THE APOSTERIORI UNCERTAINTY OF QUALITATIVE AND IDENTIFICATION ANALYSES*

Karel ECKSCHLAGER and Milan KRÁL

Institute of Inorganic Chemistry, Czechoslovak Academy of Sciences, 160 00 Prague 6 and Department of Analytical Chemistry, Prague Institute of Chemical Technology, 166 28 Prague 6

Received January 30th, 1984

It is shown that the uncertainty remaining after a qualitative, identification, chemical or instrumental analysis depends on the structure of the reagent and of the reacting component, on the physico-chemical properties of the arising substance and, in the case of an instrumental proof, also on the properties of the device used.

In the preceding paper of this series[1] it has been shown how the aposteriori uncertainty of the results of a qualitative or identification analysis is connected with selectivity and it has been demonstrated that the aposteriori uncertainty of the results of a quantitative analysis explicitly depends on their accuracy and unbiasedness. The entire exposition was presented on the basis of a black-box approach to the input-output relation of an analytical system. The advantage of this approach is its independence of the chemical or physical substance of the whole analytical process and of the properties of the device used and, therefore, also a fairly general validity of conclusions emerging from it. In fact we usually know relatively a great deal about the substance of an analytical process, *i.e.*, about a process of creating information about the chemical composition by the use of specific methods and types of devices so that, in the pure black-box approach to their evaluation or optimization, we do not utilize some very valuable clues, often to the detriment of the practical utility of the results of the evaluation and optimization of analytical procedures.

In this paper we will show on an example of chemical qualitative or identification analysis how the uncertainty after analysis depends on the structures of the reagent and of the analyte and on the physico-chemical properties of the analyte or of the substance the rise of which is the basis of the proof or of the identification. We will compare the ascertained links with the case of a physico-chemical or pure physical instrumental qualitative or identification analysis, which has been studied earlier[2];

* Part XVIII in the series Theory of Information as Applied to Analytical Chemistry; Part XVII This Journal *47*, 1580 (1982).

in this latter case the aposteriori uncertainty depends rather on the properties of the device creating the analytical signal. Thus we will attempt to proceed from the pure black-box concept of an analytical system towards a chemical or physico-chemical exposition of the causes of the rise of the uncertainty after analysis.

THEORETICAL

In chemical qualitative or identification analysis the grounds of the process of obtaining an analytical signal that contains information about the qualitative composition of the analyzed sample is the chemical reaction between the analyte and the reagent or the interaction of the analyte with energy. In the former case it is desirable that such a product arise by a chemical reaction that is obvious through its colouring or through forming a heterogeneous phase; in the latter case when the interaction is followed by a device the signal in a position characteristic for the identity of the component to be proved must have intensity distinguishable from the random noise. Apart from these conditions some others have to be fulfilled as well so that a particular qualitative proof can be carried out; however, for our exposition only the mentioned basic conditions are important.

The input-output relation for instrumental qualitative or identification analyses with two-dimensional signals has been studied in details by Cleij and Dijkstra[2]; yet in chemical analysis this relationship somewhat varies, especially on account of the diversity of the output from an analytical system. The signal in the output from such a system in the case of a pure chemical proof is one-dimensional (the change of colouring or the rise of heterogeneous phase) and it either appears (it takes on intensity $Y = 1$) or it does not appear ($Y = 0$). The conditional probability that characterizes the input-output relation of this system[1] will be denoted $P(Y \mid X_i)$; thus, *e.g.*, the probability of obtaining a signal ($Y = 1$) in the output if the component X_i is present will read $P(Y = 1 \mid X_i)$ and the probability of missing a signal will be $P(Y = 0 \mid X_i)$. Obviously, for a given component X_i, $P(Y = 1 \mid X_i) + P(Y = 0 \mid X_i) = 1$ and, subsequently, it is sufficient to deal only with $P(Y = 1 \mid X_i)$ in next exposition. Both these probabilities depend on the concentration of X_i and for the purpose of the qualitative analysis we will consider the range of the concentrations limited by the values $x_{0,i}$ (the highest concentration for which we never obtain a signal) and $x_{1,i}$ (the lowest concentration for which a signal always appears). It means that $P(Y = 1 \mid X_i) = 0$ whenever the concentration of X_i lies within $\langle 0, x_{0,i})$ and $P(Y = 1 \mid X_i) = 1$ when the concentration is greater than $x_{1,i}$. The dependence of the frequency of the occurrence of the signal on the concentration in the range $\langle x_{0,i}, x_{1,i})$ was studied by Liteanu and coworkers[3-6]. Conditional probabilities $P(X_i \mid Y = 1)$, *i.e.*, probabilities of the *i*th component being present (in concentration x_i) in the sample, given a signal $Y = 1$ in the output, can be in chemical qualitative analyses determined in a similar way as was shown by Cleij and Dijkstra[2] or by Lite-

