Luisa Arlotti

Sulla massimizzazione dell'entropia di grado β nel caso discreto

Atti della Accademia Nazionale dei Lincei. Classe di Scienze Fisiche, Matematiche e Naturali. Rendiconti, Serie 8, Vol. **64** (1978), n.3, p. 285–291.

Accademia Nazionale dei Lincei

<http://www.bdim.eu/item?id=RLINA_1978_8_64_3_285_0>

Atti della Accademia Nazionale dei Lincei. Classe di Scienze Fisiche, Matematiche e Naturali. Rendiconti, Accademia Nazionale dei Lincei, 1978.

Scienza dell'informazione. — *Sulla massimizzazione dell'entropia di grado* β *nel caso discreto* (*). Nota di LUISA ARLOTTI, presentata (**) dal Socio D. GRAFFI.

SUMMARY. — The probability distribution of a discrete scheme is determined which maximizes the entropy of degree β, the average value of a given random variable being fixed.

1. INTRODUZIONE

Come è noto, Jaynes [1] ha dedotto la legge di distribuzione di Boltzmann della meccanica statistica postulando un principio di massimo che è possibile enunciare nella forma seguente ([2], p. 120): «La legge di equilibrio statistico di un sistema è quella che rende massima l'entropia termodinamica del sistema, tenuto conto delle condizioni alle quali esso deve soddisfare». Allo scopo Jaynes ha identificato l'entropia termodinamica con l'entropia o incertezza di Shannon. Sempre identificando l'entropia termodinamica con l'entropia di Shannon, da una formulazione più generale del principio di massima incertezza, B. Forte [3] ha dedotto la statistica di Fermi–Dirac e successivamente Tellaroli [4] quella di Bose–Einstein.

Si può tuttavia osservare che il concetto di entropia termodinamica e quello di incertezza shannoniana sono ben distinti. Ha pertanto senso, ammettendo valido il principio di Jaynes, ricercare quelle distribuzioni di probabilità che massimizzano altri tipi di entropia. Nel presente lavoro viene determinata la distribuzione che massimizza l'entropia di grado β nel caso discreto; in una successiva Nota viene ritrovata la distribuzione di Boltzmann, come distribuzione limite. È opportuno osservare che la ricerca di questo massimo non riesce agevole coi metodi usuali; perciò ho dovuto ricorrere agli accorgimenti esposti in questa Nota.

2. MASSIMIZZAZIONE DELL'ENTROPIA DI GRADO β

Sia (A , P) uno schema probabilistico finito, cioè un insieme finito di eventi: $A = A_1, A_2, \cdots, A_n$, con le rispettive probabilità $P = p_1, p_2, \cdots, p_n$ con $p_i \geq 0$, $\sum_{i=1}^{n} p_i = 1$.

Si definisce (Havrda–Charvat [5]) entropia di grado β dello schema (A , P) β essendo un parametro reale positivo diverso da uno, la grandezza rappre-

(*) Lavoro eseguito nell'ambito dei Gruppi di ricerca matematica del C.N.R.

(**) Nella seduta dell'11 marzo 1978.

sentata dalla funzione

$$H_\beta(p_1, p_2, \cdots, p_n) = \frac{1}{2^{1-\beta} - 1}\left(\sum_{i=1}^{n} p_i^\beta - 1\right).$$

Evidentemente risulta $\lim\limits_{\beta \to 1} H_\beta(p_1, p_2, \cdots, p_n) = -\sum\limits_{i=1}^{n} p_i \log_2 p_i$.

Sia $(u_1, u_2, \cdots, u_n)$ l'insieme finito dei valori di una variabile casuale U; non sarà restrittivo supporre $u_1 = 0$, $u_i < u_{i+1}\ \forall i < n$. Vogliamo riconoscere che, dato il valor medio $\bar{U}$ della variabile casuale U nell'intervallo $]0, u_n[$, esiste ed è unica la distribuzione di probabilità $\{\bar{p}_i\}$ tale che:

$$a)\quad \forall i : \bar{p}_i \geq 0 \qquad b)\quad \sum_{i=1}^{n} \bar{p}_i = 1 \qquad c)\quad \sum_{i=1}^{n} \bar{p}_i u_i = \bar{U}$$

e per la quale l'entropia H_β assume il valore massimo.

