EXPERIMENTAL AND INHERENT UNCERTAINTIES IN THE INFORMATION THEORETIC APPROACH *

Y. ALHASSID

Kellogg Radiation Laboratory, California Institute of Technology, Pasadena, California 91125, USA

and

R.D. LEVINE **

Department of Chemistry, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

Received 25 February 1980; in final form 14 April 1980

A bound on the uncertainty of the Lagrange multiplier due to experimental scatter in the input frequencies is provided. An inequality relating experimental and inherent uncertainties is derived. The product of the inherent uncertainties of a constraint and its conjugate variable is shown to have a minimal value of unity.

1. Introduction

The experimental determination of the frequencies of different outcomes is necessarily subjected to some scatter. In the maximum entropy formalism [1,2] one attempts to fit the measured frequencies by a theoretical distribution computed by the maximum entropy (subject to constraints) procedure. The quality of the fit can only be improved by including additional constraints. There may come a point however where the addition of further constraints serves only to fit the noise. Given the magnitude of the relative error in the individual frequencies, it was previously shown possible to identify that point [2]. Loosely speaking one can say that constraints need be added until the average squared fractional deviance between the experimental frequencies and the theoretical probabilities is below the average squared fractional error in the frequencies. This latter quantity will play an essential role in the present discussion and will be denoted by s^2,

$$s^2 = \sum_i f_i(\delta f_i/f_i)^2 . \tag{1}$$

Here f_i is the frequency of the ith outcome and $\delta f_i/f_i \equiv \delta \ln f_i$ is the fractional error.

Even when the number of constraints has been determined there remains the problem of what range of mean values of the constraints (or, equivalently, what range of values for the Lagrange multipliers) is consistent with the given noise level in the observed frequencies. In other words, the practical question is how to assign theoretically well-defined error bars to the Lagrange parameters; error bars which reflect the experimental errors in the frequencies.

In addition to the experimental uncertainty in the mean value of a constraint there is also an inherent uncertainty, measured by its variance. This inherent uncertainty reflects the inevitable fluctuations that can occur in any finite number of observations. We shall show that the experimental and inherent uncertainties are related via the inequalities (19). Moreover we shall show that the inherent uncertainty in a con-

* Work supported by the Air Force Office of Scientific Research, Grant AFOSR 77-3135, and the US–Israel Binational Science Foundation (BSF).

** Permanent address: Department of Physical Chemistry, The Hebrew University, Jerusalem, Israel.

straint and in its conjugate variable satisfies an uncertainty relation. The minimal value of the uncertainty product occurs when the distribution is of maximal entropy subject to that constraint.

The mathematical development in sections 3 and 4 will assume, for simplicity, that a single constraint (in addition to normalization) was found sufficient to account for the observed frequencies to within their error bars. All the results can however be extended to the case of several constraints.

Applications of the practical bound for the uncertainty of the Lagrange parameter, eq. (19b), are reported in two papers in press [3,4].

2. The fundamental inequality

The fundamental inequality relates the inherent uncertainty (the variance) and the uncertainty due to the variation of the frequencies (which is due to experimental scatter or for any other reason).

Let $\{f_i\}$ be the experimental distribution and A an observable whose value in the state i is $A(i)$. The variance of A, $\Delta^2(A)$, is defined as usual, by

$$\Delta^2(A) = \sum_i f_i [A(i) - \langle A \rangle]^2, \tag{2}$$

where $\langle A \rangle$ is the average value

$$\langle A \rangle = \sum_i A(i) f_i. \tag{3}$$

Let δf_i be the uncertainty in f_i, such that normalization is conserved, $\Sigma \delta f_i = 0$. Changing the f_i by δf_i changes $\langle A \rangle$ by $\delta \langle A \rangle$,

$$\delta \langle A \rangle = \sum_i A(i) \delta f_i. \tag{4}$$

Since typically we only know the absolute magnitude but not the sign of δf_i, (4) does not, in general, offer a practical way to estimate $\delta \langle A \rangle$.

