Note di Matematica Vol.1, 43-52 (1981)

ENTROPIES ON INDEPENDENT OCCUPATION SPACES

Carlo SEMPI(°)

Sunto. *Si studiano alcune proprietà dell'entropia sopra gli spazi d'occupazione indipendenti mettendo in rilievo la relazione tra questa entropia e l'entropia di Shannon di ogni sito. L'ipotesi di indipendenza consente di rafforzare alcuni risultati di* [8] *e di* [9] *. In particolare, per l'esistenza del limite nel teorema 3 si sostituisce la richiesta che l'insieme tenda all'infinito nel senso di van Hove con quella più semplice che tenda all'infinito il numero di siti nell'insieme. Si determinano altresì due distribuzioni che danno massima entropia, secondo il principio di Jaynes.*

1. - INTRODUCTION.

Occupation spaces have been widely used (see, e.g., [9], [4]) in statistical mechanics to study systems as spin models, alloys, lattice gases,... . In all these cases a *n-dimensional occupation space* may be described as $\prod_{i=1}^{n} \Omega_i$ where $\Omega_i = \{0,1,\dots,r_i-1\}$ $(i=1,2,\dots,n)$, i.e. as the set

(°) Dipartimento di Matematica, Università di Lecce, 73100 Lecce.

of $\prod_{i=1}^{n} r_i$ finite sequences $(x_i, x_2, \ldots, x_n)$ with $x_i \in \Omega_i$. The case $r_i = 2$ $(i=1,2,\ldots,n)$ is the one most commonly considered; we shall therefore deal with it first. Each component can then assume the value 0 or 1. A physical meaning is attached to the occupation space in the following way: one has a set of n points, called sites, each of wich can either be occupied or not, or occupied by a particle of spin equal to $\frac{1}{2}$ or to $-\frac{1}{2}, \ldots$. The value of each component of a finite sequence is also given a meaning: say, in the spin model, $x_i = 1$ means that the spin of the particle in the i-th site is $\frac{1}{2}$, whilst $x_i = 0$ means that the same particle has spin equal to $-\frac{1}{2}$. We shall assume the components of every sequence to be independent, to wit we shall deal only with independent occupation spaces.

Let us consider a system of n sites and let p_i be the probability that $x_i = 1$, and hence $1 - p_i$ the probability that $x_i = 0$.

The probabilities on the $N = 2^n$ points of the n-dimensional occupation space X_n are given by

$$P_j = \prod_{i \in O} p_i \prod_{k \in U} (1 - p_k) \qquad (j=1,2,\ldots N)$$

where O and U are the set of occupied and unoccupied sites respectively (i.e., the sites for which $x_i = 1$ or $x_i = 0$, respectively). The Shannon's entropy associated with the system is

$$H_N(P_1, P_2, \ldots, P_N) = -\sum_{j=1}^{N} P_j \log P_j, \tag{1}$$

where, as in the following, the logarithms are in basis 2. As is easy to show by induction, (1) equals

$$\mathcal{H}_n(p_1,p_2,\dots,p_n) := -\sum_{i=1}^{n} \{p_i \log p_i + (1-p_i)\log(1-p_i)\}$$

(2)

$$= \sum_{i=1}^{n} H_2(p_i, 1-p_i),$$

that is the entropy of the system is the sum of the Shannon's entropies associated with each site. This result was to be expected in view of the assumption of independence and the properties of (1) (see [1]).

In the general case if p_{ij} is the probability that $x_i = j$ $(i=1,2,\dots,n,$ $j = 1,2,\dots,r_i)$, so that $p_{ij} \geq 0$ $(1 \leq i \leq n;\ 1 \leq j \leq r_i)$, $\sum_{j=1}^{r_i} p_{ij} = 1$ $(1 \leq i \leq n)$, then the independence assumption yields for the entropy of the system with $N' = \prod_{i=1}^{n} r_i$ possible states

$$\mathcal{H}_n(\underline{p}_1,\underline{p}_2,\dots,\underline{p}_n) := H_{N'}(P_1,P_2,\dots,P_{N'}) =$$

(3)

$$= \sum_{i=1}^{n} H_{r_i}(p_{i1},p_{i2},\dots,p_{ir_i}),$$

where $\underline{p}_i$ represents the probability vector $\underline{p}_i = (p_{i1},p_{i2},\dots,p_{ir_i})$ $(i=1,2,\dots,n)$. In the following the vector notation will be reserved to denote (3).

