match no. 18 pp. 261-271 1985

COMPLETE ENTROPIC MEASURES

J.Brickmann[+] and R. D. Levine
The Fritz Haber Molecular Dynamics Research Center
The Hebrew University, Jerusalem 91904, Israel

(received: February 1985)

The familiar (or 'Shannon') entropy is not sufficient for comparing two probability distributions. A more general notion is the entropy of a distribution with respect to a reference (a 'measure'). Given two distributions, a set of references (or measures) which is sufficient is defined. Using this set one can compare the two distributions, that is establish a partial order relation. The set of measures is necessary and sufficient for deciding whether one distribution is more mixed than the other or that the two are incomparable.

INTRODUCTION

The need to compare two distributions arises in many branches of physics. An obvious example is nonequilibrium phenomena where one may wish to know which distribution is further away from equilibrium. There are however other situations. Take, for example, the distribution of intensities in the absorption spectrum of a molecule[1,2]. The example is instructive since on physical grounds it was argued[1,2] that the measures (in the technical sense discussed below) to be used depend on the distribution itself.

[+]Permanent address: Institute for Physical Chemistry,
Technical University Darmstadt,
Petersenstr. 20, D-6100 Darmstadt,
W. Germany

That a partial ordering of distributions can be introduced is shown by the concept of mixing character[3,4]. Given two discrete normalized probability distributions $p = (p_1,\dots,p_n)$ and $p' = (p'_1,\dots,p'_n)$ we assume them to be arranged in descending order ($p_{i+1} \leq p_i$). Then p' is said to be more mixed than p, in symbols

$$m[p'] \succ m[p] \tag{1}$$

if and only if all the partial sums satisfy

$$\sum_{i=1}^{r} p'_i \leq \sum_{i=1}^{r} p_i \tag{2}$$

for all r between 1 and n. The analogous definition for continuous distributions is given in section 2 and for quantum mechanical states we refer the reader to the literature[5,6].

The concept of increasing mixing character is meant to convey increasing disorder with the uniform distribution being most mixed [3-6]. There is however no obvious relation between the condition (2) and the familiar notion of entropy. To be sure, (2) implies

$$S[p'] \geq S[p] \tag{3}$$

where S is the usual entropy,

$$S[p] = -\sum_{i=1}^{n} p_i \ln p_i \tag{4}$$

but the converse does not hold. It has often been conjectured that (2) is equivalent to (3) if (3) holds for some 'complete' set of entropies. But at least for the Renyi entropies [7] that apparently is not the case [8].

The entropy (4) is however a special case in that it is the entropy relative to a uniform measure. In general, the entropy of a distribution p relative to some normalized reference distribution q is given by[7,9-12] the non-negative quantity

$$DS[p \mid q] = -\sum_{i=1}^{m} p_i \ln(p_i/q_i) \tag{5}$$

The reference distribution q can be understood as a measure [13]. An even more general interpretation is provided by the 'grouping property' [14] of the entropy (4). Consider the usual case where the index i does not refer to an 'elementary' event but is a grouping together of several such events. (By elementary events we mean those which are equiprobable at the most statistical distribution). Then each elementary event carries two indices (say i and j), the first identifying the group (i) and the second index (j) identifies the state within the group. The probability of any state i, j can then be written in the obvious notation $p_{ij} = p_i p(j \mid i)$. Then

$$\begin{aligned} S[p] &= -\sum_{ij} p_{ij} \ln p_{ij} \\ &= -\sum_{i}\Big(p_i \ln p_i + p_i(-\sum_{j} p(j \mid i) \ln p(j \mid i)\Big) \\ &= -\sum_{i} p_i \ln(p_i/g_i) \end{aligned} \tag{6}$$

where the non-negative measure g_i,

$$\ln g_i = -\sum_{j} p(j \mid i) \ln p(j \mid i), \tag{7}$$

is the 'effective degeneracy' of the group of states i (g_i is bounded from above by the number of states in the group). Introducing the total effective number of states, $N = \sum_i g_i$ and normalising $q_i = g_i/N$

$$S[p] = \ln N - DS[p \mid q]. \tag{8}$$

For a given distribution q, lnN is the maximal value of $S[p]$. Hence $DS[p \mid q]$ measures how far the entropy is below its maximal value. For this reason it is sometimes known as 'the entropy deficiency'. Note that the measure DS is a special representation of a mixing distance [15].

