JOURNAL OF COMPUTATIONAL PHYSICS 23, 446–451 (1977)

Note

The Truncated Kullback Statistic as an Objective Function for Fitting Partially Known Probability Vectors

The Kullback statistic [1, 2], I_k, is a positive definite [3, 4] measure of the difference between two normalized probability vectors $P^{(1)}$ and $P^{(2)}$,

$$I_k(P^{(1)}, P^{(2)}) = \sum_{i=1}^{n} p_i^{(1)} \log_2 (p_i^{(1)}/p_i^{(2)}), \tag{1}$$

where

$$P^{(1)} = (p_1^{(1)},..., p_n^{(1)})^T$$

and

$$P^{(2)} = (p_1^{(2)},..., p_n^{(2)})^T.$$

It may be defined as the negative of the expectation value of the surprisals U_i, $U_i = -\log_2(p_i^{(1)}/p_i^{(2)})$. The Kullback statistic has attracted much interest in physics and chemistry [5–9]. The Kullback statistic has already been used in the context of interest here, which is to fit probability vectors [10].

A contemporary problem in chemical physics is the determination of the probabilities of the quantum states, 0 to n, of the products of a chemical reaction [11]. The range, 0 to n, of quantum states is determined by the energy available to the reaction (which is readily calculated [12]). Experimentally it may be possible to detect only a subset of the final states and to measure only the relative probabilities of, e.g., the states s to n ($s > 0$). We may then calculate the experimental values of the ratios of the probabilities of states $s + 1$ to n to the probability of state s; these are the experimental probability ratios q_j, $j = s + 1$ to n. (If, on account of physical reasons, the probability ratios of adjacent states s to n are determined experimentally, the ratios $s + 1$ to n can be calculated only at the expense of error propagation.)

Theoretically it is possible in principle to calculate the complete probability vector $(p_0^{(2)}, p_1^{(2)},..., p_s^{(2)}, p_{s+1}^{(2)},..., p_n^{(2)})^T$ and, from this, the theoretical values for the probability ratios, $p_{s+1}^{(2)}/p_s^{(2)},..., p_n^{(2)}/p_s^{(2)}$, which may be compared with the experimental values, q_j. If the theoretical method involves unknown variables (related to, say, the potential between separating products) we must find the "best" values of these variables. The "best" values of the variables are those values which minimize the distance between the calculated and experimental probability vectors or minimize the full Kullback statistic (1) [1–4, 10]. However, in the general case, we do not know the experimental probability vector. Only the probability ratios $q_{s+1},..., q_n$ are known.

Copyright © 1977 by Academic Press, Inc.
All rights of reproduction in any form reserved.

ISSN 0021-9991

One method of determining optimum values of the variables is to minimize the sum of squares of differences, $\sum R_k{}^2$, between observed and calculated probability ratios,

$$\sum R_k{}^2 = \sum_{k=s+1}^{n} (q_k - q_k^{(\mathrm{calcd})})^2,$$

where

$$q_k^{(\mathrm{calcd})} = p_k^{(2)}/p_s^{(2)}, \qquad k = s + 1 \text{ to } n.$$

The optimum probability vector derived by minimizing this objective function will be denoted by $P^{(2)}(\sum R_k{}^2)$.

If we can express the relative probability ratios as ratios to a common denominator (as described above), an alternative objective function, the truncated Kullback statistic, may be used. We can incorporate the known probability ratios, q_j, in a normalized probability vector using A defined in

$$\sum_{i=0}^{s} p_i^{(2)} + A \sum_{j=s+1}^{n} q_j = 1.$$

The truncated Kullback statistic, I_k', is simply

$$I_k' = \sum_{j=s+1}^{n} p_j^{(1)} \log_2 (p_j^{(1)}/p_j^{(2)}), \tag{2}$$

where

$$p_j^{(1)} = Aq_j, \qquad j = s + 1 \text{ to } n.$$

The optimum probability vector derived by minimizing I_k' will be denoted by $P^{(2)}(I_k')$.

Generally speaking, the Kullback statistic (1) assigns small weights to surprisals of low probabilities. The truncated Kullback statistic (2) omits from the summation (1) unknown probabilities, thereby restricting the measure of uncertainty to known (relative) probabilities. When the unknown probabilities are small, we have found that the truncated Kullback statistic may be superior to a sum of squares of residuals as an objective function. That is, according to the criteria outlined above, the optimum probability vector, $P^{(2)}(I_k')$, derived by minimizing I_k' with respect to the variables in a theoretical model, may be closer (as measured by distance or the full Kullback statistic (1)) to the experimental or true probability vector than $P^{(2)}(\sum R_k{}^2)$.

