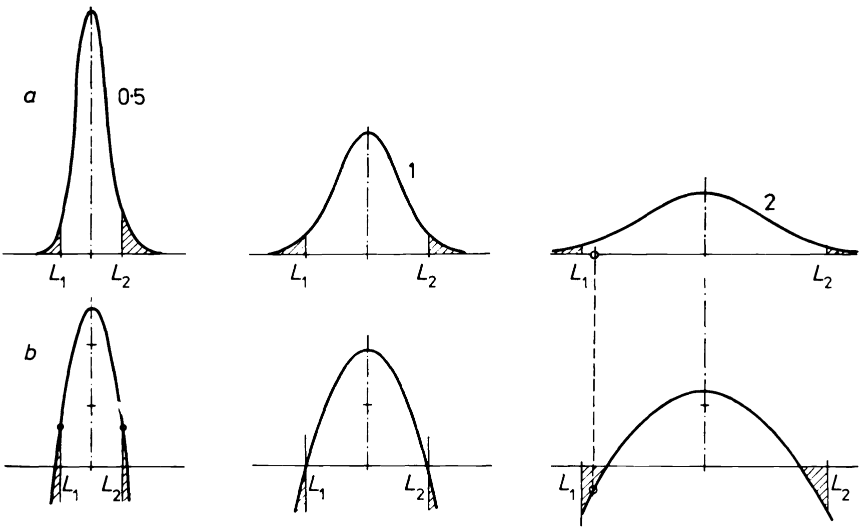


FIG. 2

Dependence of information gain on the result of quantitative analysis. *a* Probability density for $\sigma = 0{\cdot}5, 1{\cdot}0$ and $2{\cdot}0$. *b* $I(r, p, p_0)$ for the corresponding σ; L_1 and L_2 are confidence interval limits for $(1 - \alpha) = 0{\cdot}95$

misinform us; for details see ref.[31]. Fig. 2 for $\mu = 2$ demonstrates a case where the result, which lies within the confidence interval, can have a negative information gain.

The extended divergence measure according to Eq. (*1.14*) can serve as a "warning" model of results involving a bias, hence of a case that should never occur in practice. Owing to the fact that it can also take negative values, $I(r, p, p_0)$ can also serve as a characteristics for assessing to what extent the analytical results approach the truth[58,59], i.e. the actual, objectively existing analyte content of the sample. The use of the divergence measure and the extended divergence measure facilitated the assessment of the information properties of methods of "classical" (chemical) analysis[22], instrumental (physico-chemical) analysis[19,26], methods of analytical quality control[23], etc. Their major importance, however, rests in their role of measures for assessing the results and methods of multicomponent analysis, as will be shown in Chapter 4.

In conclusion of this paragraph, summarize the main results and practical consequences following for analytical practice from information and system-theoretical considerations and quantifications:

1) The simplest characteristics of information properties of accurate results of quantitative chemical analysis is the information gain $I(p, p_0)$; according to Eqs (*3.3*) and (*3.10*), this gain is determined by variance reduction, i.e. the a posteriori to a priori distribution variance ratio. Although the problem is thereby simplified considerably, this measure can be useful in practice.

2) The information gain of accurate results of quantitative analysis $I(p, p_0)$ is also affected by how the result obtained, μ, agrees with the value which is expected in advance, based on theory for instance. For example, if the continuous rectangular distribution $U(x_1, x_2)$ is adopted as the a priori distribution, then the information gain according to Eq. (*3.3*) for $x_1 + 3\sigma \leqq \mu \leqq x_2 - 3\sigma$ depends on $w = (x_2 - x_1)/\sigma$ solely; for other μ values it also depends, according to Eq. (*3.4*), on the difference between μ and x_1 or x_2. If the a priori distribution is normal, then the information gain according to Eq. (*3.14*) depends on variance reduction and on the difference between the mean values of the a priori distribution μ_0 and the a posteriori distribution μ.

3) Information gain expressed by Eq. (*3.7*) for the standard deviation estimate depends on the number of parallel determinations n_s from which the standard deviation was estimated, and on the number of parallel determinations n_a from which the average, which is reported as the final result, is determined. Information gain grows with increasing n_a and n_s values; this growth is rapid at low n values and slow at high n values. From this it follows that performing too many parallel determinations is of no practical value.

4) Information gain of results that can involve a bias, i.e. $I(r, p, p_0)$ according to Eqs (*3.8*), (*3.9*), (*3.15*)–(*3.17*), depends on variance reduction and on the value of the mean error δ or on its statistical significance. The dependences of $I(r, p, p_0)$ on σ and on $z(\alpha) = \delta/\sigma$ for a priori rectangular and normal distributions are shown in Fig. 1, demonstrating that the dependences for the two cases of a priori distribution are similar.

5) The $I(r, p, p_0)$ value depends on the $k = (\sigma_r/\sigma)^2$ ratio, which characterizes the reliability of the reference material (standard) employed for testing the method. If we have $\sigma_r \leqq (1/4)\,\sigma$, $k \leqq 0{\cdot}0625$ and its effect is negligible; at higher k values, however, $I(r, p, p_0)$ decreases appreciably.

6) Information gain $I(p, p_0)$ expressed in divergence measure terms is a model of a case where it is assumed that $\delta = 0$; information gain $I(r, p, p_0)$ expressed in extended divergence measure terms is a model of a case where the occurrence of a nonzero bias is admitted. A case can, however, occur where we prove experimentally that the mean error δ is zero; then, if $k < 1$, the information gain $I(r, p, p_0)$ is higher than $I(p, p_0)$ for which this zero mean error is only assumed. The difference depends on the k value which characterizes the "quality" of the standard used for testing the accuracy of the results and can be as high as 0·5 natural units ("nits").

7) If the calibration dependence is homoskedastic and $\delta = \text{const.}$ for all analyte constants $\mu \in \langle x_1, x_2 \rangle$, the information gain is independent of μ, whereas if it is heteroskedastic, i.e. δ depends on μ, the information gain changes with analyte content, particularly for a nonzero mean error. In well-optimized methods, $I(r, p, p_0)$ attains its maximum for $\mu \approx (1/2)(x_1 + x_2)$, and it is high enough also for $\mu = x_i$ $(i = 1, 2)$.

8) Relevance of results of single-component analysis is primarily given by their information gain. If the ratio $w = (x_2 - x_1)/\sigma$ (for rectangular a priori distribution) or σ_0/σ (for normal a priori distribution) is low, i.e. the results are little precise against the a priori assumption and at the same time they involve a statistically significant mean error, $I(r, p, p_0)$ can attain negative values. Then a situation occurs where such poor results misinform us rather than provide us with relevant information. It is, however, evident that accurate results, even not very precise ones, cannot bring about a zero or even negative information gain[55]. This can be interpreted so that accurate results converge (faster or more slowly according to how they reduce uncertainty) to the "true" value; inaccurate results, however, converge (the faster the more precise they are) to an incorrect value, not corresponding to the reality. Therefore, all provisions referred to as good laboratory practice[6], quality data assurance[7,8] or automated LIMS (Laboratory Information Management System), are aimed at eliminating or at least reducing the hazard that such error occurs (see also ref.[36]).

9) For results of quantitative analysis involving mean error statistically significant at $(1 - \alpha)$ level to always have a positive information gain, I_0 (i.e. $I(r, p, p_0)$ for $\delta = 0$) must always be $I_0 < (1/2)\, z(\alpha)^2$ where $z(\alpha)$ is the critical value of the normal distribution at that significance level at which insignificance of the mean error δ can be assured. For the conventional $(1 - \alpha) = 0.95$ level we must have $I_0 < 1.92$ nit, which is not hard to achieve in analysis of the majority component.

3.2. *Trace Analysis*

Trace analysis, qualitative or quantitative, has some specific features as compared with the detection or determination of the majority components, and these features have largely unfavourable consequences.

1) The a posteriori uncertainty is enhanced due to factors such as sample contamination by analyte, analyte losses by adsorption on the vessel walls, etc.; these phenomena occur during analyses of higher contents too, they are, however, relatively so low that they can be neglected. Some of the factors, contamination for example, cannot be always described by a suitable mathematical model.

2) The a posteriori distribution largely cannot be regarded normal; it is distorted due to the facts that the signal corresponding to analyte concentration cannot be discerned from the background or the signal-to-noise ratio is low, the background is not zero, etc. All these factors must be taken into account when choosing a suitable a posteriori distribution.

3) The relative value of the standard deviation is usually high, the standard deviation estimate inaccurate, and thus even a high mean error seems statistically insignificant. Sometimes it is even claimed that in trace analysis the term "bias" loses meaning. We may or may not agree with this opinion; anyway, results of trace analysis must be handled in a way different from that for results of determination of higher contents, just because of the low relative precision associated with trace analysis.

4) Sensitivity (of the method, of the detector, etc.) plays a major role. It is defined by Eq. (*1.3*). In the determination of majority components its effect is by far not as significant. Calibration alone is associated with a high increase in result uncertainty; suitable reference materials (standards) are mostly lacking, and the matrix effect if often high.

5) Unless the determination of trace analytes is selective, majority components interfere with multicomponent trace analysis, their signals being considerably higher than those of the analytes. Removal of the main portion of the matrix or preconcentration of trace components are operations usually bringing about distortion of the results.

The so-called detection limit is very important in all trace analysis considerations, e.g. in ref.[51]. Kaiser[60] defined it as the analyte concentration which corresponds to the lowest signal y_D discernible from the baseline or from the blank signal y_0 as

$$y_D = y_0 + 3\sigma(y_0) \tag{3.26}$$

i.e. as the concentration

$$x_D = y_D/S = x_0 + 3\sigma(x_0)\,, \tag{3.27}$$

where S is sensitivity, x_0 is analyte concentration corresponding to signal y_0 and $\sigma(x_0)$ is the standard deviation of determination of concentration x_0.