The question posed in this paper is: given two distributions p and p' arranged in descending order, can one determine a set of measures $q^{(r)}$ such that

$$m[p'] \succ m[p] \Leftrightarrow DS[p|q^{(r)}] \geqslant DS[p'|q^{(r)}], \quad \text{all } r. \qquad (9)$$

THEORY

We consider continuous distributions on the interval $[0,1]$. The extension to other intervals can be performed by a suitable transformation. The discrete distributions are then a special case (step functions). The partial order amongst such continuous distributions is

$$m[p'] \succ m[p] \Leftrightarrow \int_0^1 g(p'(\tau))d\tau \geqslant \int_0^1 g(p(\tau))d\tau \qquad (11a)$$

for all convex functions g. It was shown by Ruch and Mead [16] that it is sufficient to consider a subclass, M^+, of all convex functions

$$M^+ = \left\{ g_\lambda(x) = -(x-\lambda)^+ \equiv -\max(0,x-\lambda) \,|\, 0 \leq \lambda \right\}.$$

The equivalence (11a) can then be replaced by

$$m[p'] \succ m[p] \Leftrightarrow \int_0^1 [p(\tau)-\lambda]^+ d\tau \geqslant \int_0^1 [p'(\tau)-\lambda]^+ d\tau ,$$

$$\text{for all } \lambda \geqslant 0. \qquad (11b)$$

Associated with any given probability density $p \in L^1([0,1])$ there is a density P which is monotonically decreasing and satisfies the relation

$$p(\tau) = P(\phi(\tau)) \qquad (10)$$

where $\phi(\tau)$ is a measure preserving transformation on the interval $[0,1]$, [16].

Using such a transformation we can replace p and p' by their monotonically decreasing arrangements P and P' respectively. In the discrete case this transformation is of course just a permutation of the indices. One can now state (11c), a definition equivalent to (11a) and (11b)

$$m[p'] \succ m[p] \Leftrightarrow \int_0^\lambda P(\tau)d\tau \geqslant \int_0^\lambda P'(\tau)d\tau, \text{ all } \lambda \text{ in } [0,1]. \qquad (11c)$$

Two distributions p and p' for which the inequalities in (11b or c) are valid only for a particular set of values of λ are called 'non comparable', i.e. neither $m[p'] \succ m[p]$ nor $m[p] \succ m[p']$ holds.

With these preliminaries under way we come to our central result relating mixing character to the entropy. Theorem:

$$m[p'] \succ m[p] \Leftrightarrow \int_0^1 P\ln(P/Q_\lambda)d\tau \geqslant \int_0^1 P'\ln(P'/Q_\lambda)d\tau,$$

$$\text{all } \lambda \text{ in } [0,1]. \qquad (12)$$

Here $p \equiv P(\phi(\tau))$, $p' = P'(\phi'(\tau))$ with ϕ and ϕ' two measure preserving transformations on the interval $[0,1]$ such that $P(\tau)$ and $P'(\tau)$ are non increasing. Defining ΔP_λ and ΔS_λ by

$$\Delta P_\lambda = \int_0^\lambda (P-P')d\tau \qquad (13)$$

$$\Delta S_\lambda = \int_0^\lambda (P\ln P - P'\ln P')d\tau, \qquad (14)$$

the reference distributions Q_λ are given as

$$Q_\lambda = \begin{cases} 1 \text{ for } \Delta P_\lambda = 0 \text{ and otherwise by} \\ N_\lambda \exp[(\Delta S_\lambda/\Delta P_\lambda)-1], & \tau \leq \lambda \\ N_\lambda \exp[(\Delta S_\lambda - \Delta S_1)/\Delta P_\lambda], & \tau > \lambda \end{cases} \qquad (15)$$

with the normalization

$$N_\lambda = \{\lambda \exp[(\Delta S_\lambda/\Delta P_\lambda)-1] + (1-\lambda)\exp[(\Delta S_\lambda - \Delta S_1)/\Delta P_\lambda]\}^{-1}. \qquad (16)$$

Proof. Consider first where $\Delta P_\lambda \neq 0$ for $\lambda < 1$. We show that the r.h.s. of (12) is equivalent to the r.h.s. of (11c). From (12)

$$I_\lambda(P) \equiv \int_0^1 P\ln(P/Q_\lambda)d\tau > I_\lambda(P') \equiv \int_0^1 P'\ln(P'/Q_\lambda)d\tau. \tag{17}$$

Dividing the interval $[0,1]$ into $[0,\lambda]$ and $(\lambda,1]$ leads to

$$I_\lambda(P)-I_\lambda(P') = A(P)-A(P') + B(P)-B(P') > 0 \tag{18}$$

with

$$\begin{aligned} A(P) &= \int_0^\lambda P\ln P d\tau - \left[(\Delta S_\lambda/\Delta P_\lambda)-1\right] \int_0^\lambda P d\tau \\ B(P) &= \int_\lambda^1 P\ln P d\tau - \left[(\Delta S_\lambda - \Delta S_1)/\Delta P_\lambda\right] \int_\lambda^1 P d\tau. \end{aligned} \tag{19}$$