The simple bootstrap model of M. J. Berry has been used to calculate the probability vector of the vibrational states of a diatomic product of a chemical reaction [13–15]. There are two variables to be optimized and one parameter in this model. The variables are the internuclear separation of the diatomic product at the saddle point of the potential surface, $r_e{}^{\ddagger}$, and the coupling constant, α, between the diatomic product and the continuum of resonant energy levels of the polyatomic fragment [15]. The parameter, τ, is the time for the fragments to separate measured in half-lives of the decay of the highest significantly populated vibrational level of the diatomic [15]. Previous grid searches have demonstrated that this model is unimodal in $r_e^{\ddagger}$ and α [15, 16]. Accordingly, for low-accuracy optimization, we used Golden Sections [17] to find the values of $r_e^{\ddagger}$ and α which minimize I_k' or $\sum R_k^2$.

Three examples are presented to discuss the utility of the truncated Kullback statistic. The relative probabilities (given in Table II under "Ratios") for obtaining the diatomic product in its vibrational levels are all expressed relative to the probability of the lowest vibrational level ($v' = 0$). In [13, Example 1] and [16, Example 3] the product is hydrogen fluoride formed from the reaction of fluorine atoms with molecular hydrogen and oxygen atoms with vinyl fluoride, respectively. In [15, Example 2] the product is hydrogen chloride formed by flash photolysis of vinyl chloride.

TABLE I

Optimized Parameters

	Example 1		Example 2		Example 3	
Objective function	I_k'	ΣR_k^2	I_k'	ΣR_k^2	I_k'	ΣR_k^2
r[a]	1.55	1.38	1.88	1.90	1.47	1.16
α[b]	—	—	−3.9	−3.8	−4.3	−4.8
τ[c]	—	—	1.6	1.6	4.6	3.2

[a] The bond length of the diatomic at the transition state [15]. The Golden Section is less than 0.05 Å.

[b] The Briggs logarithm of the intercontinuum coupling constant in cm [15]. The Golden Section is less than 0.5.

[c] The time for relaxation [15]. The step size of the initial value problem is 0.2.

It is clear from the results displayed in Table I that using the truncated Kullback statistic as an objective function can result in a minimum different from that obtained by using a sum of squares of residuals for known probability ratios. The optima obtained using I_k' for Examples 1 and 2 are close to those chosen by Berry [13, 15].

In Examples 1 and 3, only one probability is missing, so the experimental ratios may be normalized to give the experimental probability vectors $P^{(\mathrm{exp})}$. $P^{(\mathrm{exp})}$ may then be compared with $P^{(2)}(I_k')$ and $P^{(2)}(\sum R_k^2)$ by calculating the distance $| P^{(\mathrm{exp})} - P^{(2)} |$ and the full Kullback statistic (1). These results are recorded in Table II.

TABLE II

Optimum Probability Vectors

	Example 1				Example 2				Example 3				
	Experimental		Calculated		Experimental		Calculated			Experimental		Calculated	
v'	Ratios	$P^{(\mathrm{exp})}$	$P^{(2)}(I_k')$	$P^{(2)}(\Sigma R_k^2)$	v'	Ratios	$P^{(2)}(I_k')$	$P^{(2)}(\Sigma R_k^2)$	v'	Ratios	$P^{(\mathrm{exp})}$	$P^{(2)}(I_k')$	$P^{(2)}(\Sigma R_k^2)$
0		0.029	0.008	0.029	0		0.250	0.241	0		0.711	0.794	0.717
1	5.20	0.149	0.121	0.234	1	0.86	0.207	0.203	1	0.32	0.227	0.161	0.230
2	17.7	0.505	0.586	0.587	2	0.77	0.185	0.185	2	0.069	0.049	0.037	0.048
3	11.1	0.318	0.286	0.150	3	0.70	0.167	0.170	3	0.018	0.013	0.008	0.005
					4	0.55	0.129	0.134	4		0	0.001	0.000
					5		0.060	0.065					
					6		0.002	0.002					
I_k'			1.6×10^{-2}	1.5×10^{-1}			3.4×10^{-5}	3.1×10^{-4}				2.9×10^{-4}	8.0×10^{-3}
ΣR_k^2			4.3×10^{3}	5.2×10^{1}			4.7×10^{-3}	4.3×10^{-4}				1.4×10^{-2}	1.4×10^{-4}
$\lvert P^{(\mathrm{exp})} - P^{(2)} \rvert$			9.4×10^{-2}	2.1×10^{-1}								1.1×10^{-1}	1.0×10^{-2}
$I_k(P^{(\mathrm{exp})}, P^{(2)})$			3.9×10^{-2}	1.4×10^{-1}								2.8×10^{-2}	6.5×10^{-3}

