

Complessità statistica

Giovanni Pistone

Collegio Carlo Alberto,
Via Real Collegio 30,
10024 Moncalieri

giovanni.pistone@carloalberto.org

Complessità e struttura sono due termini che vengono usati sia in senso generico, senza una precisa definizione, sia in senso tecnico. Entrambi i termini sono dunque di solito accompagnati, nel discorso scientifico, da una qualificazione che ne precisa l'ambito di applicazione.

La nozione di struttura in matematica designa un sistema assiomatico di oggetti e di relazioni. La definizione assiomatiche ha lo scopo di mettere in evidenza gli isomorfismi tra diverse strutture, cioè la possibilità di trasformare gli oggetti dell'una negli oggetti dell'altra conservandone le relazioni reciproche. Il caso classico è la relazione tra geometria della retta e algebra dei numeri reali. Con un sistema di coordinate cartesiane ogni punto della retta corrisponde ad un numero reale e viceversa. Inoltre l'operazione di somma di due segmenti diventa la somma delle coordinate. Questo permette di trattare la geometria con l'algebra o anche di interpretare l'algebra con la geometria. I numeri reali nascono nella Grecia classica sulla base di operazioni geometriche e, con Cartesio (1596-1650), i concetti geometrici cominciano ad essere tradotti sistematicamente in concetti algebrici.

Il concetto scientifico di complessità è molto più moderno. Nel linguaggio corrente un oggetto complesso è un oggetto di cui non si possono dare regole semplici di funzionamento. In matematica sono stati fatti molti tentativi di definizione della complessità, come la complessità computazionale di un algoritmo, complessità di Kolmogorov, complessità dei sistemi dinamici caotici, complessità dei sistemi biologici. Per questi usi si possono vedere altre relazioni di questa serie. Qui ci occupiamo della complessità dei sistemi a molti componenti di cui interessa il comportamento collettivo: è la complessità *statistica*.

Recentemente il termine complessità ha avuto un grande favore nell'ambito della gestione e della politica: l'Unione Europea e altri organismi di promozione delle scienze applicate hanno favorito la nascita di reti di centri di ricerca sulla complessità. Uno di queste è la rete ASSYST (1) di cui citeremo alcune ricerche più avanti. L'Unione Europea probabilmente finanzia un grande progetto di ricerca sulla complessità, con applicazione principale ai sistemi sociali quali la sanità e i trasporti. In questo contesto la ricerca matematica discute un certo numero di metodologie e concetti che si spera suscettibili di sviluppo (e di essere finanziati).

Interazioni, grafi, complessi simpliciali

Un ingrediente dei modelli matematici applicati ai sistemi sociali è un concetto che viene dalla statistica, quello di *interazione*. Consideriamo un *effetto* che dipende dai livelli di vari *fattori*. In un modello senza interazioni il livello dell'effetto dipende dal livello di ciascuno dei fattori e la quantità di effetto di ogni fattore non dipende dal livello degli altri. In formule, l'effetto Y ha un livello di base m ed è proporzionale a ciascuno dei livelli dei fattori X1 e X2,

$$Y = m + a X1 + b X2.$$

Se faccio variare X2 tenendo fisso X1 al livello x1, l'effetto è

$$Y = (m + a x_1) + b X_2$$

ed è come se il livello di riferimento fosse passato da m a (m + a X1).

Se c'è interazione (nel senso statistico del termine) allora l'effetto della variazione di X2 dipende dal livello di X1 e l'effetto della variazione di X1 dipende dal livello di X2. La formula lineare usata sopra non è più valida. La statistica cerca allora di misurare l'interazione separando gli *effetti semplici* dei fattori dall'*effetto dell'interazione* scrivendo formule del tipo

$$Y = m + a X_1 + b X_2 + c X_1 X_2.$$

Le interazioni si possono rappresentare graficamente. Un sistema con due fattori A e B senza interazioni si rappresenta con due punti. Se c'è interazione con due punti uniti da un segmento. Se ci sono tre fattori A, B, C, la situazione si complica. Possono essere assenti le interazioni, possono esserci solo interazioni a due a due, ci possono essere interazioni triple. Si pensi ad un comitato di tre persone A, B, C, che decidono

- 1) senza parlarsi (A, B, C); oppure
- 2) scambiandosi telefonate a due a due (AB, AC, BC); oppure
- 3) discutendo in una conferenza cui partecipano tutti (ABC).

Oppure si pensi ad una persona A e due amanti C e D. Ci sono varie situazioni di relazione possibili:

- 1) B e C hanno rapporti solo con A (AB, AC); oppure
- 2) B e C si conoscono (AB, AC, BC); oppure
- 3) si incontrano tutti insieme (ABC).

Le dinamiche psicologiche sono molto diverse nei tre casi!

