Report n.25

Attrazione ed entropia

Vincenzo BRUNO

Pisa,1989
Questa ricerca è stata in parte finanziata dal Ministero della Pubblica Istruzione.

1 - In questi ultimi tempi la teoria e le applicazioni dei processi probabilistici si sono evoluti continuamente. Molti fenomeni, naturali, economici e sociali, fino a ieri sviluppati con metodi deterministici, vengono oggi osservati con questi ultimi modelli.
In questo contesto rientrano le analisi condotte con il concetto di entropia. Tale concetto, sviluppatosi nella fisica e nella teoria dell'informazione, si usa attualmente in molte altre branche del sapere. Gli elementi fondamentali della predetta teoria si rifanno al concetto d'incertezza, incertezza che esiste, quasi sempre, in ogni campo della vita. Quanto maggiore è la probabilità che si verifichi un evento, tanto più piccola è l'incertezza dello stesso. Con probabilità uguale a uno, si ha la certezza che si verifichi l'evento. Se una distribuzione è discreta con probabilità, $p_1, p_2, \ldots\ldots, p_n$, l'entropia di Shannon si definisce come segue :

$$H_n = -k \sum_{i=1}^{n} p_i \log p_i \qquad (1)$$

ove k è una costante; per k = 1, si ha

$$H_n = -\sum_{i=1}^{n} p_i \log p_i = \sum_{i=1}^{n} p_i \frac{\log 1}{p_i} \qquad (2)$$

$$0 \leq H_n \leq \log n$$

L'entropia ha il campo di variazione compreso fra 0 (massima certezza) e log n (massima incertezza) (1) .

2 - Com'è noto si può analizzare la variabilità bidimensionale dei fenomeni osservando la variabilità della tabella così come i fenomeni vengono disposti. La disamina di tale variabilità può essere condotta sia attraverso la "teoria della connessione", sia mercè la "teoria dell'entropia". Con la prima si tenta di misurare la variabilità della tabella effettiva con quella ricavata in un'ipotesi d'indipendenza probabilistica, con la seconda si misura il grado di variabilità che si delinea nel posizionare il fenomeno ad una cella o ad un'altra della tabella in oggetto (2).
Lo studio di due o più variabili in termini entropici è fecondo di risultati, tant'è che si perviene ad aspetti, in un certo senso nuovi, di coentropia (3) . Riteniamo utile, in questa sede, riesaminare i concetti

di attrazione e di repulsione cui danno luogo due particolari serie entropiche.

3 - Sia data una tavola (tab.1) della forma 2x2, detta tavola a quattro quadranti o tavola tetracorica, dove

Tabella 1

A \ B	B_1	B_2 non B_1	Totale
A_1	a	α	$a + \alpha = C$
A_2 non A_1	β	b	$\beta + b = D$
Totale	$a+\beta = E$	$\alpha + b = F$	$N = C+D = E+F$

A_1 e B_1 ; A_2 e B_2 sono modalità che esprimono la medesima qualità.
A_1 e B_2 ; A_2 e B_1 sono modalità che indicano qualità diverse.
$a + \alpha$; $\beta + b$; sono le frequenze di B rispetto ad A
$a + \beta$; $\alpha + b$; sono le frequenze di A rispetto a B

Con la predetta tabella è possibile analizzare le "relazioni interne' fra due caratteri (4). Consideriamo le frequenze relative lungo le righe (5); esse danno luogo a due serie così disposte :

$$p_{11} = \frac{a}{C} \quad ; \quad p_{12} = \frac{\alpha}{C} \tag{3}$$

$$p_{21} = \frac{\beta}{D} \quad ; \quad p_{22} = \frac{b}{D} \tag{4}$$

La tavola (tab.2) che ne deriva sarà :

Tabella 2

B / A	B_1	B_2 non B_1	Totale
A_1	p_{11}	p_{12}	1
A_2	p_{21}	p_{22}	1
Totale	$p_{11}+p_{21}$	$p_{12}+p_{22}$	2

Nella tab.2, si ha $p_{i,j} \geq 0$.

