INFORMATION GAIN AS UNCERTAINTY REDUCTION*

Karel ECKSCHLAGER

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

Received February 24, 1989
Accepted May 19, 1989

Dedicated to Professor Otto Exner on the occasion of his 65th birthday.

It is shown that the various information content or information gain measures can be expressed in terms of the difference between the a priori and a posteriori uncertainties. Measures based on uncertainty expressed by means of variance are compared with those where the uncertainty also involves bias of the results, and it is shown that measures of the two kinds find somewhat different application.

Assessed in terms of uncertainty reduction, experimental information content is usually expressed as

$$I = H_0 - H\,, \tag{1}$$

where H_0 is the a priori uncertainty, existing before the experiment, and H is the a posteriori uncertainty, remaining after the experiment has been performed. These uncertainties can be treated in various ways, according to what we want to ascertain in the experiment. Some measures differing basically only in the way of expressing uncertainty are compared in monograph[1].

In the present paper, information gain measures are compared according to whether the uncertainty considered is only dependent on the variance of the results or if bias is also involved, and fields of their applicability are discussed.

THEORETICAL

If the result of measurement or quantitative analysis is regarded[2] as a continuous random variable ξ, the desired information is the average value of the results $E[\xi] = \mu$. Uncertainty of this information is given by variance $V[\xi] = \sigma^2$ and by bias, the latter being characterized by the mean error $\delta = (\mu - X)$, where X is the true value of the experimentally established quantity.

* Part XXIV in the series Theory of Information as Applied in Analytical Chemistry; Part XXIII: Collect. Czech. Chem. Commun. *54*, 3031 (1989).

Uncertainty can be expressed by using Shannon's entropy, which for a continuous random variable ξ with a probability density $p(x) > 0$, $x \in \langle x_1, x_2 \rangle$, for which

$$\int_{x_1}^{x_2} p(x)\,\mathrm{d}x = 1\,, \tag{2}$$

is given by the relation

$$H(p) = -\int_{x_1}^{x_2} p(x) \log p(x)\,\mathrm{d}x\,. \tag{3}$$

For details see, e.g., refs[1,3].

If both the a priori and a posteriori distributions are normal, $N(\mu_0, \sigma_0^2)$ and $N(\mu, \sigma^2)$, respectively, and $\sigma_0 \geqq \sigma$, then the information gain from the results of measurement is

$$I = \log(\sigma_0/\sigma)\,. \tag{4}$$

In terms of variance reduction $R = (\sigma/\sigma_0)^2$ $(R \in (0, 1\rangle)$, Kateman[4] expresses information as

$$I = -(1/2)\log R\,. \tag{5}$$

In his paper[5], which is one of the first to use information theory for the evaluation of results of chemical experiments, Exner has shown that information obtained by experimental verification of theory is

$$I = -\log\psi\,, \tag{6}$$

where

$$\psi = [\sum_i (x_i - \mu_0)^2 / \sum_i (x_i - \bar{x})^2]^{1/2}\,, \tag{7}$$

in which μ_0 is the value predicted by the theory; the problem of the agreement of the experimental result x_i with theory is also discussed in ref.[5] from the information point of view.

For the a priori uniform distribution $U(x_1, x_2)$ and the a posteriori normal distribution $N(\mu, \sigma^2)$, the information content is

$$I = \log[(x_2 - x_1)/(\sigma\sqrt{(2\pi e)}]\,, \tag{8a}$$

which can be modified to

$$I = \log[(\sigma_0 2\sqrt{3})/(\sigma\sqrt{(2\pi e)})] = -(1/2)\log R + \log 2\sqrt{(3/(2\pi e))} \tag{8b}$$

taking into account the fact that variance of the a priori uniform distribution is $\sigma_0^2 = (x_2 - x_1)^2/12$.

Thus, in Kateman's conception, information content expressed as the entropy difference can be written as

$$I = -(1/2)\log R + A\,. \quad (9)$$

For identical a priori and a posteriori distributions we have $A = 0$, but, for instance, in Eq. (*8b*) we have $A = \log[2\sqrt{(3/(2\pi e))}] < 0$, hence, the difference of the entropies (*8*) is lower than the information gain I according to Eq. (*5*), given by the variance reduction solely.

The concept of information as uncertainty reduction determined not only by variance but also by the bias of the results of measurement or quantitative analysis has led to the use of the divergence measure for expressing the measurement information, and more recently to the introduction of the so-called extended divergence measure[2,6,7], which is convenient for evaluating results and methods of quantitative analysis. A particular case of the divergence measure is the transinformation $T(\xi, \eta)$, employed[1,8] for evaluating the process of acquiring information about a random quantity ξ by measuring another quantity η, the two quantities being correlated.