Using the Cauchy–Schwarz inequality:

$$\left| \sum_i a_i b_i \right| \leqslant \left(\sum_i a_i^2 \right)^{1/2} \left(\sum_i b_i^2 \right)^{1/2}, \tag{5}$$

with equality if and only if a_i is proportional to b_i, it follows that

$$|\delta \langle A \rangle| = \left| \sum_i A(i) \delta f_i \right| = \left| \sum_i [A(i) - \langle A \rangle] \delta f_i \right|$$

$$= \left| \sum_i [A(i) - \langle A \rangle] f_i \delta \ln f_i \right|$$

$$\leqslant \left[\sum_i f_i [A(i) - \langle A \rangle]^2 \right]^{1/2} \left[\sum_i f_i (\delta \ln f_i)^2 \right]^{1/2}$$

$$= \Delta(A) s, \tag{6}$$

with equality if and only if $\delta \ln f_i = \delta f_i / f_i$ is proportional to $A(i) - \langle A \rangle$,

$$-\delta f_i / f_i = \delta \lambda [A(i) - \langle A \rangle], \tag{7}$$

$$-\delta \ln (f_i / f_i^0) = \delta \lambda [A(i) - \langle A \rangle]. \tag{8}$$

In (7) and (8), $\delta\lambda$ is used to denote the coefficient of proportionality and f_i^0 is any constant distribution.

3. The practical inequalities

The inequality (6) offers a practical bound for the error, $\delta \langle A \rangle$, in the mean value of A. The input is the inherent uncertainty, $\Delta(A)$, of A [cf. (2)] and s which can be computed [cf. (1)] using only the squared fractional error, $(\delta f_i / f_i)^2$, in the frequencies. The bound is tightest when f_i is a distribution of maximal entropy subject to $\langle A \rangle$ as a constraint [cf. (12) below].

The inequality (6) is valid for any observable A and any set of frequencies $\{f_i\}$. Say however that it proves possible to fit f_i to within the error bars by a theoretical distribution p_i which is of maximal entropy subject to A as the constraint,

$$p_i = p_i^0 \exp[-\lambda A(i) - \lambda_0]. \tag{9}$$

Here λ_0 is that function of λ which ensures that $\{p_i\}$ is normalized

$$\exp(\lambda_0) = \sum_i p_i^0 \exp[-\lambda A(i)], \tag{10}$$

and we proceed to write down three known identities which follow from (10):

$$\langle A \rangle_p \equiv \sum_i A(i) p_i = -\partial \lambda_0 / \partial \lambda, \tag{11}$$

$$\partial \ln p_i / \partial \lambda = -[A(i) - \langle A \rangle_p], \tag{12}$$

and, denoting the variance of A for the distribution $\{p_i\}$ by $\Delta_p^2(A)$,

$$\partial\langle A\rangle_p/\partial\lambda = \sum_i A(i)p_i\,\partial\ln p_i/\partial\lambda = \Delta_p^2(A), \tag{13}$$

where we used (12).

A rough and ready derivation of the second practical inequality can now be given as follows. Let $\{p_i\}$ be the theoretical fit for $\{f_i\}$. Then, from (13), the error bar, $\delta\lambda$, on the Lagrange parameter is estimated from

$$\delta\lambda \approx (\partial\lambda/\partial\langle A\rangle_p)\delta\langle A\rangle_p = \delta\langle A\rangle_p/\Delta_p^2(A). \tag{14}$$

Now, by construction, $\langle A\rangle_p = \langle A\rangle_f$ and, when f is changed to $f' \equiv f + \delta f$ it can still be fitted by a distribution of maximal entropy, $p' \equiv p + \delta p$ such that * $\langle A\rangle_{p'} = \langle A\rangle_{f'}$. Hence, $\delta\langle A\rangle_p$ is just $\delta\langle A\rangle$ as defined in (4). Using the first inequality, (6), in (14)

$$\delta\lambda \leqslant s\Delta_f(A)/\Delta_p^2(A) \approx s/\Delta_p(A). \tag{15}$$

In section 4 we shall argue that a variance can be defined for λ and that an alternative form of the inequality (15) is

$$\delta\lambda \leqslant s\Delta(\lambda), \tag{16}$$

where [using (22) in (15)]

$$\Delta^2(\lambda) \equiv \Delta_p^2(\partial\ln p/\partial\langle A\rangle) = \sum_i p_i(\partial\ln p_i/\partial\langle A\rangle)^2. \tag{17}$$