It is the purpose of this note to study a few of the properties of H_n and to set the applications of these entropies to statistical mechanics in the framework of information theory.

2. - PROPERTIES OF H_n

Let $\Gamma_n := \{(p_1, p_2, \ldots, p_n), p_i \geq 0, \sum_{i=1}^{n} p_i = 1\}$ denote the set of complete n-ary discrete probability distributions.

THEOREM 1. The mapping $H_n : \prod_{i=1}^{n} \Gamma_{r_i} \longrightarrow R$ is, for $n = 2,3,\ldots,$

(H1) symmetric: $H_n(\underline{p}_1, \underline{p}_2, \ldots, \underline{p}_n) = H_n(\underline{p}_{\pi(1)}, \underline{p}_{\pi(2)}, \ldots, \underline{p}_{\pi(n)})$, where $(\pi(1), \pi(2), \ldots \pi(n))$ is any permutation of $(1,2,\ldots,n)$;

(H2) non-negative: $H_n(\underline{p}_1, \underline{p}_2, \ldots, \underline{p}_n) \geq 0$,

(H3) decisive: $H_n(\underline{p}_1 \underline{p}_2, \ldots, \underline{p}_n) = 0$ if and only if $\underline{p}_i = \underline{c}(r_i)$ $(i=1,2,\ldots n)$ where $\underline{c}(r_i) \in \Gamma_{r_i}$ is such that one component equals 1 while the remaining ones vanish;

(H4) expansible: $H_n(\underline{p}_1, \underline{p}_2, \ldots, \underline{p}_{n-1}, \underline{c}(r_n)) = H_{n-1}(\underline{p}_1, \underline{p}_2, \ldots, \underline{p}_{n_1})$;

(H5) maximal: $H_n(\underline{p}_1, \underline{p}_2, \ldots, \underline{p}_n) \leq \sum_{i=1}^{n} \log r_i$;

(H6) continuous: $\lim_{\underline{p}_n \to \underline{c}(r_n)} H_n(\underline{p}_1, \underline{p}_2, \ldots, \underline{p}_{n-1}, \underline{p}_n) = H_{n-1}(\underline{p}_1, \underline{p}_2, \ldots, \underline{p}_{n-1})$.

Proof. All the assertions are obvious consequences of (3) and of the properties of Shannon's entropy.

REMARKS 1. Along with (H1), H_n has a second, more obvious property of symmetry arising from the symmetry of Shannon's entropy with respect to the interchange of the components of any of the probability vectors p_i's.

2. If $r_i = r$ $(i = 1,2,,\dots,n)$ then (H5) reads

(H5') $$H_n(p_1,p_2,\dots,p_n) \leq n \log r.$$

3. - APPLICATIONS TO STATISTICAL MECHANICS.

We shall now show that the independence assumption allows one to strengthen proposition 6 and 7 in |8|, and, at one and the same time, to simplify their proofs.

In order to use the same notation as in |8| we set

$$S(\Lambda) := H_n(\underline{p}_1,\underline{p}_2,\dots,\underline{p}_n),$$

where Λ denotes the set of sites:

THEOREM 2. The following properties hold:

(S1) $$0 \leq S(\Lambda) \leq \sum_{i=1}^{n} \log r_i;$$

(S2) $$\Lambda \subset \Lambda' \Rightarrow 0 \leq S(\Lambda) - S(\Lambda') \leq \sum_{i=n+1}^{n'} \log r_i,$$

where Λ is formed by the sites numbered 1 to n and Λ' by those numbered 1 to n', with $n' > n$;

(S3) $S(\Lambda_1 \cup \Lambda_2) + S(\Lambda_1 \cap \Lambda_2) = S(\Lambda_1) + S(\Lambda_2)$ (strong additivity).

Proof. (S1) follows immediately from (3), (H2) and (H5).

(S2) $0 \leq \sum_{i=n+1}^{n'} H_{r_i}(p_{i1},p_{i2},\dots,p_{ir_i}) = S(\Lambda) - S(\Lambda) \leq \sum_{i=n+1}^{n'} \log r_i$, where

the last inequality follows from the maximality of Shannon's entropy.