Two cases must be considered in trace analysis:

a) Analyte content is lower than the detection limit, $\mu < x_D$, and signal is not discernible from noise. According to papers[10,11,18,48,61], the information gain then is

$$I(p, p_0) = \ln(x_2/x_D)\,, \tag{3.28}$$

where x_2 is the maximum analyte content expected beforehand, $x_1 = 0$, and x_D is the detection limit for the method used and analyte determined. In practice we always have $x_2 < x_D$ because we would not choose an analytical method that does not enable us to determine the analyte which, as we know in advance, is present in a concentration $\mu_a \leqq x_2$.

b) Analyte content $\mu_a \geqq x_D$, signal is higher than noise, but the signal-to-noise ratio is low and the signal intensity probability distribution is not symmetric. The distribution of the results of trace analysis then cannot be considered symmetric either. It can be described[11,61] by shifted lognormal distribution or by truncated normal distribution.

The probability density of the lognormal distribution shifted to the detection limit $x_D \geqq 0$ is

$$p(x) = \begin{cases} 0 & x \leqq x_D \\ [(x - x_D)\,\sigma\sqrt{2\pi}]^{-1} \exp\{-(1/2)\,[[\ln(x - x_D) - \mu]^2/\sigma^2]\} \end{cases}. \tag{3.29}$$

Putting $\mu = \ln q \,.\, x_D$ we have for the information gain of the result of trace analysis

$$I(p, p_0) = \ln(x_2/x_D)\,[1/(\sigma q\sqrt{2\pi e})]\,, \tag{3.30}$$

and since the $q\sigma$ product is low for analyte content approaching x_D, this gain is invariably higher than that for case *a*) according to Eq. (*3.28*). The probability density for normal distribution truncated in point $z_D = (x_D - \mu)/\sigma$ is

$$p(x) = \begin{cases} 0 & x \leqq x_D \\ \{[1 - F(z_D)]\,\sigma\sqrt{2\pi}\}^{-1} \exp\{-(1/2)\,[(x - \mu)/\sigma]^2\} \end{cases} \tag{3.31}$$

and the information gain of results with this distribution is

$$I(p, p_0) = \ln \frac{x_2 - x_1}{[1 - F(z_D)]\, \sigma \sqrt{2\pi e}} + \frac{1}{2}\left[\frac{z_D f(z_D)}{1 - F(z_D)}\right], \qquad (3.32)$$

where $f(z_D)$ and $F(z_D)$ are the values of the frequency and distribution functions of normal distribution for z_D, respectively. Eq. (*3.32*) is a particular case of Eq. (*3.4*) for $z_1 = z_D$ and for z_2 so high that $F(z_2) \approx 1$ and $f(z_2) \approx 0$. It is clear that as the analyte content increases, the qx_D product in Eq. (*3.30*) approaches unity and in Eq. (*3.32*), when $f(z_D)$ as well as $F(z_D)$ are very low for $z_D \leqq -3$, these relations transform into

$$I(p, p_0) = \ln [x_2/(\sigma \sqrt{2\pi e})], \qquad (3.33)$$

i.e. into the information gain for normal distribution according to Eq. (*3.3*) for $x_1 = 0$.

In trace analysis, the σ value depends considerably on the sensitivity and on the way calibration is performed. This will be dealt with in Chapter 5. Here we demonstrate another effect of sensitivity: relations (*3.28*), (*3.30*), (*3.32*) and (*3.33*) are only meaningful provided that the sensitivity is such that $x_2 \leqq y_{\max}/S$; otherwise at the maximum measurable signal $y_{\max}$ the analyte content x_2 cannot be determined since the signal is beyond scale. If the sensitivity is too low, i.e. $S \ll y_{\max}/x_2$, the σ value, e.g. in Eq. (*3.33*), at a standard deviation of signal measurement σ_y is unnecessarily high. For details see refs[9,19,37].

The limit of determination is important in quantitative trace analysis. It is defined as

$$x_Q = x_0 + w\sigma(x), \qquad (3.34)$$

which is analogous to the definition of the detection limit according Eq. (*3.27*), the standard deviation of determination $\sigma(x)$ being used. The choice of the w value, however, is not unique. Usually, $w = 10$, which according to Currie[62] is associated with the fact that at this value, the precision of results of analyte determination at a concentration x_0 is given by the relative standard deviation value of $\sigma_{rel} = 100 \,.\, (\sigma/\mu_a) = 10\%$, which can be regarded acceptable. Currie's definition led to understanding the limit of determination as the lowest result that can be determined at a sufficient precision. Liteanu and Rica[51] use other precision characteristics as well, e.g. the entropy to express the minimum tolerable reliability; these authors too, however, fail to regard the accuracy of the reult. In his monograph[63], Bayermann defines the limit of determination as the lowest result that is precise and accurate enough to be regarded as a satisfactory estimate of the true analyte content. A positive information gain value can be adopted as a condition of a sufficiently precise and accurate result. The information gain is zero if $(x_2 - x_1)/\sigma = (2\pi e)^{1/2} \exp \,.\, [(1/2)\, z(\alpha)^2]$, where $z(\alpha)$ is the critical value of the normal distribution for the α

level at which the mean error δ is statistically significant. The determination of the analyte concentration $\mu_a < x_2$, $\mu_a \in \langle x_0, x_3 \rangle$, $x_3 \geqq x_2$, then invariably has a positive information gain, so that putting $x_1 = x_0$ we have for the lower limit of nonzero information content

$$x_I = x_2 = x_0 + \sigma(2\pi e)^{1/2} \exp\left[(1/2)\, z(\alpha)^2\right] = x_0 + A(\alpha)\, \sigma\,, \tag{3.35}$$

which is an analogy of the definition of the limit of determination according to Eq. (*3.34*) for $w = A(\alpha)$. In practice, however, particularly during the determination of very low contents, it is difficult to decide whether the mean error is really a bias. Therefore, it has been suggested in ref.[31] that the w value be determined according to the scheme

$$w = \begin{cases} 10 & 0 \leqq z(\alpha) < 1{\cdot}33 \\ A(\alpha) = 4{\cdot}133 \exp\left[(1/2)\, z(\alpha)^2\right] & 1{\cdot}33 \leqq z(\alpha) \leqq 1{\cdot}96 \end{cases} . \tag{3.36}$$

The case $z(\alpha) < 1{\cdot}96$ should never occur in practice because it implies that the results involve an error statistically significant at a level $(1 - \alpha) \geqq 0{\cdot}95$; such error should always be eliminated or at least reduced by modifying the working or calibration procedure or the like. The $A(\alpha)$ values are tabulated in ref.[31].

The detection limit x_D and the nonzero information content limit x_I enable us to distinguish between three regions of trace concentrations μ_a, viz.

1) $\mu_a < x_D$; the presence of analyte cannot be proved by the analytical method in question because the corresponding analytical signal has not appeared. We only know that μ_a can lie within the limits $0 \leqq \mu_a < x_D$, the information content of this being given by Eq. (*3.28*).

2) $x_D \leqq \mu_a < x_I$; the presence of analyte can be proved by the analytical method in question, quantitative determination, however, cannot be carried out without the hazard that the result may have a zero information gain.

3) $\mu_a \geqq x_I$; the analyte can be determined, and if the results do not involve an error statistically significant at a level $(1 - \alpha) < 0{\cdot}95$, they invariably have a positive information gain, which can be expressed by Eq. (*3.33*).

For example, in trace analysis using a method whose detection limit is $x_D = 4 \,.\, 10^{-4}\%$ and performed at a standard deviation $\sigma = 2 \,.\, 10^{-5}\%$, and expecting beforehand no more than $x_2 = 10^{-3}\%$ analyte, the information gain is $I(p,p_0) = \ln(10^{-3}/4 \,.\, 10^{-4}) = 0{\cdot}92$ natural units for $\mu_a < x_D$, $I(p, p_0) = \ln(10^{-3}/2{\cdot}5 \,.\, 10^{-5} \sqrt{2\pi e}) = 2{\cdot}27$ natural units for $\mu_a \gg x_D$, and $I(p, p_0) = \ln(10^{-3}/2{\cdot}5 \,.\, 10^{-5} \,.\, \sqrt{2\pi e}) + (1/2)(-0{\cdot}5 \,.\, 0{\cdot}3521/0{\cdot}3085) - \ln 0{\cdot}3085 = 3{\cdot}16$ natural units for $\mu_a = 4{\cdot}125\%_0$, at which $z = -0{\cdot}5$.

It will be clear that for different methods of determination of a certain analyte in different materials, or for different procedures, calibrations or instrument sensitivities the x_D and x_I values will be different. In this case – as Currie demonstrated – the "a priori", i.e. mean or limiting, value of some metrological characteristics valid for the given analytical method has to be distinguished from its "a posteriori", true, actual value for the given procedure (including the sampling, sample handling, calibration, result calculation, etc.) of detection or determination of a given analyte in the sample of a given matrix. The a priori values of characteristics play a role, e.g., in the choice of the optimum method[9]; the a posteriori, actual values supplement the analytical result and are of importance in procedure optimization[9], quality data assurance[7,8] and in all cases where attaining really relevant analytical information is a matter of importance[7,36,39].

All conclusions drawn for identification, qualitative or quantitative analysis apply to trace analysis as well; in addition, however, the following circumstances – largely unfavourable – play a role too:

1) Information gained by trace analysis can answer the question "what", if this is identification analysis performed near the detection limit of the method employed; otherwise it answers – with a higher or lower uncertainty – the question of "how much" analyte is present.

2) Information content of results of identification of analyte present in an amount approaching the detection limit can be expressed by using entropy for the conditional probability according to Eq. (*2.3*) or (*2.8*) or, if several analytes are identified simultaneously, by using equivocation according to Eq. (*2.9*).