Per la dimostrazione ci serviranno le seguenti proposizioni:

A) Sia $\{p_i\}$ una distribuzione di probabilità con $p_i > 0\ \forall i$, $\sum\limits_{i=1}^{n} p_i = 1$. Siano $(u_1, \cdots, u_n)$ e $(v_1, \cdots, v_n)$ due n–ple di reali tali che $u_i < u_{i+1}$, $v_i < v_{i+1}\ \forall i < n$. Allora se è $n > 1$ risulta

$$\sum_{i=1}^{n} p_i u_i v_i > \left(\sum_{i=1}^{n} p_i u_i\right)\left(\sum_{i=1}^{n} p_i v_i\right).$$

La dimostrazione si può eseguire per ricorrenza. Sia $n = 2$. In tal caso si ha:

$$p_1 u_1 v_1 + p_2 u_2 v_2 - (p_1 u_1 + p_2 u_2)(p_1 v_1 + p_2 v_2) =$$
$$= p_1 p_2 (u_2 - u_1)(v_2 - v_1) > 0.$$

Ammessa la diseguaglianza vera per m, sia $n = m + 1$. In questa ipotesi posto

$$q_i = \frac{p_i}{1 - p_{m+1}} > 0\ (i = 1, \cdots, m) \quad , \quad u = \sum_{i=1}^{m} q_i u_i < u_{m+1},$$
$$v = \sum_{i=1}^{m} q_i v_i < v_{m+1},$$

è

$$\sum_{i=1}^{m+1} p_i u_i v_i = (1 - p_{m+1}) \sum_{i=1}^{m} q_i u_i v_i + p_{m+1} u_{m+1} v_{m+1} >$$
$$> (1 - p_{m+1})\, uv + p_{m+1} u_{m+1} v_{m+1} > \left(\sum_{i=1}^{m+1} p_i u_i\right)\left(\sum_{i=1}^{m+1} p_i v_i\right).$$

B) Sia $\{p_i\}$ una distribuzione di probabilità con $p_i > 0\ \forall i$, $\sum\limits_{i=1}^{n} p_i = 1$; siano $(u_1, \cdots, u_n)$ e $(v_1, \cdots, v_n)$ due n–ple di reali con $u_i < u_{i+1}$, $v_i > v_{i+1}$,

$\forall i < n$. Allora se è $n > 1$ risulta

$$\sum_{i=1}^{n} p_i u_i v_i < \left(\sum_{i=1}^{n} p_i u_i\right)\left(\sum_{i=1}^{n} p_i v_i\right).$$

Per la dimostrazione basterà porre $w_i = -v_i$ ed applicare la Proposizione A alle n-ple $(u_1, \cdots, u_n)$, $(w_1, \cdots, w_n)$.

Indicati dunque con $(u_1, \cdots, u_n)$ i valori della variabile casuale U, con $u_1 = 0$, $u_i < u_{i+1}\ \forall i < n$, valgono i lemmi seguenti

LEMMA 1. *Si consideri l'equazione*

$$\text{(1)} \qquad Y_\beta(\gamma) = \bar{U}$$

con $0 < \beta < 1$, *avendo definito* $\forall \gamma > -\frac{1}{u_n}$ $q_i(\gamma)$ *e* $Y_\beta(\gamma)$ *mediante le formule*

$$\sum_{j=1}^{n} (1 + \gamma u_j)^{1/(\beta-1)} q_i(\gamma) = (1 + \gamma u_i)^{1/(\beta-1)} \ (i = 1, \cdots, n) \quad Y_\beta(\gamma) = \sum_{i=1}^{n} q_i(\gamma) u_i .$$

Comunque fissato $\bar{U} \in]0, u_n[$ *l'equazione* (1) *ammette una ed una sola soluzione* $\bar{\gamma}$.

Dimostrazione. La funzione $Y_\beta(\gamma)$ è derivabile in $]-1/u_n, +\infty[$, *e*, posto $v_i = u_i(1 + \gamma u_i)^{-1}$, risulta

$$Y'_\beta(\gamma) = (\beta - 1)^{-1} \left[\sum_{i=1}^{n} q_i u_i v_i - \left(\sum_{i=1}^{n} q_i u_i\right)\left(\sum_{i=1}^{n} q_i v_i\right)\right].$$

In virtù della Proposizione A si può affermare che è $Y'_\beta < 0$ e quindi che la funzione Y_β è decrescente. Da ciò e dall'essere

$$\lim_{\gamma \to -(1/u_n)} Y_\beta(\gamma) = u_n \qquad \lim_{\gamma \to +\infty} Y_\beta(\gamma) = 0$$

l'asserto segue immediato.