(S3) If Λ_1 is composed by the sites numbered 1 to n and Λ_2 by those numbered $m+1$ to r, with $m \leq n < r$, then, by (3)

$$S(\Lambda_1 \cup \Lambda_2) + S(\Lambda_1 \cap \Lambda_2) = \sum_{i=1}^{r} + \sum_{i=m+1}^{n} H_{r_i}(p_{i1},p_{i2},\dots,p_{ir_i}) =$$

$$= \sum_{i=1}^{n} + \sum_{i=m+1}^{r} H_{r_i}(p_{i1},p_{i2},\dots,p_{ir_i}) = S(\Lambda_1) + S(\Lambda_2).$$

COROLLARY. (S4) $S(\Lambda_1 \cup \Lambda_2) < S(\Lambda_1) + S(\Lambda_2)$ (subadditivity).

REMARKS 1. In the case $r_i = 2$ $(i = 1,2,\dots,n)$ one has

(S1') $S(\Lambda) \leq n$ and (S2') $S(\Lambda') - S(\Lambda') \leq n' - n$,

which are the same upper bounds as given in [9].

2. (S3) improves on the property of strong subadditivity of [9].

3. The same expressions, (strong) additivity and subadditivity, have here a meaning different from that usual in information theory ([1]); therefore caution should be paid in using these terms. Here additivity is meant with respect to the set of sites, there with respect to probabilities.

Let the state ρ be defined as in [9]. Then one can prove as in [6] or in [10], the following theorem

THEOREM 3. If $r_i \leq r$ $(i = 1,2,\dots,n)$, then the limit

$$s(\rho) = \lim_{n\to\infty} H_n/n$$

exists and $s \in [0, \log r]$.

REMARKS 1. The above result holds in particular if $r_i = r(i = =1,2,\dots,n)$.

2. Last theorem corresponds tc propositon 7 in [8].Moreover, no mention is here made of convergence in the sense of van Hove ([9]) that was required in [8]. For theorem 3 tc hold it suffices that the number of sites in the set Λ tends tc infinity. This is certainly a simpler condition to check than the convergence in the sense of van Hove.

4. - MAXIMUM PPRINCIPLE.

We shall give two simple applications of Jaynes's maximum entropy principle ([5],[7],[2]) to H_n.

Again we shall first deal with the special case $r_i = 2$ $(i=1,2,\dots,n)$. H_n will be maximized subject to the restriction that $P=\sum_{i=1}^{n} p_i$ be assigned. Assigning P means prescribing the total number of occupied sites. As is well known, one must maximize

$$(4)\quad H_n(p_1,p_2,\dots,p_n) + \alpha \sum_{i=1}^{n} p_i = - \sum_{i=1}^{n} \{p_i \log p_i + (1-p_i) \log (1-p_i)\} + \alpha p_i$$

where α is the Lagrange multiplier. Expression (4) is the sum of function each of which depends on only one of the p_i's. Thus it suffices separately to maximize each of the functions $f:[0,1] \longrightarrow R$

$$f(p_i) := -p_i \log p_i - (1-p_i) \log (1-p_i) + \alpha p_i \quad (i=1,2,\dots,n).$$

The condition $f'(p_i) = 0$ yields $p_i = 2^{\alpha} /(1+2^{\alpha})$ $(i = 1,2,\dots,n)$. The value of the constant α is easily determined by means of the condition $P = \sum_{i=1}^{n} p_i$; one has $p_i = P/n$ $(i = 1,2,\dots,n)$. It is immediate tc check that these probability actually maximize $\tilde{H}_n$ under the given condition.