3) Information gain of results of trace analyses must be expressed for various continuous probability distributions, basically according to the relation between the true analyte content μ_a and metrological characteristics of the method such as its detection limit x_D and limit of determination x_Q or the nonzero information gain limit x_I. Only three cases can occur:

a) $\mu_a < x_D$; analytical signal cannot be discerned from background noise, the probability distribution of the results is uniform and the information content is given by Eq. (*3.28*). This result only informs us that analyte can be present at any concentration from zero up to the detection limit.

b) $x_D \leqq \mu_a < x_Q$ or x_I; analytical signal appears, the signal-to-noise ratio, however, is low, and the shifted lognormal or truncated normal distribution must be considered. Information content is given by Eq. (*3.30*) or (*3.32*). Although quantitative determination is possible, the result is so uncertain that the information gain may be zero even at a low mean error. A result from this region thus must be only regarded as a semiquantitative estimate of the true analyte content.

c) $\mu_a > x_I$; the result distribution can be regarded normal, information content is given by Eq. (*3.33*), and information gain is positive unless the results involve a mean error statistically significant at a $(1 - \alpha)$ level for which $z(\alpha) > \ln [(x_2 - x_1)^2/(2\pi e^k \sigma^2)]^{1/2}$.

4) Metrological backup of calibration and quality data assurance procedures is more difficult in trace analysis than in majority component analysis; it is largely necessary to subtract the blank, whose value can be variable, and contamination or analyte losses can hardly be forecast even in a "pure" laboratory. Relative standard deviation of results is usually high, and only high mean errors δ can be disclosed as significant by statistical testing. Thus, there is a point in the opinion that the notion of bias loses sense in trace analysis.

5) Sensitivity, of the instrument detector for instance, plays a more important part here than in qualitative or quantitative analysis of higher contents.

6) In view of what has been said in *1*) through *3*), the sharp difference between qualitative and quantitative analysis vanishes within the analyte content region of $\mu_a \in \langle 0, x_l \rangle$. Results of trace analysis frequently do not enable us more than to give (rather wide) limits within which the true analyte content lies; so they are rather semiquantitative.

7) Since the relative precision of results of trace analysis is usually low, it is appropriate to carry out more parallel determinations than for majority components. Also quality data assurance in routine trace analysis must be more stringent than in routine analyses of higher analyte contents.

4. MULTICOMPONENT ANALYSIS

The amount of information obtained by multicomponent analysis can be regarded as a sum of information contents of the individual determinations according to Eq. (*1.15*) but it must be borne in mind that this relation holds true exactly for independent results only, which, however, is not the case with multicomponent qualitative, identification or quantitative chemical analysis. If nothing else, the results often are all based on the same batch taken, the measurement is preceded by the same sampling and decomposition, sample handling, etc.. which makes the occurrence of a "common" bias possible. Thus, the true amount of information is largely lower than according to Eq. (*1.15*), particularly if the a posteriori uncertainty of the individual results is increased by imperfect selectivity. However, it is always possible to determine the "total", i.e. maximum, potential information content of the whole sequence of signals (spectrum, chromatogram, polarogram, etc.) in multicomponent analysis as[10,38]

$$M_P = [(z_{max} - z_{min})/\Delta z] \ln [(y_{max} - y_{min})/(\sigma_y \sqrt{2\pi e})] , \qquad (4.1)$$

where z_{min}, z_{max}, y_{min} and y_{max} are the lowest and highest, respectively, values of position z and intensity y recorded by the instrument, and Δz is the smallest distance between adjacent signals at which the signals do not affect one another. These distances are different for different signal shapes (profiles)[3,9,10,33].

The application of information theory is more important in multicomponent analysis[33,38] than in single-component analysis because in addition to accuracy and precision, it enables the selectivity, relevance of results, redundancy of the procedure, effect of the signal processing method, etc., to be included in the analytical method assessment and in optimization criteria[64].

Information gain attained by component separation (e.g. in chromatography, electrophoresis, etc.) or by signal resolution (e.g. on a prism or grating in emission spectrography) – irrespective of whether identification, qualitative or quantitative analysis is concerned – can be determined for $i = 1, 2, \ldots, n$ components A_i, signals $y(z_j)$ being measured in positions $j = 1, 2, \ldots, k$, as

$$I_{\mathrm{sep}}^{(j)} = \log n + \sum_{i=1}^{n} a_{ij} \log a_{ij}\,, \quad j = \mathrm{const}\,. \tag{4.2}$$

This quantity always lies within the interval $I_{\mathrm{sep}}^{(j)} \in \langle 0, \log n \rangle$; the maximum value is attained if $-\sum_{i=1}^{n} a_{ij} \log a_{ij} = 0$ when perfect separation is achieved, whereas $I_{\mathrm{sep}}^{(j)} = 0$ if $-\sum_{i=1}^{n} a_{ij} \log a_{ij} = \log n$, i.e. no separation or signal resolution for the components is achieved. We insert into Eq. (4.2)

$$a_{ij} = \begin{cases} P(A_i \mid z_j)\,, & j = \text{const.: qualitative analysis} \\ S_{ij} / \sum_{i=1}^{n} S_{ij}\,, & j = \text{const.: quantitative analysis}\,. \end{cases} \tag{4.3}$$

If a component A_i cannot be detected or determined by means of signal in position z_j, than $a_{ij} = 0$ and we put $a_{ij} \log a_{ij} = 0$. If the signal intensities are measured in k positions $j = 1, 2, \ldots, k$ $(k \geqq n)$, the amount of information obtained by the separation is

$$M_{\mathrm{sep}} = k \log n + \sum_{j=1}^{k} \sum_{i=1}^{n} a_{ij} \log a_{ij}\,. \tag{4.4}$$

If each analyte is detected or determined by measuring a single signal, we have $k = n$ and

$$M_{\mathrm{sep}} = n \log n + \sum_{j=1}^{n} \sum_{i=1}^{n} a_{ij} \log a_{ij}\,. \tag{4.5}$$

The average information gain per component, attained during the separation of n components, is

$$\bar{I}_{\mathrm{sep}} = (1/n)\, M_{\mathrm{sep}} = \log n + (1/n) \sum_{j=1}^{n} \sum_{i=1}^{n} a_{ij} \log a_{ij} \tag{4.6}$$

so that invariably $0 \leqq \bar{I}_{\text{sep}} \leqq \log n$. For perfectly separated signals we have $\bar{I}_{\text{sep}} = \log n$, whereas for no separation, $\bar{I}_{\text{sep}} = 0$. Further details can be found in ref.[33].

4.1. *Selectivity*

Selectivity, a property of fundamental importance in muticomponent analysis, was first defined by Kaiser[65]; to date, eight different definitions exist[66] but none of them is quite universal. It has been demonstrated[33] that selectivity must be regarded as a continuous property; the differentiation between "satisfactory" and "unsatisfactory" selectivities[3] is acceptable from practical aspects only. Rather, it is appropriate to differentiate between cases where selectivity is good enough not to affect the accuracy of results and cases where nonselectivity is so pronounced that signal overlap must be taken into account during results handling.

Consistent with the requirement of universal applicability of selectivity is the quantification based on a quantity analogous to entropy. For signal measurement in position z_j, this quantity is

$$H(a_j) = -\sum_{j=1}^{n} a_{ij} \log a_{ij}\,, \quad \sum_{j=1}^{n} a_{ij} = 1\,, \quad j = \text{const.}\,, \tag{4.7}$$

where a_{ij} is determined according to Eq. (*4.3*), hence, it need not be a probability. For the entire analytical procedure,

$$H(a_{ij}) = -\sum_{j=1}^{n} \sum_{i=1}^{k} a_{ij} \log a_{ij}\,. \tag{4.8}$$

The value of this quantity, which is zero for a perfectly selective procedure, is governed by the matrix $\|a_{ij}\|$, which is the matrix of conditional probabilities or normalized partial sensitivities, where $0 \leqq a_{ij} \leqq 1$. For $n = k$ this matrix is a square one, and if it is a unit diagonal matrix ($a_{ii} = 1$, $a_{ij} = 0$ for $j \neq i$) which is the case when the procedure is pefectly selective (Fig. 3*a*), we have $H(a_{ij}) = 0$; otherwise, $0 < H(a_{ij}) \leqq k \log n$ (Fig. 3*b*), or $0 < H(a_{ij}) \leqq k \log n$ (Fig. 3*c*). It is clear that improvement of the resolving power of the instrument, better resolution of the signals, "autofocusation", etc., bring about decrease in the result uncertainty and thus increase in the amount of information obatined by a multicomponent analysis. For $a_{ij} = P(A_i) \mid z_j)$, Eqs (*4.7*) and (*4.8*) express Shannon's entropy for conditional probabilities; for $a_{ij} = S_{ij}/\sum S_{ij}$, it is more appropriate to regard the quantities (*4.7*) and (*4.8*) as an analogy of entropy, because in the original Boltzmann's concept the thermodynamic entropy and in Shannon's concept the information entropy are defined in probability terms.

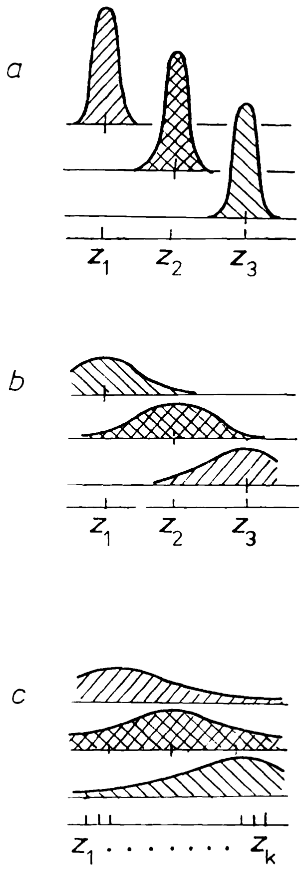
For nonselective procedures it is necessary to process a sequence of analytical signals (spectrum, chromatogram) making allowance for the signal overlap, i.e., to "separate" the interfering signals by calculation, whereupon the standard deviation

σ_y, which is relevant from the information content point of view, increases, and so does the a posteriori uncertainty of the individual determinations.