Se dai suddetti processi rileviamo le entropie, avremo :

$$H_{A_1B_1} = - p_{11} \log p_{11} = p_{11} \log (1/p_{11}) \quad (5)$$
$$H_{A_1B_2} = - p_{12} \log p_{12} = p_{12} \log (1/p_{12}) \quad (6)$$
$$H_{A_2B_1} = -p_{21} \log p_{21} = p_{21} \log (1/p_{21}) \quad (7)$$
$$H_{A_2B_2} = -p_{22} \log p_{22} = p_{22} \log (1/p_{22}) \quad (8)$$

$H_{A_1B_1}$ e $H_{A_2B_2}$ sono le entropie per i caratteri che hanno le medesime qualità, cioè simili fra loro; $H_{A_1B_2}$ e $H_{A_2B_1}$ sono le entropie per i caratteri che hanno qualità diverse, cioè dissimili (6). La tabella tetracorica (Tab.3) che ne deriva, sarà :

Tabella 3

B / A	B_1	B_2 non B_2	Totale
A_1	$H_{A_1B_1}$	$H_{A_1B_2}$	$H_{A_1B_1}+H_{A_1B_2}$
A_2 non A_1	$H_{A_2B_1}$	$H_{A_2B_2}$	$H_{A_2B_1}+H_{A_2B_2}$
Totale	$H_{A_1B_1}+$ $+H_{A_2B_1}$	$H_{A_1B_2}+$ $+H_{A_2B_2}$	N

Nella suddetta tavola presupponiamo che $H_{A_1B_1} + H_{A_1B_2}$ sia il minore numero e lo si pone uguale ad uno. Con tale somma uguale ad uno, trasformeremo la medesima tavola nella seguente (tab.4), facendo le dovute proporzioni:

Tabella 4

B / A	B_1	B_2 non B_1	Totale
A_1	m	1 - m	1
A_2 non A_1	p - m	n + m - p	n
Totale	p	n + 1 - p	n + 1

dove :

$$m = \frac{H_{A_1B_1}}{H_{A_1B_1} + H_{A_1B_2}} \qquad (9)$$

$$p = \frac{H_{A_1B_1} + H_{A_2B_1}}{H_{A_1B_1} + H_{A_1B_2}} \qquad (10)$$

$$n = \frac{H_{A_2B_1} + H_{A_2B_2}}{H_{A_1B_1} + H_{A_1B_2}} \qquad (11)$$

Dalla tabella 4 si deduce che, se n e p sono costanti, al crescere di m aumentano tutti i rapporti che si possono istituire, quindi, aumentano, generalmente, le relazioni fra le entropie di parametri simili. Cresce, in particolare, H_{A1B1} ed H_{A2B2}.

Se $p > 1$, si ha $H_{A1B1} + H_{A2B1} > H_{A1B1} + H_{A1B2}$.

Se $p < 1$, si ha $H_{A1B1} = p$ e $H_{A2B1} = 0$.

Se $p = 1$ si avrà $H_{A1B2} = H_{A2B1}$.

Se $p = n$ si avrà $H_{A2B1} = H_{A1B2}$.

Se $n = p = 1$ si ha 1/2 per tutte le quattro entropie (caso di simmetria perfetta) ed i quattro rapporti sono uguali ad m.

Per $m=1$, si ha $H_{A1B2} = 0$; per $m = 0$, si ha $H_{A1B1} = 0$.

Il valore minimo che può assumere m è 0.

Se m ha il valore massimo 1; $H_{A1B2} = 0$.

4 - L'attrazione si ottiene come segue:

$$A = \frac{H_{A_1B_1}}{H_{A_1B_1} + H_{A_1B_2}} - \frac{\dfrac{H_{A_1B_1} + H_{A_2B_1}}{H_{A_1B_1} + H_{A_1B_2}}}{n + 1} \tag{12}$$

Vi è attrazione fra simili quando nella (12) si ha:

$$\frac{H_{A_1B_1}}{H_{A_1B_1} + H_{A_1B_2}} > \frac{\dfrac{H_{A_1B_1} + H_{A_2B_1}}{H_{A_1B_1} + H_{A_1B_2}}}{n + 1} \tag{13}$$

Vi è indipendenza quando nella (12) vi è uguaglianza, fra i due rapporti ; vi è associazione fra dissimili, quando nella (12) il primo rapporto supera il secondo.