The divergence measure is easy to derive by means of the Kerridge–Bongard inaccuracy measure, which for a continuous distribution is

$$H(r, p) = H(r) + D(r, p) = -\int_{x_1}^{x_2} r(x)\log p(x)\,dx\,, \quad (10)$$

where the "error term"

$$D(r, p) = -\int_{x_1}^{x_2} r(x)\log[r(x)/p(x)]\,dx \quad (11)$$

is a measure of divergence of the true distribution $r(x)$ and the experimentally established distribution $p(x)$. The divergence measure of the information content

$$I(p, p_0) = H(p, p_0) - H(p) = \int_{x_1}^{x_2} p(x)\log[p(x)/p_0(x)]\,dx = D(p, p_0) \quad (12)$$

is used if the results of measurement or quantitative analysis are unbiassed, whereas the extended divergence measure

$$I(r; p, p_0) = H(r, p_0) - H(r, p) = \int_{x_1}^{x_2} r(x)\log[p(x)/p_0(x)]\,dx \quad (13)$$

is employed if the results may be biassed.

For both the a priori and a posteriori distributions normal, we have in natural logarithm terms

$$I(p, p_0) = \ln(\sigma_0/\sigma) + (1/2)\left[(\mu - \mu_0)^2/\sigma_0^2 + (\sigma^2 - \sigma_0^2)/\sigma_0^2\right] \quad (14a)$$

which, denoting the second right-hand term A_1, can be written in a contracted form as

$$I(p, p_0) = -(1/2)\ln R + A_1 \,. \qquad (14b)$$

While A_1 can be positive, zero or negative, expression (*14*) is always non-negative.

The difference between the result μ and the value μ_0 established, e.g., by preliminary measurement or predicted by theory, can be looked upon as a measure of surprise by the result obtained. Kuhn[9] has discussed how unexpected experimental results can stimulate extension or even revision of established theories; so we consider the information-theoretical treatment of the measure of surprise from a result worthwhile.

For a priori uniform and a posteriori normal distributions, the divergence measure (*12*) leads to Eqs (*8*); no measure of surprise occurs here because we know in advance that $\mu \in \langle x_1, x_2 \rangle$.

Relation (*12*) is useful for expressing information content of an analytical signal or information gain from a result of measurement[2]; for expressing the information gain from a quantitative analysis where imperfect analytical operations, matrix effect or inadequate calibration give rise to a systematic error (bias) $\delta = (\mu - X)$ ($|\delta| > 0$), the extended divergence measure[2,6,7] according to Eq. (*13*) must be used. For the a priori, a posteriori and true distributions all normal, $N(\mu_0, \sigma_0^2)$, $N(\mu, \sigma^2)$ and $N(X, \sigma_r^2)$, respectively[7], we have in natural logarithm terms

$$I(r; p, p_0) = \ln(\sigma_0/\sigma) + (1/2)\left[(\mu - \mu_0)^2/\sigma_0^2 - (\mu - X)^2/\sigma^2 + k(\sigma^2 - \sigma_0^2)/\sigma_0^2\right], \qquad (15a)$$

where $k = (\sigma_r/\sigma)^2$. Denoting the second right-hand term A_2 we obtain in the contracted form

$$I(r; p, p_0) = -(1/2)\ln R + A_2 \,. \qquad (15b)$$

Again, A_2 can be positive, zero or negative. It will be clear that for $\delta = 0$ and $k = 1$, A_2 reduces to A_1 in Eq. (*14b*). For the a priori distribution uniform and the a posteriori and the true distributions normal ($N(\mu, \sigma^2)$ and $N(X, \sigma_r^2)$, respectively), the information gain in natural logarithm terms is

$$I(r; p, p_0) = \ln\left[(x_2 - x_1)/(\sigma\sqrt{(2\pi e^k)})\right) - (1/2)(\delta/\sigma)^2 \qquad (16a)$$

or

$$I(r; p, p_0) = \ln\left[(\sigma_0\, 2\sqrt{3})/(\sigma\sqrt{(2\pi e^k)})\right) - (1/2)\, z^2 \qquad (16b)$$

or

$$I(r; p, p_0) = -(1/2)\ln R + A_3 \,, \qquad (16c)$$

where the term

$$A_3 = \ln \left[(2 \sqrt{3}) / (\sqrt{(2\pi e^k)}) - (1/2)\, z^2 \right] \qquad (17)$$

is positive at low $k = (\sigma_r/\sigma)^2$ and $z = \delta/\sigma$ values and negative at high k and z values. For $k = 1$ and $\delta = 0$, relations (*16*) reduce to relations (*8*).