LEMMA 2. *Si consideri di nuovo l'equazione* (1), *supponendo ora* $\beta > 1$ *e definendo* $q_i(\gamma)$ *e* $Y_\beta(\gamma)\ \forall \gamma > -1/u_2$ *mediante le formule*

$$\sum_{j=1}^{n} [\sup(0, 1 + \gamma u_j)]^{1/(\beta-1)} q_i(\gamma) = [\sup(0, 1 + \gamma u_i)]^{1/(\beta-1)} \qquad (i = 1, \cdots, n)$$

$$Y_\beta(\gamma) = \sum_{i=1}^{n} q_i(\gamma) u_i .$$

Comunque fissato $\bar{U} \in \left]0, \frac{\sum_{i=1}^{n} u_i^{\beta/(\beta-1)}}{\sum_{i=1}^{n} u_i^{1/(\beta-1)}}\right[$ *l'equazione* (1) *ammette una ed una sola soluzione* $\bar{\gamma}$.

Dimostrazione. Si suddivida l'intervallo I di definizione di Y_β negli $n-1$ intervalli I_k $(k = 1, \cdots, n-1)$ così definiti

$$I_k = \left]-1/u_{k+1}, -1/u_{k+2}\right] \; (k = 1, \cdots, n-2) \qquad I_{n-1} = \left]-1/u_n, +\infty\right[$$

e, fissato k, si consideri la restrizione di Y_β a I_k. Essendo in I_k

$$q_i(\gamma) = \frac{(1+\gamma u_i)^{1/(\beta-1)}}{\sum_{j=1}^{k+1} (1+\gamma u_j)^{1/(\beta-1)}} \qquad (i \le k+1); \qquad q_i(\gamma) = 0 \qquad (i > k+1)$$

tutte le q_i, e quindi Y_β, sono ivi continue e derivabili. Posto $v_i = u_i(1+\gamma u_i)^{-1}$ risulta

$$Y'_\beta(\gamma) = (\beta-1)^{-1}\left[\sum_{i=1}^{k+1} q_i u_i v_i - \left(\sum_{i=1}^{k+1} q_i u_i\right)\left(\sum_{i=1}^{k+1} q_i v_i\right)\right].$$

Per la Proposizione A si può affermare che Y_β è crescente in ciascun intervallo I_k. Poichè però Y_β è continua anche nei punti $-1/u_k$ $(k = 3, \cdots, n)$ essa risulta continua e crescente in tutto I ed assume quindi ogni valore compreso fra i limiti

$$\lim_{\gamma \to -(1/u_2)} Y_\beta(\gamma) = 0 \qquad \lim_{\gamma \to +\infty} Y_\beta(\gamma) = \frac{\sum_{i=1}^{n} u_i^{\beta/(\beta-1)}}{\sum_{i=1}^{n} u_i^{1/(\beta-1)}}$$

e ciascuno una sola volta.

LEMMA 3. *Si consideri di nuovo l'equazione* (1), *supponendo ancora* $\beta > 1$ *e definendo ora* $q_i(\gamma)$ *e* $Y_\beta(\gamma)$ $\forall \gamma \in \left]-u_{n-1}, 0\right]$ *mediante le formule*

$$\sum_{j=1}^{n} [\sup(0, \gamma + u_j)]^{1/(\beta-1)} q_i(\gamma) = [\sup(0, \gamma + u_i)]^{1/(\beta-1)} \qquad (i = 1, \cdots, n)$$

$$Y_\beta(\gamma) = \sum_{i=1}^{n} q_i(\gamma) u_i .$$

Comunque fissato $\bar{U} \in \left[\frac{\sum_{i=1}^{n} u_i^{\beta/(\beta-1)}}{\sum_{i=1}^{n} u_i^{1/(\beta-1)}}, u_n\right[$ *l'equazione* (1) *ammette una ed una sola soluzione* $\bar{\gamma}$.