Se la (12) risulta uguale ad 1, si può avere :

$$\frac{H_{A_1B_1}}{H_{A_1B_1} + H_{A_1B_2}} \gtreqless \frac{1}{2} \qquad (14)$$

L'indice di attrazione, in termini entropici, è, per $p > 1$:

$$I_1 = \frac{\dfrac{H_{A_2B_1} + H_{A_2B_2}}{H_{A_1B_1} + H_{A_1B_2}} \cdot \dfrac{H_{A_1B_1}}{H_{A_1B_1} + H_{A_1B_2}} + \dfrac{H_{A_1B_1}}{H_{A_1B_1} + H_{A_1B_2}} - \dfrac{H_{A_1B_1} + H_{A_2B_1}}{H_{A_1B_1} + H_{A_1B_2}}}{\dfrac{H_{A_2B_1} + H_{A_2B_2}}{H_{A_1B_1} + H_{A_1B_2}} + 1 - \dfrac{H_{A_1B_1} + H_{A_2B_1}}{H_{A_1B_1} + H_{A_1B_2}}} \qquad (15)$$

La (15) dopo diversi sviluppi diventa :

$$I_1 = \frac{H_{A_1B_1} \quad (N-1) - H_{A_2B_1}}{(H_{A_1B_1} + H_{A_1B_2})\,(H_{A_2B_1} + H_{A_2B_2})} \quad ; \qquad (16)$$

Il predetto indice sarà per $p < 1$;

$$I_2 = \frac{\dfrac{H_{A_2B_1} + H_{A_2B_2}}{H_{A_1B_1} + H_{A_1B_2}} \cdot \dfrac{H_{A_1B_1}}{H_{A_1B_1} + H_{A_1B_2}} + \dfrac{H_{A_1B_1}}{H_{A_2B_1} + H_{A_1B_2}} - \dfrac{H_{A_1B_1} + H_{A_2B_1}}{H_{A_1B_1} + H_{A_1B_2}}}{\dfrac{H_{A_2B_1} + H_{A_2B_2}}{H_{A_1B_1} + H_{A_1B_2}} \cdot \dfrac{H_{A_1B_1} + H_{A_2B_1}}{H_{A_1B_1} + H_{A_1B_2}}} \tag{17}$$

La (17) dopo successivi sviluppi, sarà:

$$I_2 = \frac{H_{A_1B_1}(H_{A_2B_1} + H_{A_2B_2}) - H_{A_2B_1}(H_{A_1B_1} + H_{A_1B_2})}{H_{A_1B_1}(H_{A_2B_1} + H_{A_2B_2}) + H_{A_2B_1}(H_{A_2B_1} + H_{A_2B_2})} \tag{18}$$

Per n e p costanti, l'associazione cresce al crescere di m e raggiunge il valore 1 (avendo attrazione perfetta, quando nell'indice I_1 il rapporto corrispondente ad m è uguale a 0 e nell'indice I_2, $m= p$). I due indici, I_1 e I_2, saranno zero, quando $m= p/n+1$. In tale caso si ha indifferenza fra i gruppi di attrazione.

Quando è $p=n=1$. cioè quando le entropie delle due popolazioni sono in numero uguale, si avrà $I_1 = 2m-1$, per cui

$$I_1 = 2\,\frac{H_{A_1B_1}}{H_{A_1B_1} + H_{A_1B_2}} - 1 = \frac{H_{A_1B_1} - H_{A_1B_2}}{H_{A_1B_1} + H_{A_1B_2}} \tag{19}$$