I sistemi a più componenti e interazioni al più doppie si rappresentano bene con grafi. Un *grafo* è costituito da una lista di vertici e da una lista di coppie di vertici che rappresentano le interazioni doppie esistenti nel sistema. La teoria matematica dei grafi è molto ricca ed è oggetto di ricerca contemporanea. (2)

Alcuni esempi che sono oggetto di intensa attività di ricerca da parte di ricercatori della rete ASSYST. Il traffico automobilistico in una strada ad una corsia prevede interazioni solo tra due auto vicine ed è rappresentabile con un grafo. Ma questo è solo un modello approssimativo nel caso di una autostrada a più corsie, dove il guidatore tiene conto sia delle auto che gli sono davanti che di quelle che lo seguono: la rappresentazione con un grafo non è adatta perché sono presenti interazioni multiple. Un secondo esempio sono i sistemi sociali a rete, come quelli che si hanno considerando tutti gli scambi a coppie nella rete Internet. Si pensi ad esempio agli scambi di messaggi di posta elettronica. Un terzo esempio sono gli oggetti che si muovono in uno *sciame*. Ogni uccello dello sciame regola il proprio movimento sulla base dell'osservazione della posizione degli altri uccelli. Queste interazioni a coppie producono un comportamento coerente particolare.

Nel caso di interazioni multiple si preferisce la rappresentazione con un complesso simpliciale. Un *complesso simpliciale* è una collezione di sottoinsiemi di un insieme tale che, se contiene un insieme, allora contiene tutte le sue parti. Ad esempio

$$A, B, C, D, AB, AC, BC, ABC, AD$$

è un sistema di quattro persone A, B, C, D. le prime tre formano un gruppo ABC ed in particolare

hanno rapporti a due a due AB, AC, BC. La quarta persona D ha rapporti solo con A. Questa struttura si rappresenta graficamente disegnando un quadrato di vertici A, B, C, D, evidenziando il triangolo pieno ABC e aggiungendo il segmento AD. I lati del triangolo ABC costituiscono il bordo del simpleso ABC. Questo linguaggio permette di descrivere efficacemente tutti i tipi di interazione e inoltre approfitta di una teoria matematica consolidata.(3)

La statistica tradizionale, cioè quella iniziata dal lavoro di Ronald Fisher (1890-1962), considera sistemi non complessi nel senso che possono sì avere interazioni, ma sono comunque sistemi di causa (X_1 e X_2) ed effetto (Y). Il concetto di interazione resta comunque valido e utile anche nella teoria della complessità: una definizione possibile di tipo fisico è che un sistema complesso è quello in cui l'effetto delle interazioni è prevalente rispetto agli effetti dovuti ai singoli componenti. Sempre dalla fisica ci viene un ulteriore approfondimento, basato sulla osservazione che sistemi complessi presentano più scale dimensionali.

Sistemi multiscala

I sistemi complessi di interesse attuale presentano anche relazione tra effetti fisici che vivono in varie scale dimensionali, dal microscopico al macroscopico. Non si tratta solo di considerare contemporaneamente fenomeni diversi, l'uno di scala microscopica e l'altro di scala macroscopica, ma di considerare effetti microscopici e macroscopici nello stesso fenomeno. Sono concetti nati in Fisica Statistica, inizialmente come teoria del calore. Un gas è costituito da molte molecole, il moto di ciascuna delle quali è caratterizzato da quantità meccaniche: posizione, velocità, massa, dimensione. Le molecole interagiscono urtandosi e questo determina un comportamento collettivo macroscopico che è descritto da quantità fisiche macroscopiche: quantità di calore, temperatura.

Si tratta di una teoria classica che risale alle ricerche di Ludwig Boltzmann (1844-1906) e Willard Gibbs (1839-1903).(4) Genericamente, si può dire che la Fisica Statistica è la Statistica applicata alla Fisica, ed alcuni manuali dicono proprio questo. Ma questo non rende l'idea della grande svolta concettuale avvenuta nella fisica a cavallo tra '800 e il '900, simbolizzabile nella generale accettazione dell'esistenza degli atomi. La contraddizione tra la meccanica newtoniana e la termodinamica viene risolta distinguendo due livelli di descrizione, uno microscopico (newtoniano) e uno macroscopico (termodinamico). Le proprietà macroscopiche si spiegano e calcolano come comportamenti medi in presenza di molte particelle, collegando così, quantità fisiche come la quantità di moto con quantità termodinamiche come la temperatura.

Ci sono dunque due livelli di realtà che non corrispondono a due descrizioni imprecise e contraddittorie, ma corrispondono ad una struttura complessa, con leggi che permettono di legare l'una all'altra. Nell'ambito di questo programma di ricerca emergono alcuni concetti che saranno cruciali nelle ricerche sulla complessità.