The selectivity quantification according to Eq. (*4.8*) is universal for qualitative, identification and quantitative analysis and is also applicable to quantitative analysis when overdetermined systems of equations are solved, i.e. k equations of the form $y(z_j) = \sum_i S_{ij}x_i$ in n unknowns, $k > n$, using the least squares condition

$$\sum_{j=1}^{k}(\bar{x}_i - x_{ij})^2 = \min; \quad \left(\bar{x} = (1/k)\sum_{j=1}^{k} x_{ij}\,, \quad i = \text{const.}\,, \quad j = 1, 2, \ldots, k\right).$$

Signals: Matrix $\|a_{ij}\|$ Entropy:



$$\begin{array}{c|c} & \begin{array}{ccc} z_1 & z_2 & z_3 \end{array} \\ \begin{array}{r} i = 1 \\ 2 \\ 3 \end{array} & \left\|\begin{array}{ccc} a_{11} & 0 & 0 \\ 0 & a_{22} & 0 \\ 0 & 0 & a_{33} \end{array}\right\| \end{array}$$

$H = 0$

$$\begin{array}{c|c} & \begin{array}{ccc} z_1 & z_2 & z_3 \end{array} \\ \begin{array}{r} i = 1 \\ 2 \\ 3 \end{array} & \left\|\begin{array}{ccc} a_{11} & a_{12} & 0 \\ a_{21} & a_{22} & a_{23} \\ 0 & a_{32} & a_{33} \end{array}\right\| \end{array}$$

$0 < H < n$.
. $\log n$
$k = n$

$$\begin{array}{c|c} & \begin{array}{ccccccc} z_1 & z_2 & z_3 & \cdots & z_{k-2} & z_{k-1} & z_k \end{array} \\ \begin{array}{r} i = 1 \\ 2 \\ 3 \end{array} & \left\|\begin{array}{ccccccc} a_{11} & a_{12} & a_{13} & \cdots & a_{1k-2} & a_{1k-1} & a_{1k} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2k-2} & a_{2k-1} & a_{2k} \\ a_{31} & a_{32} & a_{33} & \cdots & a_{3k-2} & a_{3k-1} & a_{3k} \end{array}\right\| \end{array}$$

$0 < H < k$.
. $\log n$
$k < n$

FIG. 3

Entropies for various kinds of analytical signals and for various $\|a_{ij}\|$ matrices

Such systems of equations are now solved by using computers; the precision of determination of the individual components can be obtained as well.

Eqs (*4.1*), (*4.4*) and (*4.5*) for the amount of information obtained by multicomponent analysis, and the characteristics of selectivity, i.e. entropy according to Eq. (*4.8*), are suitable for results obtained simultaneously using a single-channel instrument. Results obtained by means of multichannel instruments would apparently call for a different system and information theoretical approach.

4.2. *Relevance and Redundancy*

Any information – including that obtained by multicomponent analysis – has its pragmatic aspect and its content. These two properties, which are not entirely independent, are of importance in assessing the relevance of results in solving a certain analytical problem. In multicomponent analysis, the kind and number of components that can be determined simultaneously are given by the analytical method; information thus obtained can have different relevance in solving different analytical problems. Therefore, the so-called amount of exploitable information is sometimes used[11,39], defined as

$$M_E = \sum_{i=1}^{n} I_i k_i \,, \tag{4.9}$$

where the coefficient of relevance of information about component A_i in solving a given problem is $k_i \in (0, 1\rangle$. If more than one method must be employed to obtain the required information ($j = 1, 2, \ldots, m$ where $m > 1$), the amount of information obtained by a parallel combination of methods[10] is given by

$$M_E = \sum_{i=1}^{n} \sum_{j=1}^{m} I_{ij} k_{ij} \,, \tag{4.10}$$

where I_{ij} is information gain attained in the j-th method and concerning the i-th analyte, and $k_{ij} = \langle 0, 1\rangle$ is the corresponding relevance coefficient. It is clear that we invariably have $0 < M_E \leqq M$, irrespective of whether the results have been obtained by a single method of multicomponent analysis or by several methods of single-component or multicomponent analysis. Details concerning the amount of information deduced from a combination of several methods are given in monograph[10], Paragraph 6.9., where series (successive) and parallel (simultaneous) combinations of analytical methods are differentiated.

The relevance coefficient for i-th analyte can be regarded either as constant, i.e. the static model is treated, or as variable with information content of the result, i.e. the dynamic model is introduced, where

$$\mathrm{d}k_i/\mathrm{d}I_i = f(k_i) \,. \tag{4.11}$$

Two types of dynamic model have found practical application, viz.

a)
$$k_i = \begin{cases} 0 & (I < I_N) \\ k_{i,\max}\{1 - \exp[-a(I - I_N)]\} & (I \geqq I_N) \end{cases} . \qquad (4.12)$$

This model (see Fig. 4*a*) can be used if information is only relevant in case that its content I is equal to or higher than the required information content I_N. Relation (*4.12*) for $a < 0$ is the solution of the equation $dk_i/dI_i = ak + b$; $b > 0$.

b)
$$k_i = k_{i,\max}\{1 + \exp[-a(I - I_{1/2})]\}^{-1} . \qquad (4.13)$$

This model (see Fig. 4*b*) suits well if we require that $I_{i,\min} \leqq I \leqq I_{i,\max}$; $I_{1/2} = (1/2) \cdot (I_{i,\min} + I_{i,\max})$. At $I_{1/2}$ we have $k_i = (1/2)\, k_{i,\max}$. Relation (*4.13*) is the solution of the differential equation $dk_i/dI_i = k(a - bk)$ for $0 < a \leqq b$. Equation (*4.13*) is sometimes referred to as Robertson's growth law.

The static model is actually a particular case of the dynamic model for $dk_i/dI_i = 0$, or $I \gg I_N$ in Eq. (*4.12*) or $I \ll I_{\max}$ in Eq. (*4.13*). When using the static model, the k_i values must be determined in advance for all analytes A_i, $i = 1, 2, \ldots, n$; this also applies to the $k_{i,\max}$ values in the dynamic model. A possible choice[38,39] is $k = 1$ for highly relevant analyte, $k = 0{\cdot}75$ for relevant analyte, $k = 0{\cdot}5$ for low-relevant analyte, and $k = 0{\cdot}25$ for potentially relevant analyte.

It is convenient to consider the relevance coefficient to be a value of the function of membership of information on the *i*-th analyte A_i in the fuzzy subset[66] of informa-

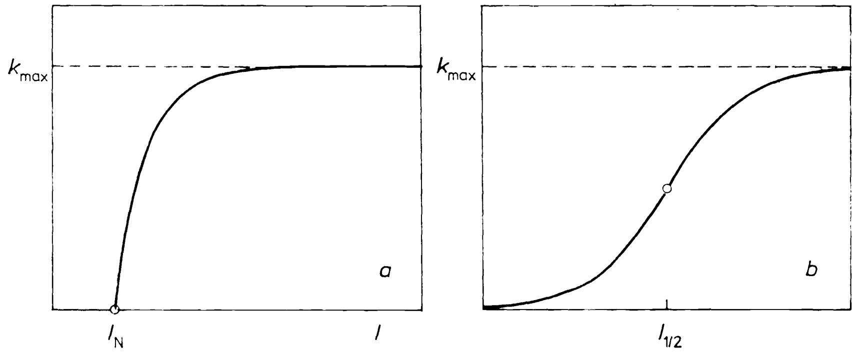


FIG. 4

Dynamic models for relevance coefficient. *a* Model according to Eq. (*4.12*), $I_N = 2$, $a = 1$; *b* model according to Eq. (*4.13*), $I_{1/2} = 6$, $a = 8$

tion relevant to the solution of the given problem. This enables us to address the case of mutually dependent pieces of information; for instance, for two equally relevant but factually entirely independent pieces of information with contents of I_1 and I_2 we have

$$M_E = I_1 k_1 + I_2 k_2 = k \sum_{i=1}^{2} I_i \,, \tag{4.14}$$

where $k = k_1 = k_2$. If, however, information on one component involves some information on another component, Eq. (*4.14*) holds true but $k \leqq k_1 = k_2$. Conversely, if information on component A_1 on its own and information on component A_2 on its own do not make decision possible but rather rich information on A_1 and A_2 contributes to correct decision, then $k \geqq k_1 = k_2$. Irrelevance of result of determination of a component X_i can occur either if the information content is lower than as required for the decision or at $k_i = 0$. The case $I(r, p, p_0) \leqq 0$, however, can also occur during the use of the static or any dynamic model if condition (*3.25*) is satisfied.

If a component can be determined by more methods $j = 1, 2, \ldots, m$, we choose that method for which I_{ij} is highest. When we want to determine some analyte X_i by more, e.g. two, different methods and the difference between the results μ_{ij}, $j = 1, 2$, is not statistically significant e.g. at the $(1 - \alpha) = 0{\cdot}95$ level, we speak about convergence of results of different methods. If the chemical or physico-chemical basis of the two methods is different, such convergence brings valuable evidence of likelihood of the results. In practice, convergence of results of different methods is sought in interlaboratory assays[36].