In [13, Example 1], the distance between $P^{(\mathrm{exp})}$ and the optimum vector obtained using the truncated Kullback statistic, $P^{(2)}(I_k')$, is smaller than the distance between $P^{(\mathrm{exp})}$ and the optimum vector, $P^{(2)}(\sum R_k^2)$, obtained using the sum of squares of residuals as the objective function (Table II). Furthermore, the full Kullback statistic (1) for $P^{(2)}(I_k')$ compared to $P^{(\mathrm{exp})}$ is smaller than for $P^{(2)}(\sum R_k^2)$. In the particular case of Example 1, the distance between experimental and calculated vectors itself could have been used as the objective function, but, in the more typical case [11], such as [15, Example 2], only a partial sum can be performed. Subject to this restriction the truncated Kullback statistic gives a better optimum vector for Example 1. In [16, Example 3] the truncated Kullback statistic produces an optimum vector inferior to that found using a sum of squares of residuals. In Example 3, two of the three known probabilities are small so the natural weighting of the Kullback statistic is distorted. In Example 2, the minima derived from the two objective functions cannot be distinguished to within the precision of optimization.

Using the truncated Kullback statistic as the objective function may give ratios of probabilities in poorer agreement with experiment than those obtained by using a sum of squares of residuals. The theoretically significant quantity is the true probability vector, however. If the experimentally unknown probabilities are not large it is clear that using the truncated Kullback statistic as an objective function can lead to a better calculated probability vector and, therefore, truer values of the unknown parameters in the theory.

ACKNOWLEDGMENTS

We thank the Dalhousie Computer Centre for facilitating these calculations. We acknowledge support from the Defence Research Board through Grant 9530-115 to W. E. Jones.

REFERENCES

1. S. KULLBACK, *Ann. Math. Statist.* **22** (1951), 79.
2. S. KULLBACK, "Information Theory and Statistics," Wiley, New York, 1959.
3. L. BRILLOUIN, "Science and Information Theory," Academic Press, New York, 1962.
4. A. J. KHINCHIN, "Mathematical Foundations of Information Theory," Dover, New York, 1957.
5. F. SCHOGL, *Z. Physik* **249** (1971), 1.
6. A. HOBSON AND B. K. CHENG, *J. Statist. Phys.* **7** (1973), 301.
7. M. TRIBUS AND R. ROSSI, *J. Statist. Phys.* **9** (1973), 331.
8. R. B. BERNSTEIN AND R. D. LEVINE, *Adv. At. Mol. Phys.* **11** (1975), 215.
9. U. DINUR AND R. D. LEVINE, *Chem. Phys.* **9** (1975), 17.
10. U. DINUR AND R. D. LEVINE, *Chem. Phys. Lett.* **31** (1975), 412.
11. T. CARRINGTON AND J. C. POLANYI, Chemiluminescent Reactions, *in* "MTP International Review of Science: Chemical Kinetics" (J. C. Polanyi, Ed.), Physical Chemistry, Series 1, Vol. 9, Chap. 5, Butterworth, London, 1972.
12. D. H. MAYLOTTE, J. C. POLANYI, AND K. B. WOODALL, *J. Chem. Phys.* **57** (1972), 1547.
13. M. J. BERRY, *Chem. Phys. Lett.* **27** (1974), 73.

14. M. J. Berry, *Chem. Phys. Lett.* **29** (1974), 329.
15. M. J. Berry, *J. Chem. Phys.* **61** (1974), 3114.
16. W. E. Jones, G. Matinopoulos, J. S. Wasson, and M. Liu, in preparation.
17. M. J. Box, D. Davies, and W. H. Swann, "Non-Linear Optimization Techniques," Oliver and Boyd, Edinburgh, 1969.

Received: May 24, 1976; revised: September 13, 1976

W. E. Jones, G. Matinopoulos, and J. S. Wasson

Department of Chemistry
Dalhousie University
Halifax, Nova Scotia, Canada B3H 4J3