At statistically highly significant errors δ, expressions (*15*) and (*16*) take negative values, which can be interpreted[2,7,10] as a case where incorrect results misinform us. Relations for the information gain, expressed by means of the divergence measure (*12*) or the extended divergence measure (*13*), can be also interpreted in terms of the variance reduction (*5*); however, quantities characterizing, e.g., the moment of surprise $(\mu - \mu_0)$, the quality of the standard employed for verifying the experimental method $k = (\sigma_r/\sigma)^2$ (refs[2,7]), error δ or its statistical significance, characterized by the critical value of the normal distribution $z(\alpha)$ at the $(1 - \alpha)$ significance level, i.e. by the value $z(\alpha) = z = \delta/\sigma$ from Eq. (*16b*), play here a role as well.

DISCUSSION AND CONCLUSIONS

Although initially derived as divergence, i.e. dissimilarity of the a priori and a posteriori distributions, the divergence measure can be regarded as a difference of uncertainties (Eq. (*12*)), the a priori uncertainty being expressed in terms of the Kerridge–Bongard inaccuracy measure (*10*) and the a posteriori uncertainty, in terms of the entropy (*3*). Extension of this measure to relation (*13*) is given by the use of the inaccuracy measure for expressing both the a priori and a posteriori uncertainties. Against Kateman's concept[4] of information as the logarithm of the variance reduction, the extended divergence measure involves also the result accuracy aspect. If the concept of an objectively existing truth which we want to approach during the process of empiric knowledge gaining is adopted not only for analytical chemistry, as suggested by Malissa[11], but for any experimental activity, we must consider highly important that property of the extended divergence measure enabling us to evaluate results also with respect to their accuracy, i.e., according to how they agree with the true value, construed[2] as the expected value of the true distribution $r(x)$.

Expressing information gain by means of variance reduction can be regarded suitable in cases where the experiment is metrologically backed up so perfectly that no bias of the results needs to be assumed. To an extent, relations (*15*) and (*16*) for $\delta \neq 0$ can be treated as a model of a case which should not occur in practice and should be prevented by using appropriate methods of data quality control[12,13]. Actual results of measurement or quantitative analysis may contain error, and so the extended divergence measure can be of use when selecting a suitable analytical or measuring method, during the optimization of an experimental procedure, etc. Of practical utility is the fact that this measure enables us to estimate, approximately at least, such a priori[14] metrological characteristics as the results must assume to provide actual information gain. In this manner, for instance, the limit of nonzero information content[15] has been determined as the lower limit of applicability of an analytical method, the highest tolerable value, statistically significant at the $(1 - \alpha)$ level of significance, of the relative standard deviation[16] of biassed results, etc. Making use of the extended divergence measure, it is possible to choose the most suitable calibration procedure; etc. As to other properties of the divergence and extended divergence measures and to practical conclusions that can be derived from them, the reader is referred to the preceding papers in this series[2,6,7,15,17] and to ref.[10]. The essence of all these cases is the application of general mathematical models to the solution of particular problems of experimental practice, which is a modern trend in chemometrics.

REFERENCES

1. Eckschlager K., Štěpánek V.: *Analytical Measurement and Information.* Research Studies Press, Letchworth 1985.
2. Eckschlager K.: Collect. Czech. Chem. Commun. *54*, 3031 (1989).
3. Peters J.: *Einführung in die allgemeine Informationstheorie.* Springer, Berlin 1967.
4. Kateman G.: Anal. Chim. Acta *191*, 125 (1986).
5. Exner O.: Collect. Czech. Chem. Commun. *31*, 3222 (1956).
6. Eckschlager K.: Collect. Czech. Chem. Commun. *47*, 1580 (1982).
7. Eckschlager K., Fusek J.: Collect. Czech. Chem. Commun. *53*, 3021 (1988).
8. Frank I., Veress G., Pungor E.: Hung. Sci. Instr. *54*, 1 (1982).
9. Kuhn T. S.: *The Structure of Scientific Revolutions.* University of Chicago Press, Chicago 1962.
10. Danzer K., Eckschlager K., Wienke D.: Fresenius Z. Anal. Chem. *327*, 312 (1987).
11. Malissa H.: Fresenius Z. Anal. Chem. *331*, 236 (1988).
12. Kateman G., Pijpers F. W.: *Quality Control in Analytical Chemistry.* Wiley, New York 1981.
13. Taylor J. K.: Anal. Chem. *53*, 1588 (1981); *55*, 600 A (1983).
14. Currie L. A.: Anal. Chem. *40*, 586 (1968).
15. Eckschlager K.: Collect. Czech. Chem. Commun. *53*, 1647 (1988).
16. Eckschlager K.: Chem. Prum. *39*, 583 (1989).
17. Eckschlager K., Štěpánek V.: Collect. Czech. Chem. Commun. *50*, 1359 (1985).

Translated by P. Adámek.