In (17), the set of values $\{\partial\ln p_i/\partial\langle A\rangle\}$ is regarded as an observable whose variance can then be computed. Noting that $\Sigma\partial p_i/\partial\lambda = 0$ it also follows from (12) and (13) that

$$\Delta_p^2(A) = \sum_i p_i(\partial\ln p_i/\partial\lambda)^2 = \Delta_p^2(\partial\ln p/\partial\lambda). \tag{18}$$

When the distribution of maximal entropy fits the frequency data to within the error bars we infer [2], that but for the inevitable scatter, the "true" distribution is one of maximal entropy. Hence, both bounds are saturated and their final practical form is

$$\delta\langle A\rangle \leqslant s\Delta(A), \tag{19a}$$

$$\delta\lambda \leqslant s/\Delta(A), \tag{19b}$$

or in a symmetric fashion

$$\delta\lambda\delta\langle A\rangle \leqslant s^2. \tag{19c}$$

Eqs. (19) are the practical working expressions. $\Delta(A)$ is computed by (2), s by (1) and $\delta\lambda$ and $\delta\langle A\rangle$ are to be interpreted as absolute values.

The central practical point is that the bounds (19) can be computed in terms of the squared fractional error in the frequencies and are valid irrespective of magnitude of the error. To stress this point consider the possibility that while $\{f_i\}$ is not the prior distribution, the prior distribution, $\{p_i^0\}$ does lie within the error range $f_i + \delta f_i$. We show explicitly in the appendix that one would then conclude that $|\delta\lambda| > |\lambda|$, or in words that $\lambda = 0$ (which is the prior distribution) does lie within the acceptable range for λ.

4. The uncertainty relation

But for the direction of the inequality, (19c) looks like an uncertainty relation between two conjugate variables. In this section we indeed show that the inherent uncertainties do satisfy an uncertainty relation.

In order to avoid any potential misunderstanding it should be stressed however that there is nothing quantum mechanical about our result. While our proof uses an observable A with a discrete set of values it can easily be modified for a continuous distribution. All that we do, is offer one more reason why, in the maximal entropy formalism one can speak of "conjugate" variables [5].

Let $\{q_i\}$ be some distribution such that the values q_i may depend on the mean value $\langle A\rangle_q$ of A. Now,

$$1 = \partial\langle A\rangle_q/\partial\langle A\rangle_q = \sum_i A(i)\partial q_i/\partial\langle A\rangle_q. \tag{20}$$

Using again the Cauchy–Schwarz inequality, and repeating the steps that led to (6), we obtain **

$$\Delta_q(A)\Delta_q(\partial\ln q/\partial\langle A\rangle) \geqslant 1. \tag{21}$$

Equality obtains in (21) when $\{q_i\}$ is that distribution which is of maximal entropy subject to $\langle A\rangle_q$ as the constraint, or

* It is important to note that it is here that we are using the assumption that the distribution of maximal entropy subject to A fits f to within its error bars.

** When the inequality (21) is written as $\Delta_q^2(A) \leqslant 1/\Delta_q^2 \times (\partial\ln q/\partial\langle A\rangle)$ it is seen to be a special case of the "Cramer–Rao" inequality [6], which is used in statistics to estimate the variance of A.

$$\Delta_p(A)\Delta_p(\partial \ln p/\partial\langle A\rangle) = 1. \quad (22a)$$

Using (18) we can also write

$$\Delta_p(\partial \ln p/\partial\lambda)\Delta_p(\partial \ln p/\partial\langle A\rangle) = 1. \quad (22b)$$

Finally, although we shall not prove it here, one can define $\Delta^2(\lambda)$, the variance or inherent uncertainty of λ such that $\Delta(\lambda) = \Delta_p(\partial \ln p/\partial\langle A\rangle)$ or

$$\Delta(A)\Delta(\lambda) = 1. \quad (22c)$$

Multiplying both sides of (22c) by s^2 we recover the practical result (19c). Indeed, the derivation of the practical inequality can also be rephrased as follows: If $\{f_i\}$ and its uncertainty can be well fitted by $\{p_i\}$ and its variation we infer that the experimental scatter is due primarily to the uncertainty in the value of $\langle A\rangle$. Hence

$$s = \Delta(\partial \ln p/\partial\langle A\rangle)\delta\langle A\rangle. \quad (23)$$

On multiplying both sides of (22a) by $\delta\langle A\rangle$ and using (23), the inherent uncertainty relation (22a) implies the practical bound (19a).