More important for its physical meaning is the following result. Let a system of non-interacting particles be characterized by the energy levels $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$; any single level can accomodate at most r particles with $1 \leq r \leq +\infty$. The case $r = 1$ corresponds to the Fermi-Dirac statistics, the case $r = +\infty$ to the Bose-Einstein statistics, whilst any other value of r corresponds to the intermediate statistics ([3]). If the probability p_i of the energy level ε_i is known, p_{ij} represents the probability of finding j particles in the state of energy ε_i. Let us introduce the random variable "number of particles per state" N which takes the values $n_{ij} = j$ $(i = 1,2,\dots,n;\ j = 1,2,\dots,r)$ and the random variable "energy per state" E which takes the value $e_{ij} = j\,\varepsilon_i$ $(i=1,2,\dots,n;\ j=1,2,\dots,r)$. Assume that the mean values of E and N are given:

(5) $$\sum_{i=1}^{n} \sum_{j=0}^{r} p_{ij}\, j\, \varepsilon_i = \langle \varepsilon \rangle ,$$

(6) $$\sum_{i=1}^{n} \sum_{j=0}^{r} p_{ij}\, j = \langle N \rangle$$

Conditions (5) and (6) give E and N the role of macroscopic observables. Maximizing (3) subject to (5) and (6), one obtains the Fermi-Dirac and the Bose-Einstein statistics together with the intermediate ones. One must, in fact, maximize

$$\sum_{i=1}^{n} \sum_{j=0}^{r} (-p_{ij} \log p_{ij} - \alpha p_{ij}\, j - \beta p_{ij}\, j\, \varepsilon_i);$$

an easy calculation leads to

$$p^*_{ij} = \exp(-\alpha j - \beta \varepsilon_i) \left\{ \sum_{k=0}^{r} \exp(-\alpha k - \beta \varepsilon_i k) \right\}^{-1}. \tag{7}$$

We refer to [2] for a proof of the existence and uniqueness of parameters α^* and β^* such that (7) satisfies conditions (5) and (6) and for a proof that (7) with $\alpha = \alpha^*$ and $\beta = \beta^*$ does indeed maximize (3) subject to (5) and (6). From (7) one can calculate the occupation number $< n_i >$ for each level ε_i

$$< n_i > = \sum_{j=0}^{n} j \, p^*_{ij} = \frac{\exp(-\alpha - \beta \varepsilon_i)}{1-\exp(-\alpha - \beta\varepsilon_i)} - \frac{(r+1)\exp\left[-(r+1)(\alpha + \beta \varepsilon_i)\right]}{1-\exp\left[-(r+1)(\alpha + \beta \varepsilon_i)\right]}$$

In order to obtain the Fermi-Dirac distribution it suffices to set $r = 1$

$$< n_i >_{F-D} = \frac{1}{\exp(\alpha + \beta \varepsilon_i)+1}$$

whilst to obtain the Bose-Einstein distribution, r will be allowed to tend to $+\infty$

$$< n_i >_{B-E} = \frac{1}{\exp(\alpha + \beta \varepsilon_i)-1} .$$

REFERENCES

[1] J.ACZÉL and Z. DARÓCZY, *On Measures of Information and Their Characterizations*, Academic Press, New York -San Francisco London,1975

[2] B.FORTE and C. SEMPI, *Maximizing Conditional Entropies: a Derivation of Quantal Statistics*, Rend. Mat.(6) 9,551-566(1976)

[3] A.M.GUÉNAULT and D.K.C.MACDONALD, *Intermediate Statistics*, Molecular Physics 5,525-530 (1962)

[4] K.HUANG, *Statistical Mechanics*, Wiley, New York-London-Sydney,1963.

[5] E.T.JAYNES, *Information Theory and Statistical Mechanics*, Phys. Rev. 106, 620-630 (1957).

[6] A.I.KHINCHIN, *Mothematical Foundations of Information Theory*, Dover, New York, 1957

[7] R.M.LEWIS, *A Unifying Principle in Statistical Mechanics*, J.Math.Phys. 8, 1448-1459 (1967).

[8] D.W.ROBINSON and D.RUELLE, *Mean Entropy States in Classical Statistical Mechanics*, Comm.Math.Phys.,5, 288-300 (1967)

[9] D.RUELLE, *Statistical Mechanics. Rigorous Results*, Benjamin, New York --Amsterdam, 1969

[10] P.WALTERS, *Entropy Theory. Introductory Lectures*, Springer, Berlin --Heidelberg -New York, 1975 (LNM 458)

Lavoro pervenuto alla Redazione il 18 Novembre 1980
ed accettato per la pubblicazione il 2 Marzo 1981
su parere favorevole di P.Benvenuti e B.Forte