La formula generale dell'indice di repulsione è:

$$I'_1 = \frac{\dfrac{H_{A_1B_1}}{H_{A_1B_1}+H_{A_1B_2}} \cdot \dfrac{H_{A_2B_1}+H_{A_2B_2}}{H_{A_1B_1}+H_{A_1B_2}} + \dfrac{H_{A_1B_1}}{H_{A_1B_1}+H_{A_1B_2}} - \dfrac{H_{A_1B_1}+H_{A_2B_1}}{H_{A_1B_1}+H_{A_1B_2}}}{\dfrac{H_{A_1B_1}+H_{A_2B_1}}{H_{A_1B_1}+H_{A_1B_2}}} \qquad (20)$$

La (20) dopo varî sviluppi diventa

$$I'_1 = \frac{H_{A_1B_1} H_{A_2B_2} - H_{A_1B_2} H_{A_2B_1}}{(H_{A_1B_1}+H_{A_1B_2})(H_{A_1B_1}+H_{A_2B_1})} \qquad (21)$$

con riferimento alla tabella 4; se $m=0$ si ha la repulsione perfetta.

5 - L'indice di Yule (Q) diventa, dopo facili passaggi :

$$Q = \frac{H_{A_1B_1} H_{A_2B_2} - H_{A_1B_2} H_{A_2B_1}}{H_{A_1B_1} H_{A_2B_2} + H_{A_1B_2} H_{A_2B_1}} \quad ; \qquad (22)$$

Il coefficiente ω è dato, invece, dalla formula :

$$\omega = \frac{1 - \sqrt{\dfrac{H_{A_1B_2}\, H_{A_2B_1}}{H_{A_1B_1}\, H_{A_2B_2}}}}{1 + \sqrt{\dfrac{H_{A_1B_2}\, H_{A_2B_1}}{H_{A_1B_1}\, H_{A_2B_2}}}} \qquad ; \qquad (23)$$

Gli indici Q e ω coicidono con quello del Benini soltanto nel caso di simmetria della tabella tetracorica.

E' acquista la seguente notazione :

$$E' = \frac{H_{A_1B_1}\, H_{A_2B_2} - H_{A_1B_2}\, H_{A_2B_1}}{N} \qquad ; \qquad (24)$$

L' indice di rassomiglianza s fra due caratteri, proposto dal Gini, è la media geometrica degli indici di reversione assoluta di B ad A e di A a B (7). Si ha, pertanto:

$$s = \sqrt{r_1\, r_2} \qquad (25)$$

Nella (25), sarà :

$$r_1 = \frac{H_{A_1B_1}}{H_{A_1B_1} + H_{A_2B_1}} - \frac{H_{A_1B_2}}{H_{A_1B_2} + H_{A_2B_2}} \qquad ; \qquad (26)$$

$$r_2 = \frac{H_{A_1B_1}}{H_{A_1B_1} + H_{A_1B_2}} - \frac{H_{A_2B_1}}{H_{A_2B_1} + H_{A_2B_2}} \qquad ; \qquad (27)$$

Sviluppando la (26) e la (27) si avrà :

$$r_1 = \frac{H_{A_1B_1} H_{A_2B_2} - H_{A_1B_2} H_{A_2B_1}}{(H_{A_1B_1} + H_{A_2B_1})(H_{A_1B_2} + H_{A_2B_2})} \quad ; \qquad (28)$$

$$r_2 = \frac{H_{A_1B_1} H_{A_2B_2} - H_{A_1B_2} H_{A_2B_1}}{(H_{A_1B_1} + H_{A_1B_2})(H_{A_2B_1} + H_{A_2B_2})} \quad ; \qquad (29)$$

Ne deriva che :

$$s = (H_{A_1B_1} H_{A_2B_2} - H_{A_1B_2} H_{A_2B_1}) \cdot$$

$$\cdot \sqrt{\frac{1}{(H_{A_1B_1} + H_{A_2B_1})(H_{A_1B_2} + H_{A_2B_2})(H_{A_1B_1} + H_{A_1B_2})(H_{A_2B_1} + H_{A_2B_2})}}$$

6 - Nel presente studio si sono analizzate le attrazioni dei fenomeni entropici. Le entropie, prese a due a due, vengono osservate alla luce degli indici di attrazione e di rassomiglianza. I predetti indici tendono a misurare le attrazioni (o le repulsioni) fra le singole entropie (cioè dei loro stadi d'incertezza) dei vari fenomeni. E' chiaro che le capacità di aggregazione delle predette meritano particolari approfondimenti (che si conta di sviluppare) se sono prese a tre a tre od a più di tre.