Another important property of experimentally acquired information is its redundancy, which for results of single-component analysis can be defined as

$$r = 1 - I/I_{\max} = (I_{\max} - I)/I_{\max} = I_l/I_{\max} \tag{4.15}$$

and for results of multicomponent analysis as

$$r = 1 - M/M_{\max} = (M_{\max} - M)/M_{\max} = M_l/M_{\max} \,, \tag{4.16}$$

where I, M are the actually attained, $I_{\max}$, $M_{\max}$ the maximum attainable and I_l, M_l the lost or "excessively" acquired information gain or amount of information, respectively. We can also write $I_l = rI_{\max}$ $(r \in \langle 0, 1 \rangle)$. When a series of parallel determinations are carried out, redundancy can be, according to refs[10,12,67], expressed by inserting in Eq. (*4.15*) as

$$r = (nI_1 - I_n)/nI_1 \,, \tag{4.17}$$

where I_i is information gain attained by carrying out i $(i = 1 \text{ or } n,\ n \geqq 2)$ parallel determinations. The nI_1, I_n, I_l and r values for various numbers of parallel determinations, $\sigma = 0{\cdot}1$ and $x_2 - x_1 = 100$ are given in Table V. It is clear that for $n \leqq 2$, lost information $I_l < I_n$ whereas for n $\geqq 3$, $I_l > I_n$ and $r > 0{\cdot}5$: for three or more parallel determination, more redundant than useful information is generated. But even in a single determination, some redundancy is involved because control analysis of standard must be inserted always after k analyses[6,8,9,73]; then

$$r = [(k + 1)I_1 - kI_1]/(k + 1)I_1 \,. \qquad (4.18)$$

For instance, for $k = 1$ we have $r = 0{\cdot}500$, for $k = 2$ we have $r = 0{\cdot}333$, and for $k = 10$ redundancy is as low as $r = 0{\cdot}091$. In short, the existence of a posteriori uncertainty does not allow us to acquire reliable information free from some redundancy; e.g., it protects the results against gross errors or bias[5] provided that provisions are made such as good laboratory practice and quality data assurance[36]. Subjective confidence in results is also higher if they are fairly precise even if some redundancy is involved.

In the case of multicomponent analysis, we can substitute in Eq. (*4.16*) for $M_{\max} = M_P$ according to Eq. (*4.1*) and M according to Eq. (*1.15*):

$$r = 1 - M/M_P = (M_P - M)/M_P = (C - M)/C \,, \qquad (4.19)$$

where C is the "capacity" of the analytical system in which the multicomponent analysis occurs. This capacity can be defined in a manner similar to that in which Peters (ref.[44], p. 179) defined the capacity of a communication channel, i.e. as the maximum amount of information the analytical system can provide simultaneously. If the information gain I_N or amount of information M_N necessary for the given

TABLE V

Values of nI_1, I_n and I_l in natural units and redundancies r for various n values

n	nI_1	I_n	I_l	r
1	7·92	7·92	0·00	0·000
2	15·84	8·42	7·42	0·468
3	23·76	8·71	15·05	0·633
4	31·68	8·92	22·76	0·718
6	47·52	9·21	38·31	0·806
10	79·20	9·58	69·62	0·879

problem to be solved is known (see also Eq. (*4.12*)), then the redundancy[39] is

$$r = \begin{cases} (I - I_N)/I = (M - M_N)/M & I > I_N, M > M_N \\ 0 & I \leqq I_N, M \leqq M_N \end{cases} . \tag{4.20}$$

But then, for $I < I_N$ or $M < M_N$ the relevance of results for the given problem is poor; therefore, both the relevance and redundancy of the results should always be considered[39]. If the redundancy, written, for instance, as

$$r = (M - M_E)/M = 1 - M_E/M \tag{4.21}$$

is too high ($M \gg M_E$), the potential of the method in multicomponent analysis is apparently made use of in an ineffective and mostly also uneconomical manner.

Redundant is also experimentally acquired information that actually could be obtained by deduction (calculation for instance) from the theory or could be retrieved from the literature.

The economic point of view is also important in practical routine acquisition of information on the chemical composition. Therefore, we use the concept[9,11,69–71] of so-called information profitability

$$P = (1/\tau)\, M \tag{4.22}$$

or information-time profitability

$$P_t = (\varepsilon_t/\tau)\, M \tag{4.23}$$

or exploitable information profitability

$$P_E = (\varepsilon_t/\tau)\, M_E\,, \tag{4.24}$$

where τ is costs that must be spent on the analysis and ε_t is the coefficient of time effectiveness of the analysis which requires time t; this coefficient can be expressed, e.g., as

$$\varepsilon_t = \begin{cases} 1 & t \leqq t_0 - t_1 \\ 1 - [(t - t_0 + t_1)/t_1]^n & t_0 - t_1 < t < t_0 \\ 0 & t > t_0\,, \end{cases} \tag{4.25}$$

where t_0 is time in which we want to know the result and t_1 is time required for the interpretation of the results. The ε_t values according to Eq. (*4.25*) are shown in Fig. 5, which also demonstrates the effect of exponent n.

Illustrative for the practice is the shape of the dependence of P_E on τ, the dynamic model for k_i being used according to Eq. (*4.12*) (Fig. 6). While information gain acquired for the costs spent increases monotonically with increasing τ (Fig. 6*a*), the dependence of P_E on τ (Fig. 6*b*) exhibits a maximum. If insufficient means are available for solving a given problem, then the results are not very relevant (region I in Fig. 6*b*), whereas excessive costs result usually in a high redundancy and lowering of P_E (region II in Fig. 6*b*). The region of optimum exploitation of costs is shown by hatching in Fig. 6*b*; it exhibits the maximum information profitableness for the given dependence of σ on τ.

The application of information theory is most important in multicomponent analysis. Although some problems have attracted attention[3,9–11,19,38,69–71], other

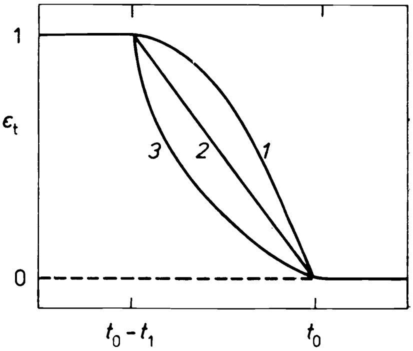


FIG. 5

Time effectivity coefficient ε_t. Value of *n*: *1* 2, *2* 1, *3* 1/2

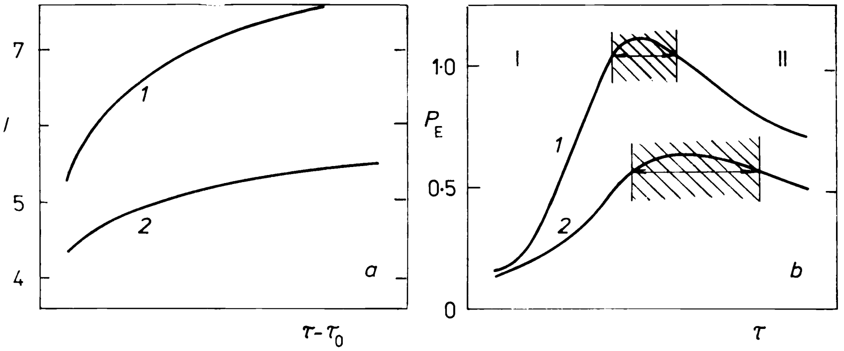


FIG. 6

Dependence of information gain (I) and information profitability (P_E) on costs $\tau = (\tau_1 - \tau_0)$, where τ_0 is minimal costs necessary for performing the analysis. *a* Dependence of $I(r, p, p_0)$ on τ. Curves: *1* $I(r, p, p_0)_1$, $\sigma^2 = 0{\cdot}05/\tau^2$, $(x_2 - x_1) = 200$; *2* $I(r, p, p_0)_2$, $\sigma^2 = 0{\cdot}10/\tau$, $(x_2 - x_1) = 100$. *b* Dependence of $P_E = (1/\tau)\{1 + \exp[-1{\cdot}2(\tau - 4)]\}^{-1}$. $I(r, p, p_0)_1$; *2* $P_E = (1/\tau) . \{1 + \exp[-(\tau - 5)]\}^{-1} I(r, p, p_0)_2$. I is the irrelevance region, II is the redundancy region, the optimum region is hatched problems of multicomponent analysis remain to be solved from the system theory and information theory points of view. These include the problem of suitable expression of the amount of information of dependent results and of results gained during the use of multichannel instruments.

In conclusion we sum up the main results following from the system and information theoretical considerations of multicomponent analysis, which are of significance in routine practice:

1) When considering the suitability of a method of multicomponent analysis for addressing a given analytical problem, the total amount of information is insufficient; the relevance of the results of determination of the individual analytes must also be taken into account. The exploitable amount of information is therefore introduced (Eqs (*4.9*), (*4.10*)) and information profitability is derived from it (Eq. (*4.24*)).

2) In addition to relevance, the redundancy of results of multicomponent analysis is an important criterion as well. Redundancy according to Eq. (*4.19*) shows to what degree the potential amount of information obtainable by the given method for the particular case is exploited, and redundancy according to Eq. (*4.21*) characterizes the excessiveness of the total acquired amount of information with respect to the amount of useful information. If the amount of information required for solving a particular problem is known, redundancy can be expressed by Eq. (*4.20*). Unless excessively high, redundancy is useful, and some redundancy is even necessary.

3) Information theory enables us to quantify the most important property of a multicomponent analysis method, viz. selectivity, by means of quantity (*4.8*), which is either entropy or its analogy defined without the use of probability. Selectivity is a "continuous" property, as characterized by Eqs (*4.7*) and (*4.8*), because it takes values $0 \leqq H(a_j) \leqq \log n$, or $0 \leqq H(a_{ij}) = k \log n$, $i = 1, 2, \ldots, n$; $j = 1, 2, \ldots, k$; $k > n$.

4) Amount of information obtained by separation, M_{sep} (Eq. (*4.4*)), or the average information gain $\bar{I}_{sep}$ (Eq. (*4.6*)) depends on the number of components separated, n, and on how the signals corresponding to the individual components are resolved. This applies to quantitative and identification multicomponent analysis, to all signal shapes and to signals from completely unresolved ones (e.g. in field flow fractionation) to perfectly resolved spectral lines. We invariably have $0 \leqq M_{sep} \leqq n \log n$, or $0 \leqq \bar{I}_{sep} \leqq \log n$, zero corresponding to no separation and the maximum corresponding to perfect separation. The relation between the signal and the $\|a_{ij}\|$ matrix is illustrated in Fig. 3.