Dimostrazione. La dimostrazione è simile a quella del Lemma precedente. In questo caso si divida l'intervallo J di definizione di Y_β negli intervalli $J_k = \left]-u_{k+1}, -u_k\right]$ $(k = 1, \cdots, n-2)$. Per ogni fissato k risulta allora in J_k

$$q_i(\gamma) = 0 \; (i \le k); \qquad q_i(\gamma) = \frac{(\gamma + u_i)^{1/(\beta-1)}}{\sum_{j=k+1}^{n} (\gamma + u_j)^{1/(\beta-1)}} \qquad (i > k).$$

Ciò implica che Y_β è continua e derivabile in J_k. Posto $v_i = (\gamma + u_i)^{-1}$ la derivata risulti ivi

$$Y'_\beta(\gamma) = (\beta - 1)^{-1} \left[\sum_{i=k+1}^{n} q_i u_i v_i - \left(\sum_{i=k+1}^{n} q_i u_i \right) \left(\sum_{i=k+1}^{n} q_i v_i \right) \right]$$

ed è negativa per la Proposizione B. Essendo poi Y_β continua anche nei punti $-u_k$ $(k = 1, \cdots, n-2)$ essa risulta continua e decrescente in J, assumendo ogni valore compreso fra

$$Y_\beta(0) = \frac{\sum_{i=1}^{n} u_i^{\beta/(\beta-1)}}{\sum_{i=1}^{n} u_i^{1/(\beta-1)}} \quad \text{e} \quad \lim_{\gamma \to -u_{n-1}} Y_\beta(\gamma) = u_n$$

e ciascuno una sola volta.

I risultati ottenuti permettono di dimostrare il seguente

TEOREMA. *Per ogni numero reale* $\bar{U} \in \,]0, u_n[$ *esiste una ed una sola distribuzione di probabilità* $\{\bar{p}_i\}$ *che, soddisfacendo le condizioni* (*a*), (*b*), (*c*) *rende massima l'entropia di grado* β. *Tale distribuzione si ottiene ponendo* $\bar{p}_i = q_i(\bar{\gamma})$ *con* $\bar{\gamma}$ *soluzione dell'equazione* $Y_\beta(\gamma) = \bar{U}$; *le funzioni* q_i *e* Y_β *hanno, a seconda dei valori di* β *e di* $\bar{U}$, *le espressioni indicate nei Lemmi 1, 2, 3.*

Dimostrazione. Poichè è evidentemente

$$a)\ \ \bar{p}_i \geq 0 \ \forall i \qquad b)\ \ \sum_{i=1}^{n} \bar{p}_i = 1 \qquad c)\ \ \sum_{i=1}^{n} \bar{p}_i u_i = \bar{U}$$

è sufficiente mostrare che per ogni distribuzione $\{p_i\}$ diversa da $\{\bar{p}_i\}$ che verifichi le stesse condizioni (*a*), (*b*), (*c*) risulta $H_\beta(p_1, \cdots, p_n) < H_\beta(\bar{p}_1, \cdots, \bar{p}_n)$

In base alle diseguaglianze

$$p_i^\beta - \bar{p}_i^\beta \leq \beta \bar{p}_i^{\beta-1}(p_i - \bar{p}_i) \qquad p_i^\beta - \bar{p}_i^\beta \geq \beta \bar{p}_i^{\beta-1}(p_i - \bar{p}_i)$$

valide rispettivamente (*) per $0 < \beta < 1$ e per $\beta > 1$, con il segno di eguaglianza se e solo se è $p_i = \bar{p}_i$, si ha

$$H_\beta(p_1, \cdots, p_n) - H_\beta(\bar{p}_1, \cdots, \bar{p}_n) = (2^{1-\beta} - 1)^{-1} \sum_{i=1}^{n} (p_i^\beta - \bar{p}_i^\beta) <$$
$$< \beta (2^{1-\beta} - 1)^{-1} \sum_{i=1}^{n} \bar{p}_i^{\beta-1}(p_i - \bar{p}_i) .$$

Per proseguire la dimostrazione è opportuno distinguere i tre casi:

Ipotesi del Lemma 1:

$$0 < \beta < 1 .$$

È in questo caso

$$\sum_{i=1}^{n} \bar{p}_i^{\beta-1}(p_i - \bar{p}_i) = \frac{\sum_{i=1}^{n}(1+\bar{\gamma}u_i)(p_i-\bar{p}_i)}{\left[\sum_{j=1}^{n}(1+\bar{\gamma}u_j)^{1/(\beta-1)}\right]^{\beta-1}} = 0$$

poichè entrambe le distribuzioni verificano le condizioni (*b*) e (*c*). Da ciò la tesi.