Noting the definitions (13),(14), we obtain from (18) via (19) and a number of simple steps, including

$$\int_0^\lambda (P-P')d\tau = - \int_\lambda^1 (P-P')d\tau, \tag{20}$$

the result

$$I_\lambda(P)-I_\lambda(P') = \int_0^\lambda (P-P')d\tau > 0. \tag{21}$$

The inequality in (12) is thus equivalent to (11c). For $\Delta P_\lambda = 0$ and for $\lambda = 1$, the r.h.s. of (12) reads

$$- \int_0^1 P'\ln P'd\tau > - \int_0^1 P\ln P d\tau \tag{22}$$

which is true if $m[p'] \succ m[p]$ since the entropy is convex. On the other hand

$$\int_0^1 (P-P')d\tau = 0 \tag{23}$$

holds in any case since P and P' are normalized. For $\Delta P_\lambda \equiv 0$ in some range of λ, say $\lambda \in [\lambda_1, \lambda_2]$, any step function

$$\Omega_\lambda = \begin{cases} \varrho_1 \text{ for } \tau \leq \lambda \in [\lambda_1, \lambda_2] \\ \varrho_2 \text{ for } \tau > \lambda \in [\lambda_1, \lambda_2] \end{cases} \quad (24)$$

with $\varrho_1 \lambda + \varrho_2(1 - \lambda) = 1$, fulfills the equality in (12). Using $Q_\lambda = \Omega_\lambda$

$$\int_0^1 \left[P\ln(P/\Omega_\lambda) - P'\ln(P'/\Omega_\lambda) \right] d\tau$$

$$= \Delta S_1 - \ln\varrho_1 \Delta P_\lambda + \ln\varrho_2 \Delta P_\lambda \quad (25)$$

$$= \Delta S_1$$

since we consider the case $\Delta P_\lambda \equiv 0$. Hence one can just as well choose $\varrho_1 = \varrho_2 = 1$ (a uniform distribution). Q.E.D.

EXAMPLES

A few examples may demonstrate the procedure of constructing reference distributions for entropic measures of the comparison of both. We take for simplicity and clarity the discrete distributions

$$p = \frac{1}{36}(16,8,4,4,2,2,0,0)$$

$$p' = \frac{1}{36}(8,8,8,8,1,1,1,1) \quad (26)$$

$$p'' = \frac{1}{36}(16,8,4,2,2,2,2,0)$$

The partial sums $\sum_{i=1}^{k} p_i$ with k = 1-8 are without the common normalization

for p: 16,24,28,32,34,36,36,36
for p': 8,16,24,32,33,34,35,36 (27)
for p'': 16,24,28,30,32,34,36,36

and obviously

$$m[p'] \succ m[p]$$
$$m[p''] \succ m[p] \quad (28)$$

while the two distributions p' and p" turn out to be incomparable.

Five reference distributions (non normalized) are necessary and sufficient as reference distributions to compare p and p'

$$q^{(1)} = (32/e,\ 8\sqrt{2})$$
$$q^{(2)} = (32/e,\ 8\sqrt{2})$$
$$q^{(3)} = (1,1)$$
$$q^{(4)} = (64/e,\ 8)$$
$$q^{(15)} = (1024/e,\ 1/4)\ .$$

In this notation $q_i^{(k)}$ takes the first value for $i \leq k$ and the second for $i > k$.

For p and p" only four reference distributions are sufficient and necessary to decide that $m[p''] \succ m[p]$ while seven distributions hav to be used to identify that p" and p' are incomparable.

The normalized distributions p and p' are shown in fig 1a where in the lower panel shown is the range the individual reference distributions $q^{(j)}$ can take. Fig. 1b shows the results for the distributions p and p". In the first case the distributions p and p' are drastically differing. The range the $q^{(i)}$ can take is also large. In the second case p and p" are approaching each other and the reference distributions are only slightly different from the uniform distribution.

DISCUSSION

We have shown that there exists a complete set of reference distribution Q_λ for two given distributions P and P' (both in non-increasing order) such that the increase of the entropy deficiency going from P to P' for all Q_λ implies an increase of the mixing character $m[P'] \succ m[P]$. Some examples were given and it turns out also from other numerical checks that the number