5) Those result properties that have been discussed in Chapter 2 for single-component identification analysis have a similar meaning in multicomponent analysis as well.

6) Result precision and accuracy in multicomponent analysis have an effect on the acquired amount of information analogous to that in single-component analysis;

this applies to the determination of higher analyte contents (Paragraph 3.1.) as well as to multicomponent trace analysis (Paragraph 3.2.).

7) The logarithm base used, e.g., in Eqs (*4.1*), (*4.2*) and (*4.4*) through (*4.8*) determines the units in which the information gain or amount is obtained, the conclusions following from the information theoretical considerations, however, are unaffected by it. Therefore, we often choose natural logarithms because in this manner the simplest relations are obtained for the a posteriori normal, truncated normal and lognormal distributions of results of analysis.

8) The application of information theory and the system approach to problems of multicomponent analysis are of higher importance than in the case of single-component analysis, they require, however, some special approaches that were unnecessary in single-component analysis. These include, in particular, problems of relevance and redundancy of the results, combinations of methods, selectivity problems, etc. Some of them have not yet been solved, e.g. the amount of information gained from results that are mutually dependent or that were obtained by using a multichannel instruments, etc.

5. TRANSFORMATION OF SIGNAL INTO ANALYTICAL INFORMATION

Analysis as a process of acquiring information on the qualitative or quantitative chemical composition of sample proceeds so that signal is first formed and this is subsequently decoded into information (Paragraph 1.1.). Signal transformation into information must meet two requirements:

1) All the information included in the analytical signal and relevant to the problem addressed, which is the correct answer to the question asked (Paragraph 1.2.), must be obtained. Various pieces of information can be extracted from the signal, e.g., concerning the kind of some components, the kind and amount of all components, of amounts of selected analytes, etc. This is given basically by if we decode the positions of signals z_j, their intensities in various positions $y(z_j)$, etc. A formal mathematical description of the process of decoding for qualitative, identification or quantitative analysis is given in monograph[11] and in ref.[72].

2) The required information on chemical composition must involve a posteriori uncertainty as low as possible, i.e., it must represent the maximum information gain. Consideration based on information theory and system theory concepts can be useful in seeking for ways to achieve this goal. The following paragraph is devoted to this.

The relation between the analytical signal set and the required information set can be regarded as a binary relation. If the analytical function is a binary relation of the carrying set of signal positions or intensities into the set of analytical information, the calibration function is a relation which is inverse to it. The set of signal positions and analyte identities contains a finite number of elements, and the set of signal intensities in given positions and of analyte contents is given by values continuous on a finite interval $\langle y_{\min}, y_{\max}\rangle$ or $\langle x_1, x_2\rangle$. The quantities z_j, y, and $y(z_j)$, which are determined by measurement, and the quantity x, which is determined from the y or $y(z)$ value, must be regarded as random quantities with continuous probability distributions. For instance, the distribution of z is $p(z)$, whose mean value is

$$\mu_z = \int_{-\infty}^{+\infty} z\, p(z)\, \mathrm{d}z$$

which can take either discrete values only (e.g. in emission spectrography) or continuous values over the $z_j \in \langle z_{\min}, z_{\max}\rangle$ region. Similarly, y and $y(z)$ have distributions $p(y)$ or $p(y(z_j))$ whose mean values

$$\mu_y = \int_{-\infty}^{+\infty} y\, p(y)\, \mathrm{d}y$$

and

$$\mu_{y(z)} = \int_{-\infty}^{+\infty} y(z_j)\, p(y(z_j))\, \mathrm{d}y(z_j)$$

attain continuous values over the intervals $\langle y_{\min}, y_{\max}\rangle$ and $\langle y(z_j)_{\min}, y(z_j)_{\max}\rangle$, respectively.

It is obvious that information on the analyte identity or content must be involved in the signal position or intensity, respectively. Therefore, Mrs Frank and coworkers[72] describe the perfectness of transformation of information on the signal into information on the chemical composition in terms of the so-called transinformation. This measure of mutual information characterizes, for instance, what information on the identity of analyte A_i is contained in the signal in position z_j; it is given as

$$T(A_i, z_j) = \sum_i \sum_j P(A_i, z_j) \log \{P(A_i, z_j)/[P_z(A_i)\, P_A(z_j)]\}\,. \qquad (5.1)$$

Similarly, information on x contained in y is characterized by transinformation for continuous conjugate distributions,

$$T(x, y) = \int_x \int_y p(x, y) \log \{p(x, y)/[p_y(x)\, p_x(y)]\}\, \mathrm{d}x\, \mathrm{d}y\,. \qquad (5.2)$$

In these equations, $P(A_i, z_j)$ is the joint probability and $p(x, y)$ the joint probability density and $P_z(A_i)$, $P_A(z_j)$ are marginal probabilities and $p_y(x)$, $p_y(x)$ are marginal probability densities. The transformation is a symmetric quantity, i.e., for instance $T(x, y) = T(y, x)$, and for two identical random quantities it is the entropy, hence $T(x, x) = H(p(x))$. It is zero if the two quantities are not mutually correlated and neither of them contains information on the other one, and it attains its maximum value if a functional dependence exists between the two quantities. It has been demonstrated[11] that transinformation is a particular case of divergence measure; some additional details can be found in ref.[44]. To a first approximation, the tightness of interdependence between y (or $y(z_j)$) and x can be characterized by the correlation coefficient, which is commonplace in analytical chemistry[5,9,11].

It is appropriate to terminologically and mathematically differentiate between the amount of information contained in the signal and in the final analytical result. Therefore, the notion of the information content is introduced, and we have

$$I_z = \ln\left[(z_{\max} - z_{\min})/(\sigma_z \sqrt{2\pi e})\right] \tag{5.3}$$

for the signal position, and

$$I_y = \begin{cases} \ln\left[(y_{\max} - y_{\min})/(\sigma_y \sqrt{2\pi e}\right) & \text{(one-component analysis)} \\ \ln\left\{\left[y(z_j)_{\max} - y(z_j)_{\min}\right]/\left[\sigma_{y(z)} \sqrt{2\pi e}\right]\right\} & \text{(multicomponent analysis)} \end{cases} \tag{5.4}$$

for the signal intensity. As to the final result, its information gain is considered, and this is expressed by means of Eqs (*3.11*) or (*3.14*) as in refs[32,34,35]

$$\begin{aligned} I_x &= \ln\left[(x_1 - x_2)/(\sigma \sqrt{2\pi e^k})\right] - (1/2)(\delta/\sigma)^2 = \\ &= (1/2)\left\{-\ln R + \ln\left[12/(2\pi e^k)\right] - z(\alpha)^2\right\}, \end{aligned} \tag{5.5}$$

where the variance reduction $R = (\sigma/\sigma_0)^2$; the remaining symbols are as in Paragraph 3.1., and $\ln\left[12/(2\pi e^k)\right] = \ln(1{\cdot}91/e^k)$. In multicomponent analysis, its information contribution is characterized by means of the amount of information, i.e., of the sum of information gains of the individual results.

It is clear that a dependence, determined by the signal decoding procedure, must exist between the information content of the signal (Eq. (*5.4*)) and the information gain (Eq. (*5.5*)). This dependence, however, is by no means simple. For instance, σ is related with σ_y and their interrelation is basically only determined by how information on y is transformed into information on x; the δ and k values in Eq. (*5.5*), however, are related with no quantity from Eq. (*5.4*), and the $\langle x_1, x_2 \rangle$ interval width is very loosely related with the $\langle y_{\min}, y_{\max} \rangle$ interval width, this relation, moreover, depending on sensitivity S according to Eq. (*1.2*) or (*1.3*).

In practical calibration, the most suitable model, i.e. the calibration dependence form f_c according to Eq. (*1.1*), must be first chosen. Frequently, the linear dependence $y = a_0 + a_1 x$ is the dependence of choice; it also corresponds to many physical rules underlying the analytical practice, such as the Lambert–Beer law, Ilkovič's equation, etc. Occasionally, linear segments of generally nonlinear dependences are employed. In chemical ("classical") analysis, where $y = bx$ ($b =$ = const. is the stoichiometric equivalent), the dependence between the standard deviation of the result σ and that of the signal measurement σ_y is given as

$$\sigma = \sigma_y / b \,. \tag{5.6}$$

In instrumental analysis, the relation between σ and σ_y is dependent on the calibration conditions even if the dependence of x on y is linear (but is found by calibration), and it can be generally written as

$$\sigma = (\sigma_y / b)\, A \,. \tag{5.7}$$

For several cases the relation between σ and σ_y is given in Table VI. Additional details are given in refs[9,10,34,37,40,41]. In the standard additions method, the mutual correlation of the signals plays a role as well; this is characterized by the correlation coefficient ϱ, which lies within the interval of $\langle -1, 1 \rangle$, and no correlation exists between the signals at $\varrho = 0$. For alternative descriptions of the dependence of σ on σ_y in the standard additions method see refs[9-11,37].