Ipotesi del Lemma 2:

$$\beta > 1 \, , \bar{U} \in \left] 0 \, , \frac{\sum_{i=1}^{n} u_i^{\beta/(\beta-1)}}{\sum_{i=1}^{n} u_i^{1/(\beta-1)}} \right[.$$

Essendo k tale che $\bar{\gamma} \in I_k$ è

$$\sum_{i=1}^{n} \bar{p}_i^{\beta-1}(p_i - \bar{p}_i) = \frac{\sum_{i=1}^{k+1}(1+\bar{\gamma}\,u_i)(p_i-\bar{p}_i)}{\left[\sum_{j=1}^{k+1}(1+\bar{\gamma}u_j)^{1/(\beta-1)}\right]^{\beta-1}} .$$

Tenendo conto che entrambe le distribuzioni di probabilità verificano le condizioni (*b*) e (*c*) e che per $i \geq k+2$ è $1 + \bar{\gamma}u_i \leq 0$ si ottiene

$$\sum_{i=1}^{n} \bar{p}_i^{\beta-1}(p_i - \bar{p}_i) = \frac{-\sum_{i=k+2}^{n}(1+\bar{\gamma}u_i)\,p_i}{\left[\sum_{j=1}^{k+1}(1+\bar{\gamma}u_j)^{1/(\beta-1)}\right]^{\beta-1}} \geq 0$$

da cui la tesi essendo $2^{1-\beta} - 1 < 0$

Ipotesi del Lemma 3:

$$\beta > 1 \, , \bar{U} \in \left[\frac{\sum_{i=1}^{n} u_i^{\beta/(\beta-1)}}{\sum_{i=1}^{n} u_i^{1/(\beta-1)}} \, , \, u_n \right[.$$

Essendo k tale che $\bar{\gamma} \in J_k$ è

$$\sum_{i=1}^{n} \bar{p}_i^{\beta-1}(p_i - \bar{p}_i) = \frac{\sum_{i=k+1}^{n}(\bar{\gamma}+u_i)(p_i-\bar{p}_i)}{\left[\sum_{j=k+1}^{n}(\bar{\gamma}+u_j)^{1/(\beta-1)}\right]^{\beta-1}} .$$

Tenendo conto che entrambe le distribuzioni verificano le condizioni (*b*) e (*c*) e che per $i \leq k$ è $\overline{\gamma} + u_i \leq 0$ si ottiene

$$\sum_{i=1}^{n} \bar{p}_i^{\beta-1}(p_i - \bar{p}_i) = \frac{-\sum_{i=1}^{k} (\overline{\gamma} + u_i)\, p_i}{\left[\sum_{j=k+1}^{n} (\overline{\gamma} + u_j)^{1/(\beta-1)}\right]^{\beta-1}} \geq 0 \,.$$

Il teorema è così completamente dimostrato.

(*) La dimostrazione è semplice; si veda ad esempio [6], p. 39.

BIBLIOGRAFIA

[1] E. T. JAYNES (1963) – *Information theory and statistical mechanics*, « Brandeis Lectures », *3*, W. A. Benjamin publ.

[2] O. ONICESCU e S. GUIAŞU (1971) – *Mécanique statistique*. Springer Verlag.

[3] B. FORTE (1966) – *Sulla massimizzazione dell'incertezza condizionata e la distribuzione di Fermi-Dirac*, « Atti Sem. mat. fis. Univ. Modena », *15*, 9–21.

[4] P. TELLAROLI (1968) – *Le distribuzioni di Bose-Einstein e di Fermi-Dirac ricavate con il principio di massima incertezza*, « Atti Sem. mat. fis. Univ. Modena », *17*, 109–117.

[5] J. HAVRDA e F. CHARVAT (1967) – *Quantification Method of Classification Processes*, « Kybernetica Cislo 1, Rocnik », *4*, 30–35.

[6] G. H. HARDY, J. E. LITTLEWOOD e G. POYA (1959) – *Inequalities*, Cambridge University Press.