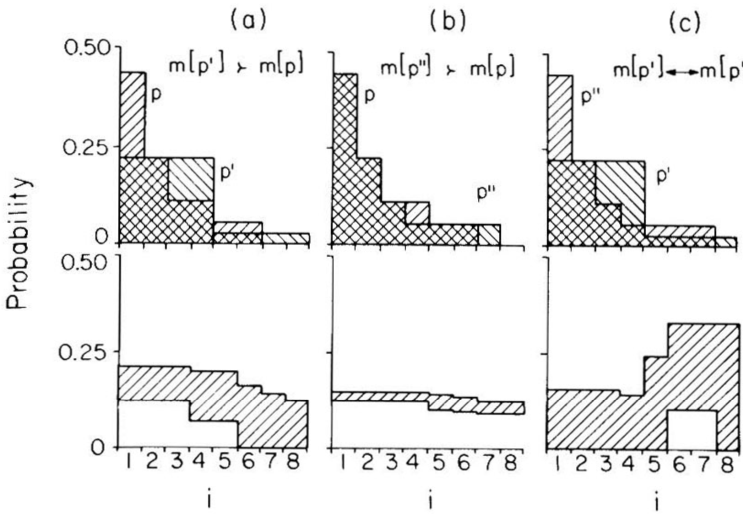

FIGURE 1

Three examples. For each the top panel is the two (discrete) distributions (each cross hatched in a different direction) and the bottom panel is the possible range of the reference distributions. (a) p' is more mixed than p. The possible range for the reference distributions is considerable but, upon actual construction it is found that they are all more mixed than either p or p'. (b) p'' is more mixed than p but p'' and p are more similar than p and p' in (a). The possible range of the reference distributions is now much narrower. In this case too all reference distributions are more mixed than either p or p'. (c) p' and p'' are not comparable. The range of the possible reference distributions is quite wide and some of them are not comparable with p' and/or p''.

of independent reference distributions, which are necessary to decide whether $m[P'] \succ m[P]$ or whether these distributions are non-comparable is decreasing when P and P' come closer to each other. A reduction of the range the Q_λ are covering is also reduced if both P and P' approach the uniform distribution. This is to be expected from the definition of the mixing character eq. (11b), where the range of λ -values which are necessary and sufficient to decide whether $m[P'] \succ m[P]$ or not is reduced to the range

$$\min(P,P') \leq \lambda \leq \max(P,P').$$

At least for discrete distributions this range has to converge to a small strip as p and p' tend towards the uniform distribution.

We are not able however to find an intuitive yet inclusive characterization of the (typically, correct) tendency of the reference distributions to become more mixed as P and P' become more mixed (cf.fig. 1a and b). It is unfortunately not true that Q_λ is necessarily not less mixed than P or P'. We shall continue to look for such a characterization.

Also not provided here is a physical interpretation of the set of complete measures Q_λ. Since the integrals I_λ that need be evaluated (cf.(12) and (17)) are entropy deficiencies (cf.(5)), $I_\lambda(P) \equiv DS[P|Q_\lambda]$ and since the entropy deficiency has a definite thermodynamic interpretation [18], such an interpretation is definitely called for.

Both these objectives while interesting in their own right are also essential preliminaries to the ultimate question: how to use the mixing character as a quantitative measure in describing systems in disequilibrium. So far, all the results can also be obtained from the maximum entropy formalism alone [4]. This is equivalent to using only a uniform measure. By introducing non uniform measures (the Q_λ's) we are opening the way for novel applications.

ACKNOWLEDGEMENT

We thank Prof. E. Ruch for discussion and Dr. Y. M. Engel for computational assistance. This work was supported by the Deutsche Forschungsgemeinschaft, Bonn and the Fonds der Chemischen Industrie, Frankfurt. The Fritz Haber Research Center for Molecular Dynamics and the stay of J.B. in Israel is supported by the Minerva Gesellschaft für die Forschung,mbH, München, BRD.

REFERENCES

1. E. J. Heller, J. Chem. Phys. 72, 1337 (1980).
2. J. Brickmann, P. Hofmann and P.C. Schmidt, Int. J. Quant. Chem. in press.
3. E. Ruch , Theor. Chim. Acta 38, 167 (1975).
4. A. Mead, Adv. Chem. Phys. 51, 113 (1982)
5. A. Wehrl, Rep. Math. Phys. 6, 15 (1974).
6. W. Thirring, Quantum Mechanics of Large Systems (Springer, New York, 1980), Sec. 2.1.
7. A. Renyi, Probability Theory (North Holland, Amsterdam, 1970).
8. W. Hasselbarth, J. Stat. Phys., submitted.
9. S. Kullback, Information Theory and Statistics (Wiley, New York, 1959).
10. A. Hobson, J. Stat. Phys. 1, 383 (1969).
11. R. D. Levine, Adv. Chem. Phys. 47, 239 (1981).
12. F. Schlögl, Z. Physk, B20, 177 (1975).
13. U. Dinur and R. D. Levine, Chem. Phys. 9, 17 (1975).
14. R. Ash, Information Theory (Wiley, New York, 1956).
15. E. Ruch, R. Schranner and T. H. Seligman, J. Chem. Phys. 69, 38 (1978).
16. E. Ruch and A. Mead, Theor. Chim. Acta 41, 95 (1976).
17. J. V. Ryff, Proc. Am. Math. Soc. 18, 1026 (1967).
18. I. Procaccia and R. D. Levine, J. Chem. Phys. 65, 3357 (1976).