In the same sense that the Lagrange parameter is conjugate to the value, $\langle A\rangle$, of the constraint, the uncertainty relation shows that the observable $\partial \ln p/\partial\langle A\rangle$ is conjugate to the observable A. This very much conforms to the role of the observable $\partial \ln p/\partial\langle A\rangle$ in the maximum entropy formalism [7,8].

5. Summary

The effect of inevitable experimental scatter in the measured frequencies on the values of the constraint and of the Lagrange parameter was discussed. The resulting uncertainty in either of these values could be bounded by a product of s, which is a measure of the experimental noise, times the inherent uncertainty, e.g. eq. (6). The bound is saturated when the distribution is one of maximal entropy. It was also shown that irrespective of the extent of noise in the experimental data, the product of the inherent uncertainties in the values of a constraint and of its conjugate variable exceeds or equals unity and is equal to unity if and only if the distribution is of maximal entropy subject to that constraint.

Appendix. An exact expression for the experimental uncertainty product

Let $\{p_i\}$ and $\{p_i'\}$ be two distributions of maximal entropy subject to A as a constraint. The value of the constraint is $\langle A\rangle_p$ and $\langle A\rangle_{p'}$, respectively and, in this appendix, $\delta\langle A\rangle \equiv \langle A\rangle_{p'} - \langle A\rangle_p$. It follows immediately from (9) that

$$\delta\langle A\rangle\delta\lambda = \sum_i (p_i' - p_i)\ln(p_i'/p_i), \quad (A1)$$

where $\delta\lambda = \lambda - \lambda'$. Note also that (A1) is non-negative and vanishes if and only if $p_i = p_i'$.

Using the inequality $\ln(1+x) < x$ (valid for $x > -1$, $x \neq 0$), it follows

$$\sum_i (p_i' - p_i)\ln(p_i'/p_i) < \sum_i (p_i' - p_i)[(p_i' - p_i)/p_i] = s^2, \quad (A2)$$

where s is defined as in (1). This offers an alternative proof of (19c).

Say now p_i' is the prior distribution. Then $\lambda' = 0$ and from (A1) and (A2) $\lambda\delta\langle A\rangle \leqslant s^2$ but $\delta\lambda$ is computed as $\delta\lambda = s^2/\delta\langle A\rangle$. Hence $\delta\lambda \geqslant \lambda$, Q.E.D.

If, due to over enthusiasm, we disregard the performance criterion in ref. [2] and do impose a constraint even when it is not warranted (due to the large error bars on the frequencies), we shall find that uncertainty in the Lagrange parameter exceeds its absolute value. See refs. [3,4] for explicit examples.

The inequality (A2) also implies that $s^2 \geqslant \delta\langle A\rangle\delta\lambda$ where equality obtains only when both sides tend to zero. It follows that the product of the inherent uncertainties retains its value even when the experiment is made precise to any desired degree (i.e. as $s \to 0$). In other words, writing (19a) and (19b) as

$$\Delta(A)\Delta(\lambda) \geqslant \delta\langle A\rangle\delta\lambda/s^2, \quad (A3)$$

the left-hand side has a finite limit (unity) as $s \to 0$.

References

[1] R.D. Levine and M. Tribus, eds., The maximum entropy formalism (M.I.T. Press, Cambridge, 1979).

[2] J.L. Kinsey and R.D. Levine, Chem. Phys. Letters 65 (1979) 413.
[3] E. Zamir and R.D. Levine, Energy Disposal in Unimolecular Elimination Reactions, to be published.
[4] E. Zamir, R.D. Levine and R.B. Bernstein, The Practice of Surprisal Inference: Products' State Distribution in the Chemiluminescent Ion Molecule Reaction of $C^+ + H_2$, to be published.
[5] R.D. Levine, J. Chem. Phys. 65 (1976) 3302.
[6] H. Cramer, Mathematical methods of statistics (Wiley, New York, 1946).
[7] Y. Alhassid and R.D. Levine, Phys. Rev. C20 (1979) 1775.
[8] R.D. Levine, J. Phys. A13 (1980) 91.