If the calibration dependence is nonlinear, linearization is achieved by computation transformation or polynomial regression is introduced in the form $y = a_0 + a_1 x + a_2 x^2 + \ldots + a_n x^n = \sum_{i=0}^{n} a_i x_i$, making effort to achieve a good fit with

TABLE VI

Dependences of σ on σ_y for various alternatives of signal intensity (y) conversion into analyte content (x)

Calibration dependence		Dependence of σ on σ_y
$y = bx$	stoichiometric dependence $b =$ const.	$\sigma = (\sigma_y/b) \sqrt{1/n}$
$y = \beta x$	calibration straight line $b = \beta$	$\sigma = (\sigma_y/b) \sqrt{[(1/n) + (1/m) + (\sigma_b/b\sigma_y)^2\, y^2]}$
	standard addition[a]	$\sigma = (\sigma_y/b)\, [(q + 1)/q] \sqrt{[2(1 - \varrho)]}$
$y = \alpha + \beta x$	calibration straight line $a = \alpha$; $b = \beta$	$\sigma = (\sigma_y/b) \sqrt{[(1/n) + (1/m) + (\sigma_b/b\sigma_y)^2\, (y - y)^2]}$

[a] $q = x_S / x_A$, where x_S is the standard addition concentration and x_A is the analyte concentration in sample, ϱ is the correlation coefficient.

the smallest possible number of coefficients a_i. The problem of an approximately linear calibration has been dealt with by Kragten and coworkers[73]. Additional details concerning suitable calibration models are given, e.g., in monographs[5,9,68,74] and in the paper[41]. Expressing the dependence of σ on σ_y is always useful because it demonstrates in which conditions the maximum information profitableness, appropriate redundancy, etc., can be reached.

The a posteriori uncertainty of the analytical result is dependent not only on σ but also on the mean error δ, or on its statistical significance which is given by the ratio $z = (\delta/\sigma)$. If the mean error cannot be determined from the theory[5,57], only the $d = (X - \bar{x})$ value can be found by analysis of a RM and assumed to be the same in the analysis proper. Since the X value is only known with a precision characterized by σ_r, the exponent $k = (\sigma_r/\sigma)^2$ appears in Eq. (*5.5*). As long as $\sigma_r \leqq (1/4)\,\sigma$, the effect of k on the information gain of the analytical result is negligible; if k increases, the information gain drops appreciably[32].

The accomplishment and metrological assurance of suitable calibration is rather difficult, particularly if the dependence of y on x is heteroskedastic[32] or if results of trace analysis are evaluated[51,61]. Usually, no more can be assured by calibration than that the standard deviation of the results is (according ref.[9]) $\sigma \approx \sigma_y$ and that the mean error is statistically insignificant at a level $(1 - \alpha) \leqq 0{\cdot}95$. Then we know no more about the information gain than that it lies within the interval (*3.19*) or (*3.20*), and by adhering to the conditions (*3.21*) through (*3.25*) – which can be achieved by using adequate calibration – we can ensure that information gain $I(r, p, p_0)$ is positive also at the lower limits of the above intervals, so that the results of analysis are associated with true information contributions.

The effect of sensitivity on the information gain is not straightforward. Sensitivity $S = 1/b$ affects the dependence of σ on σ_y on the one hand and the interrelation between $(x_2 - x_1)$ and $(y_{\max} - y_{\min})$ on the other hand. The x_1, x_2 values are given by our a priori assumption concerning the analyte content, whereas $y_{\min}$, $y_{\max}$ are instrumental parameters. It is not always possible to adjust the instrument so that $x_2 = (1/S)\,y_{\max}$; very often, either a part of the signal intensity region remains unused – largely that part in which the measurement is relatively most precise – or the analyte content cannot be determined either near x_1 (insufficient sensitivity) or near x_2 (too high sensitivity). In either case the maximum information gain attainable for the given information content of the signal is not reached. In multicomponent analysis, nearly always a compromise sensitivity must be sought so that the highest possible amount of information be gained while keeping the redundancy reasonably low.

Analytical result obtained by signal processing represents information on the composition of sample, whereas chemometric handling of the results provides information on the composition and its spatial distribution and/or time changes within that part of material reality from which the samples are taken. It is obvious that such treatment of a set of results brings about some – in some cases rather significant – information contribution given by the reduction of the variety. This can be expressed by Eq. (*2.1*), to a first approximation at least. An overview of methods of handling such sets can be found e.g. in ref.[41]; among the most important methods is the pattern recognition method. Information gain attained by performing classification by this method is discussed in Paragraph 6.15 of monograph[10]. In addition to "information compression" achieved by variety reduction, the increase in exploitable information according to Eq. (*4.14*), or the redundancy decrease according to Eq. (*4.21*), also plays a role in the information effect of results handling. These problems call for additional treatment, both mathematical (e.g. using information-theoretical and fuzzy-set theoretical approaches) and methodological and philosophical[58].

From what has been stated in this chapter, the following is of importance for analytical practice:

1) Transformation of analytical signal into analytical information proceeds either by means of a stoichiometric constant (in "classical" chemical quantitative analysis) or by means of the experimentally established calibration dependence.

2) We differentiate between the information content of the analytical signal and the information gain, or amount of information derived from the results of quantitative analysis. The relation between the information content of the signal and the information contribution of the result is determined by how the signal is decoded into the result. For the maximum relevant information to be obtained from a signal of a given content, (i) the optimum model, i.e. shape of the calibration dependence must be determined; (ii) the relation between σ and σ_y must be established for that shape, and the equation for this relation must be employed as the mathematical model in planning experiments to optimize the calibration process; and (iii) the calibration process and its metrological backup must be such that the condition $\sigma \approx \sigma_y$ be met and δ be statistically significant at a level $(1 - \alpha) = 0{\cdot}95$.

3) In choosing the optimum analytical strategy, not only a suitable analytical method, providing a satisfactory signal must be selected, but the calibration must be such that it enable all relevant information contained in the signal to be decoded.

4) A "feedback" should always function between the signal and the result; this should consist in a control of the sensitivity, or batch taken, such that – particularly in routine analysis – the results be associated with a roughly constant and sufficiently high information gain free from unreasonably high redundancy, although some redundancy is necessary.

5) The information effect associated with the handling of a set of results has two aspects: a quantitative aspect, determined by the variety reduction, and a semantic or pragmatic aspect given by the fact that it stimulates developing a chemical working hypothesis or contributes to the corroboration of a given theory[88].

6. CONCLUSIONS

Growing requirements on the number and reliability of analytical results stimulated the seeking for additional phenomena, physical ones in particular, which became the basis of new analytial techniques. This brought about not only advance in practical analytics but also methodological differentiation of analytical chemistry. While – starting from the work of Ostwald[75] – theoretical foundations of the individual analytical methods have been well elaborated, a unified basis of the whole branch of science was still lacking in the early 1970's. Malissa[2] found this basis in the information-theoretical and system approach. At that time, information theory proved to be an efficient tool in the assessment of analytical results and methods and their optimization; a great number of papers and monographs have been devoted to this topic[4,7,9–41]. When it turned out that measures introduced for the purposes of communication techniques are insufficient in metrology and in quantitative analysis, the divergence measure was proposed to serve as measurement information[48,55]. The divergence measure was later extended[34], which made it possible to assess methods and results of quantitative single-component and multicomponent analysis, to seek the optimum analytical strategy[9], and to apply a unified approach to the properties of analytical results and methods[4] and to procedures referred to as good analytical practice and quality data assurance[36]. Irrespective of whether measures taken from communication theory or the divergence measure are used, the information content of the signal or the gain attained by performing the analysis are expressed in a basically identical manner, i.e. in terms of the difference between the a priori and a posteriori uncertainties according to Eq. (*1.4*); only the ways the uncertainties are expressed different. Since logarithmic uncertainty measures are concerned, the difference between logarithms can be regarded as the logarithm of the a priori-to-a posteriori ratio of the numbers of possibilities (Eq. (*2.1*)), intervals (Eq. (*3.7*)) or variances (Eq. (*3.13*); hence, information content can be regarded as a reduction of the number of possibilities, interval width, variances, etc.

The application of information theory to identification and qualitative analysis (Chapter 2) is of particular importance in view of the fact that the results of these analyses are nominal quantities, unamenable to statistical evaluation. Moreover, it suffices to express the uncertainty by means of Shannon's entropy for the a priori and a posteriori probabilities or for conditional probabilities as used for communication purposes. The application to quantitative one – component or multicomponent analysis (Chapters 3 and 4) supplements suitably statistical assessment of results and method optimization; it is, however, based on the divergence measure or extended divergence measure, which enables the effect of not only random ("statistical") error but also of bias to be evaluated. The extended divergence measure also pointed to the importance of metrological backup of the analytical process, calibration in particular (Chapter 5). Many problems of multicomponent analysis

(Chapter 4) can be solved by means of information theory applying the system approach: information theory makes possible quantification[4] of the most important properties of multicomponent analysis and determination of the amount of information gained by separation of the components, enables us – combined with fuzzy set theory – to express the usable amount of information and the corresponding redundancy, and plays a key role in seeking the optimal compromise conditions of the multicomponent quantitative analysis process. It is clear that in multicomponent analysis the combination of information theory methods with the system approach is considerably more important than in single-component analysis. Some problems have not yet been solved satisfactorily enough; for instance, identification of the amount of information gained by multicomponent analysis (Eq. (*1.15*)) with the sum of information gains of all results actually is not fully justified. Some solution for dependent results, though not quite perfect, consists in the introduction of the concept of usable amount of information (Eq. (*4.14*)); this accounts for the semantic dependence but not for, e.g., the mutual correlation of results in the mathematico-statistical sense. Information theory may also help to advance in seeking for a suitable model of multicomponent analysis performed by using a multichannel instrument, for a model of combined ("hyphenated") techniques, etc.

At present, information theory is largely applied in relation to analytical signals, results and methods. Information effect achieved by the treatment of sets of results has been so far expressed (to a first approximation) as variety reduction. In a more comprehensive approach, the semantic and pragmatic aspects should also be addressed in addition to the quantitative aspect given by this reduction (Chapter 5).