anu and Rica[6] in the case of instrumental qualitative analysis, namely by the Bayes's rule

$$P(X_1 \mid Y_j) = \frac{P(X_1)\, P(Y_j \mid X_1)}{\sum_{i=1}^{k} P(X_i)\, P(Y_j \mid X_i)} \quad (Y_j = 0, 1). \tag{1}$$

The probabilities $P(Y = 1 \mid X_i)$ can be estimated in terms of relative frequencies of the occurrence of the signal in the presence of the sole component X_i for various concentration levels; the probabilities $P(X_i)$ are known from pre-information. The Bayes's formula yields the conditional probability of the presence of the component X_1 in concentration greater than $x_{0,1}$ if we obtain a signal $(Y = 1)$ and in the case that the other compounds $X_2, X_3, \ldots, X_k$ (in concentrations greater than $x_{0,i}$; $i = 2, 3, \ldots, k$) react in given conditions equally as the analyte X_1. For $k = 1$, *i.e.*, if only the analyte reacts, we have, of course, $P(X_1 \mid Y_j) = 1$ for any non-zero value $P(Y \mid X_1)$, *i.e.*, $P(Y = 1 \mid X_1) \neq 0$ and $P(Y = 0 \mid X_1) \neq 0$, and it decreases with increasing values of k and for not all probabilities $P(Y = 1 \mid X_i)$ and $P(Y = 0 \mid X_i)$, $i \neq 1$, being equal to zero. Thus for concentrations greater than $x_{1,1}$ the conditional probability $P(X_1 \mid Y = 1)$ depends only on the selectivity of the proof; in addition it depends on the real concentration x_i of the component to be proved when $x_{0,1} \leqq x_1 \leqq x_{1,1}$.

The uncertainty after a qualitative chemical proof can be evaluated in terms of Shannon's entropy[2,6] for conditional probabilities

$$H(X \mid Y = 1) = -\sum_{i=1}^{k} P(X_i \mid Y = 1) \operatorname{ld} P(X_i \mid Y = 1), \tag{2}$$

where ld is the binary logarithm[7]. We put $0 \cdot \operatorname{ld} 0 = 0$. This entropy can be called the compound entropy[2] and it represents the uncertainty with respect to the presence of the components X_i $(i = 1, 2, \ldots, k)$ when the proof is positive. Inequalities $0 \leqq H(X \mid Y) \leqq \operatorname{ld} k$ are always fulfilled. For $P(X_r \mid Y = 1) = 1$, *i.e.*, for a perfectly selective and unambiguous proof of the presence of the component X_r (in concentration greater than $x_{1,r}$) the compound entropy H is zero (the other $P(X_i \mid Y = 1) = 0$, $i \neq r$); in the case that all compounds $X_1, X_2, \ldots, X_k$ have equal probabilities $P(X_i \mid Y = 1)$ the compound entropy takes on its maximum, *i.e.*, $H(X \mid Y = 1) = \operatorname{ld} k$.