I casi concreti, acquistano una diversa rilevanza se osservati come processi entropici. Pertanto si sente l'opportunità di riesaminarli con alcuni indici che ne evidenzino i caratteri d'interdipendenza. Gli studi in oggetto, fecondi di futuri sviluppi, potranno chiarire aspetti del tutto nuovi e potranno dare della variabilità a più dimensioni conoscenze oggi ignote.

NOTE

(1) Cfr. ad es. :

Amato V. - Entropia e cause di morte, in "Annali dell'Istituto Universitario Navale di Napoli" , Napoli 1968.

Brambilla F. - Trattato di Statistica - La variabilità. UTET, Torino 1968.

Bruno V. - Entropia e variabilità, in "Studi di Statistica Metodologica", Quaderno n.6 - Istituto di Statistica dell'Università di Pisa, 1970.

Bruno V. - Particolari aspetti dell'incertezza statistica. Quaderno n.8 - Istituto di Statistica dell'Università di Pisa,1972

Bruno V. - La variabilità e l'incertezza statistica, con particolare riferimento alle distribuzioni paretiane dei redditi, in "Giornale degli Economisti e Annali di Economia", settembre-ottobre 1975.

Capecchi V. - Une méthode de classification fondée sur l'entropie, in "Revue française de sociologie" 1964.

De Vergottini M. - Su alcuni valori medi, indici di variabilità e di relazione. Quaderno n.6, Istituto di Statistica della Università di Pisa.

Kullbock S. - Information theory and statistics - Dover pubblications, New York, 1968.

Lauro N. - Aspetti statistici della teoria dell'informazione alla luce di un'applicazione, in "Studi di Statistica"; Quaderno n.1, Istituto di Statistica e Demografia, Università di Napoli.

Leti G. - Sull'entropia, su un indice del Gini e su altre misure dell'eterogeneità di un collettivo, in "Metron", vol.XXIV.

Leti G. - Una nota sulla misura mediante l'entropia del grado di associazione di due distribuzioni, in "Studi in onore di G. Pompilj", Oderisi Tipografia, Gubbio 1971.

(2) Brambilla F. - Trattato di Statistica, op.cit. pag.227.

(3) Astola J. and Virtanen I. - Entropy correlation coefficient, a measure of statistical dependence for categorized data. Atti del XV Congresso Europeo di Statistica - Palermo, settembre 1982.

Marasini D. - Una misura della connessione nel caso di una mutabile tripla. in "Atti della XXXII Riunione Scientifica", Sorrento 11-13 aprile 1984, vol.I.

(4) Galeotti G.- Dell'indice E di una tavola tetracorica ad un solo margine prefissato, in "Annali dell'Istituto di Statistica" Università di Bari, vol.XXX, Cacucci 1966-67.

(5) Bruno V. - Su alcuni indici delle relazioni statistiche per particolari processi stocastici, Società Italiana di Statistica - XXVIII Riunione Scientifica, Padova 20-22 maggio 1975.

(6) Bachi R. - Gli indici di attrazione matrimoniale, in "Giornale degli Economisti e Annali di Economia" Padova 1929.

Benini R. - Le combinazioni simpatiche in Demografia in : "Rivista Italiana di Sociologia" anno II fasc. III, 1898.

Benini R. - Principi di Demografia - Firenze, Barbera 1901.

Savorgnan F. - La misura dell'endogamia e dell'esogamia, in "Atti del Congresso Internazionale per gli Studi sulla Popolazione" - Roma 7-10 settembre 1931, vol.X, Roma 1934.

(7) Gini C. - Indici di omofilia e di rassomiglianza e loro relazione col coefficiente di correlazione e con altri indici di attrazione, in "Atti del Reale Istituto Veneto delle Scienze" 1914-15.