Papers in the series *Theory of Information as Applied in Analytical Chemistry*, published in this journal[12–36], are based on expressing the information content of signal and the information gain of results as the difference between the a priori and a posteriori uncertainties, which proves to suit well in practice[9,36–38,56,57,69]. In some cases of analytical methods and results, some alternative way of expressing the information contribution may appear useful. Appropriate measures then can be sought in the monographs by Peters[44], Vajda[76], Aczél and Daróczy[77], Kullback[78] or Khinchin[79]. Entropy is the subject of two papers by Katětov[80]. Of importance in analytical results handling is Kovanic's gnostic theory[81], which is based on the concept of information that is not defined in a probabilistic manner. Many stimulating ideas can be found in two monographs by Brillouin[82,83]. The importance of information theory for modern analytical chemistry has been stressed, e.g., by Currie[84], Kragten[85], Vandeginste[86], Kateman[47], and others. Basic concepts are outlined in monographs[10,11] and in papers[39,40,45]. The importance of the concept of information for the theory of experiment was first pointed-out by Nalimov[87]. The logic of scientific discovery is discussed by Popper[88].

REFERENCES

1. Janák J.: Chem. Listy *73*, 370 (1981).
2. Malissa H.: *Automation in und mit der analytischen Chemie.* Verlag der Wiener Medizin, Akademie, Wien 1972.
3. Doerffel K.: Fresenius Z. Anal. Chem. *330*, 24 (1988).
4. Eckschlager K.: Chem. Listy *83*, 1009 (1989).
5. Eckschlager K.: *Errors, Measurement and Results in Chemical Analysis.* Van Nostrand, London 1969.
6. Horwitz W.: Anal. Chem. *50*, 521A (1987).
7. Musil J., Eckschlager K.: Chem. Listy *81*, 611 (1987).
8. Obrusník I.: Chem. Listy *81*, 1256 (1987); *83*, 561 (1989).
9. Doerffel K., Eckschlager K.: *Optimale Strategien in der Analytik.* VEB Verlag für Grundstoffindustrie, Leipzig 1981.
10. Eckschlager K., Štěpánek V.: *Information Theory as Applied to Chemical Analysis.* Wiley, New York 1979.
11. Eckschlager K., Štěpánek V.: *Analytical Measurement and Information.* Research Studies Press, Letchworth 1985.
12. Eckschlager K.: Collect. Czech. Chem. Commun. *36*, 3016 (1971).
13. Eckschlager K.: Collect. Czech. Chem. Commun. *37*, 137 (1972).
14. Eckschlager K.: Collect. Czech. Chem. Commun. *37*, 1486 (1972).
15. Eckschlager K.: Collect. Czech. Chem. Commun. *38*, 1330 (1973).
16. Eckschlager K.: Collect. Czech. Chem. Commun. *39*, 1426 (1974).
17. Eckschlager K., Vajda I.: Collect. Czech. Chem. Commun. *39*, 3076 (1974).
18. Eckschlager K.: Collect. Czech. Chem. Commun. *40*, 3627 (1975).
19. Eckschlager K.: Collect. Czech. Chem. Commun. *41*, 1875 (1975).
20. Eckschlager K.: Collect. Czech. Chem. Commun. *41*, 2527 (1976).
21. Eckschlager K.: Collect. Czech. Chem. Commun. *42*, 225 (1977).
22. Eckschlager K.: Collect. Czech. Chem. Commun. *42*, 1935 (1977).
23. Eckschlager K.: Collect. Czech. Chem. Commun. *43*, 231 (1978).
24. Eckschlager K.: Collect. Czech. Chem. Commun. *44*, 2373 (1979).
25. Eckschlager K., Štěpánek V.: Collect. Czech. Chem. Commun. *45*, 2516 (1980).
26. Eckschlager K.: Collect. Czech. Chem. Commun. *46*, 478 (1981).
27. Eckschlager K., Štěpánek V.: Collect. Czech. Chem. Commun. *47*, 1195 (1982).
28. Eckschlager K.: Collect. Czech. Chem. Commun. *47*, 1580 (1982).
29. Eckschlager K., Král M.: Collect. Czech. Chem. Commun. *49*, 2342 (1984).
30. Eckschlager K., Štěpánek V.: Collect. Czech. Chem. Commun. *50*, 1359 (1985).
31. Eckschlager K.: Collect. Czech. Chem. Commun. *53*, 1647 (1988).
32. Eckschlager K., Fusek J.: Collect. Czech. Chem. Commun. *53*, 3021 (1988).
33. Eckschlager K.: Collect. Czech. Chem. Commun. *54*, 1770 (1989).
34. Eckschlager K.: Collect. Czech. Chem. Commun. *54*, 3031 (1989).
35. Eckschlager K.: Collect. Czech. Chem. Commun. *55*, 91 (1990).
36. Eckschlager K.: Collect. Czech. Chem. Commun. *55*, 2624 (1990).
37. Danzer K., Eckschlager K., Wienke D.: Fresenius Z. Anal. Chem. *327*, 312 (1987).
38. Danzer K., Eckschlager K., Matherny M.: Fresenius Z. Anal. Chem. *334*, 1 (1989).
39. Eckschlager K., Štěpánek V.: Chemlab. *1*, 273 (1987).
40. Eckschlager K., Štěpánek V., Danzer K.: J. Chemometrics *4*, 195 (1990).
41. Mocák J., Eckschlager K.: Chem. Listy *81*, 1, 126 (1987).

42. Belyaev Yu. J., Koveshnikova T. A.: Zh. Anal. Khim. *27*, 375 (1972); Belyaev Yu. J.: Zh. Anal. Khim. *32*, 2298 (1977).
43. Musil J.: Chem. Listy *78*, 731 (1984).
44. Peters J.: *Einführung in die allgemeine Informationstheorie.* Springer, Heidelberg 1967.
45. Eckschlager K., Štěpánek V.: Anal. Chem. *54*, 1115A (1982).
46. Ehrlich G., Friedrich K., Kucharkowski R., Stahlberg R.: Z. Chem. *24*, 204 (1984).
47. Kateman G.: Anal. Chem. Acta *191*, 215 (1986).
48. Eckschlager K.: Fresenius Z. Anal. Chem. *277*, 1 (1975).
49. Cleij P., Dijkstra A.: Fresenius Z. Anal. Chem. *289*, 97 (1979).
50. Eckschlager K.: Chem. Prum. *32*, 38 (1982).
51. Liteanu C., Rica I.: *Statistical Theory and Methodology of Trace Analysis.* Horwood, Chichester 1979.
52. Liteanu C., Rica I.: Anal. Chem. *51*, 1986 (1979).
53. Malissa H.: Fresenius Z. Anal. Chem. *256*, 7 (1971).
54. Danzer K.: Z. Chem. *13*, 69, 229 (1973).
55. Vajda I., Eckschlager K.: Kybernetika *16*, 120 (1980).
56. Obrusník I., Eckschlager K.: J. Radioanal. Nucl. Chem. *112*, 233 (1987).
57. Obrusník K., Eckschlager K.: Anal. Chem. *62*, 565 (1990).
58. Malissa H.: Fresenius Z. Anal. Chem. *331*, 236 (1988); *333*, 285 (1989).
59. Malissa H.: Mikrochim. Acta *1986*, I., 371.
60. Kaiser H.: Fresenius Z. Anal. Chem. *209*, 1 (1965); *231*, 80 (1966).
61. Eckschlager K., Štěpánek V.: Mikrochim. Acta *1978*, I., 107; *1981*, II., 143.
62. Currie L. A.: Anal. Chem. *40*, 586 (1968).
63. Bayermann K.: *Organische Spurenanalyse.* Thieme, Stuttgart 1982.
64. Danzer K., Hopfe V., Marx G.: Z. Chem. *22*, 332 (1982).
65. Kaiser H.: Fresenius Z. Anal. Chem. *260*, 252 (1972).
66. Otto M., Wegscheider W.: Anal. Chim. Acta *180*, 445 (1986).
67. Malissa H., Rendl J., Marr I. L.: Talanta *22*, 597 (1975).
68. Kateman G., Pijpers F. W.: *Quality Control in Analytical Chemistry.* Wiley, New York 1981.
69. Danzer K., Eckschlager K.: Talanta *25*, 725 (1978).
70. Eckschlager K.: Z. Chem. *16*, 111 (1976).
71. Danzer K.: Z. Chem. *13*, 69 (1973).
72. Frank I., Veress G., Pungor E.: Hungar. Sci. Instr. *54*, 1 (1982).
73. Kragten J., de Jagher P. C., Decnop-Weener L. G.: Anal. Chim. Acta *180*, 457 (1986).
74. Massart D. L., Dijkstra A., Kaufman L.: *Evaluation and Optimization of Laboratory Methods and Analytical Procedures.* Elsevier, Amsterdam 1978.
75. Ostwald W.: *Die wissenschaftlichen Grundlagen der analytischen Chemie.* W. Engelmann, Leipzig 1894.
76. Vajda I.: *Teória informácie a štatistického rozhodovania.* Alfa, Bratislava 1982.
77. Aczél J., Daróczy Z.: *On Measures of Information and Their Characterizations.* Academic Press, New York 1975.
78. Kullback S.: *Information Theory and Statistics.* Wiley, New York 1959.
79. Khinchin A. I.: *Mathematical Foundation of Information Theory.* Dover, New York 1957.
80. Katětov M.: Czech. Math. J. *108*, 564 (1983); *110*, 565 (1985).
81. Kovanic P.: Probl. Control Inform. Theory *13*, 259, 309 (1984).
82. Brillouin L.: *Science and Information Theory.* Academis Press, New York 1962.
83. Brillouin L.: *Scientific Uncertainty and Information.* Academic Press, New York 1966.
84. Currie L. A.: J. Res. Natl. Bur. Stand. *93*, 193 (1988).

85. Kragten G.: Presented at the conference *Compana 88, Jena 1988.*
86. Vandeginste B. G. M.: Top. Curr. Chem. *141*, 1 (1987).
87. Nalimov V. I.: *Teoriya eksperimenta.* Nauka, Moskva 1971.
88. Popper K. R.: *Logik der Forschung.* 8th: Ed. J.C.B. Mohr, Tübingen 1984.

Translated by P. Adámek.