The quality of the analytical procedure will be described by the average of the uncertainty after analysis, *i.e.*, by the equivocation

$$E = \sum_{Y_j=0}^{1} P(Y_j)\, H(X \mid Y_j)$$

$$= P(Y = 0)\, H(X \mid Y = 0) + P(Y = 1)\, H(X \mid Y = 1)$$

$$= -P(Y = 0) \sum_{i=1}^{k} P(X_i \mid Y = 0) \operatorname{ld} P(X_i \mid Y = 0) -$$

$$-P(Y = 1) \sum_{i=1}^{k} P(X_i \mid Y = 1) \operatorname{ld} P(X_i \mid Y = 1) . \qquad (3)$$

This applies in the relatively narrow range $x_{0,i} \leqq x_i \leqq x_{1,i}$ $(i = 1, 2, \ldots, k)$. The equivocation assumes its maximum when $P(Y = 0 \mid X_i) = P(Y = 1 \mid X_i) = \frac{1}{2}$ for all i and no component is preferred prior to the analysis.

The Effect of Some Properties upon the Uncertainty

In qualitative and identification chemical or instrumental analyses it is advantageous for practical purposes if: (*a*) the proof is selective or specific; (*b*) if we can prove as a small amount of the analyte as possible, *i.e.*, if the detection limit $x_{1,i}$ is as small as possible or $pD_i = -\log x_{1,i} (x_{1,i} < 1)$ is as large as possible. However, in a chemical proof we achieve high selectivity and p*D* with another means than in an instrumental proof and also the requirements upon the selectivity and the value of p*D* will differ for these two types of analyses.

The selectivity in instrumental qualitative and identification analyses is given by the possibility to discriminate the signal of the analyte from the signals of other components, in which this discrimination can be influenced by the procedure preceding the proper determination, but it is always conditioned by the properties of the device. The dependence of the aposteriori uncertainty of the results of a qualitative instrumental analysis on the selectivity of the procedure has been discussed in the last paper[1] in an interpretation proposed in[2,5]. Also in the case of a chemical analysis we can express the uncertainty in a similar way by employing the entropy in (*2*) or the equivocation in (*3*), which involve analogous mathematical relations as are those for instrumental analyses[1,2,5], yet the meaning of substituted conditional probabilities is somewhat distinct: $P(i \mid j)$ in[1] is the conditional probability of the i th component being present, given a signal in position j in the output while $P(X_i \mid Y = 1)$ from this paper is the conditional probability of the ith component being present in concentration x_i when the reaction occurs $(Y = 1)$ and $P(X_i \mid Y = 0)$ is a similar probability for the case when the reaction does not occur. In chemical analyses the selectivity or the specificity are not determined by our ability to discriminate signals of individual components but they follow from the reality that, under given conditions, only a single component $(k = 1)$ or a few components react in a particular or very similar manner; *cf.* the note on what values $H(X \mid Y)$ in (*2*) takes on in dependence upon k. Equally as in instrumental analysis the entropy characterizes the selectivity of an individual proof and the equivocation means the same for the

selectivity of the entire procedure regardless of the manner of achieving this selectivity. Thus, *e.g.*, the selectivity of the proof of cations by the means of organic reagents[8-10], *i.e.*, the feature that only a few cations or even a single one react with the reagent, is given by the presence of π-electron chromophores in their molecule, while the donor atoms are embodied into the π-electron system or are tied to it. Another time we can achieve an increase of selectivity (and thus a decrease of uncertainty after analysis) in non-selective proofs by suppressing the reactions of strange components, *e.g.*, by masking them, yet basically always only by interfering with the chemical reaction running in the course of the proof.

The detection limit of an instrumental qualitative or identification analysis is given, in its substance, by the least signal distinguishable from the noise or by the concentration x_i corresponding to such a signal. In a chemical proof a low value of the detection limit expressed, for instance, by a high value of $pD_i = -\log x_{1,i}$ depends to great extent on the sensitivity defined, *e.g.*, in Section 2.3 of the monograph[7] as the derivative

$$S_i = \frac{dE[\eta_i]}{dX_i}, \tag{4}$$

where $E[\eta_i]$ is the expected value of the signal intensity η_i. If $y_{i,min}$ is the minimum perceivable intensity of the signal then

$$pD_i = -\log x_{1,i} = -\log y_{i,min} + \log S_i \tag{5}$$

s the greater the weaker signal can be observed and the greater is the sensitivity S_i given in (*4*). The value $x_{1,i}$ can be indeed affected also by the dissociation of a colour complex compound. Both $x_{i,min}$ and S_i depend on physical properties of the product arising by the reaction. For instance, in the case of colouring $y_{i,min}$ depends on the wave length of the maximum absorption and on the spread about the maximum (*i.e.*, on "the colour shade") and both $y_{i,min}$ and S_i depend on the molar absorption coefficient for the maximum absorption. Here too, we know specific relations between the absorption curve and the colour intensity, *i.e.*, between the absorption coefficients of complex compounds and their structures[9,10]. This relationship has been studied for complex compounds of transient metals ions with organic reagents in sufficient details[11,12] and it can be employed in practice.

DISCUSSION

The uncertainty after a chemical or instrumental qualitative or identification analysis as given in (*2*) is influenced by the selectivity and limited by the sensitivity and in the same time by the minimum observable signal intensity (*5*) in dependence on the

detection limit. All these factors make themselves useful independently and they also origin from different sources. The relevance of factors affecting the aposteriori uncertainty differs, of course, according to the purpose that we carry out the proof for, in which it does not matter if we are concerned with chemical or instrumental analyses. In identification analyses the selectivity will be more important; the sensitivity does not become so much evident here, for we can usually work with sufficiently large concentrations. In contrary, in qualitative analyses when the number of possible identities is usually not large (in most cases we determine the presence or the absence of one specific compound) we often prove low concentrations of the analyte and therefore a low value of the detection limit is important. However, in either case it is relevant to work under optimum conditions; if they are not maintained both the selectivity and the detection limit can be affected.

The uncertainty remaining after an instrumental qualitative or identification analysis was carried out, is for physico-chemical methods to considerable extent and for pure physical methods almost exclusively conditioned by the properties of the device employed. In those cases when the effect of the properties of a device upon the uncertainty is little known or too complicated, the black-box approach is a useful starting point to improve the procedure, at least in the first step.

The uncertainty after a qualitative chemical proof is given by: (*a*) the structures of the reagent and of the analyte; (*b*) the structure and some properties (*e.g.*, the dissociation) of the reaction product; (*c*) the way of the course of the reaction that is the grounds of the proof. This is indeed also influenced by whether the proof is carried out in optimum conditions or not. If we use the aposteriori uncertainty, for instance, the compound entropy in (*2*) expressed in terms of conditional probabilities calculated by Bayes's theorem (*1*) as an objective function in optimizing the procedure of a qualitative proof or of an identification analysis, it is necessary for the success of this optimization to realize the conditions, upon which the uncertainty depends. The black-box approach has, in judging an analytical process, a great importance for its universality and independence of the substance of this process; yet, in practical optimization it is expedient to utilize knowledge of special properties of the procedure in the account and to do a chemical and physico-chemical discussion of the causes of the aposteriori uncertainty in such a way as we have shown in this paper in the example of a qualitative or identification proof.

REFERENCES

1. Eckschlager K.: This Journal *47*, 1580 (1982).
2. Cleij P., Dijkstra A.: Fresenius' Z. Anal. Chem. *298*, 97 (1979).
3. Liteanu C., Rica I.: Mikrochim. Acta *1973*, 745.
4. Liteanu C., Rica I.: Mikrochim. Acta *1975* II, 311.
5. Liteanu C., Rica I.: *Statistical Theory and Methodology of Trace Analysis*. E. Horwood, Chichester 1979.

6. Litanu C., Rica I.: Anal. Chem. *51*, 1986 (1979).
7. Eckschlager K., Štěpánek V.: *Information Theory as Applied to Chemical Analysis.* Wiley-Interscience, New York 1979
8. Holzbecher Z., Diviš L., Král M., Šůcha L., Vláčil F.: *Handbook of Organic Reagents in Inorganic Analysis.* E. Horwood, Chichester 1976.
9. Jørgensen C. K.: *Absorption Spectra and Chemical Bonding in Complexes.* Pergamon Press, Oxford 1962.
10. Lever A. B. P.: *Inorganic Electronic Spectroscopy.* Elsevier, Amsterdam 1968.
11. Král M.: Chem. Listy *69*, 435 (1975).
12. Král M.: Chem. Listy *71*, 1 (1977).

Translated by V. Štěpánek.