- ⤴ *Entropia, informazione* L'entropia o informazione è una quantità associata alle distribuzioni di probabilità e ai sottoinsiemi di un insieme finito. È un mezzo di calcolo fondamentale, secondo alcuni una grandezza fisica, anche indipendentemente dalla interpretazione termodinamica. Produce in fisica un nuovo principio di minimo, come lo sono la minima azione, il minimo cammino.
- ⤴ *Interazione fisica* Nella descrizione di Boltzmann le particelle interagiscono urtandosi, secondo uno dei paradigmi classici di casualità. Un altro famoso modello, quello di ferromagnetismo di Ising prevede oggetti discreti, gli *spin*, a valori ± 1 , sistemati in un reticolo regolare, che interagiscono non urtandosi, ma attraverso il campo magnetico.
- ⤴ *Locale e globale* Nel modello di Ising le interazioni sono descritte localmente, ogni oggetto interagisce con quelli vicini. Questo produce una configurazione statistica globale espressa da una formula globale. Viceversa, dalla descrizione globale si può passare alla descrizione della dipendenza locale.

L'evoluzione del sistema aggiunge una dimensione che non fa parte della descrizione standard, che riguarda lo stato stazionario.

Un esempio: le epidemie

Secondo l'informatico Peter Sloot(5) dell'università di Amsterdam, lo studio delle pandemie, quale quella dell'AIDS deve necessariamente fondarsi su una scienza della complessità che utilizza tutti gli strumenti che abbiamo descritto ed altri da sviluppare. Infatti una epidemia per sua natura si basa sulla trasmissione da individuo ad individuo, dunque sulle relazioni tra individui piuttosto che sugli individui stessi. Per questa ragione si tratta di valutare, a livello politico, se sono più efficaci investimenti sulla cura dei malati o investimenti sulla struttura della società intesi a modificare le interazioni tra gli individui.

Questo problema presenta una enorme variabilità di scala:

geni < proteine < cellule < organi < individui < popolazione.

All'interno di ogni livello si osservano dinamiche complesse e ad ogni passaggio di scala si osservano le dualità microscopico-macroscopico e locale-globale.

A valle di queste ricerche, Sloot conclude che nei grandi agglomerati urbani l'intervento sulle strutture sociali è molto più efficace del trattamento sistematico e che queste politiche sociali conviene siano accompagnate da diagnostiche mirate seguite da un opportuno trattamento medico.

Una formalizzazione matematica di complessità

Una delle definizioni più comuni di complessità è questa:

Il tutto è più della somma delle sue parti.

Ad esempio, è il sottotitolo del disco *Inneres Auge* del cantante Franco Battiato, e compare frequentemente nella pubblicistica politica e divulgativa, di solito con intenzione polemica antiriduzionista: la società è più degli individui che la compongono, un vivente è più delle cellule, il pensiero è più dei processi neurologici, Dio è più del Mondo. Questa frase viene da alcuni attribuita al filosofo classico Aristotele. Ma probabilmente, almeno nel senso attuale, ci viene dal pensiero vitalista del 18o secolo. Se si vede il mondo come stratificato in gerarchie di livelli, ciascuno dei quali con proprie regole di funzionamento, si può argomentare che ogni livello superiore si appoggia ma non è determinato dalle leggi del livello inferiore. È una concezione più forte di quella che emerge dalla fisica statistica, che riconosce sì due livelli, uno microscopico e l'altro macroscopico, ma si ingegna comunque a ritrovare le leggi del macroscopico come comportamenti collettivi dei componenti microscopici. Nell'ambito prima della sociologia, poi della psicologia cognitiva ed evolutivista e della biologia evolutivista, è nata l'idea di una *causalità discendente* che vuole ritrovare l'influenza del livello superiore sul livello inferiore. Il termine è stato introdotto nel 1974 dal sociologo americano Donald T. Campbell (1916-1996). A questo stesso programma di ricerca appartengono altri termini intesi a tematizzare i complicati rapporti tra i vari livelli e la loro dinamica temporale: *epistemologia evolutiva, emergenza*.(6)

Il matematico e informatico Nihat Ay(7), dell'Istituto Max Planck di Lipsia, ha sviluppato una definizione matematica di complessità che parte proprio dalla frase citata sopra e costituisce un bell'esempio di come funziona un modello matematico. Questo ricercatore considera un insieme di oggetti *O*, le "parti," ed una classe di altri oggetti composti *T*, destinati a rappresentare i possibili "tutto." Il sistema di cui vogliamo valutare la complessità appartiene a quest'ultima classe. La

definizione “aristotelica” si interpreta tramite due operazioni di “somma” e “parti”:

- ▲ è data una trasformazione S dagli oggetti ai composti, che, per definizione, è semplice o non complessa (la "somma");
- ▲ è data una trasformazione P dai composti agli oggetti (le "parti");
- ▲ è data una funzione distanza $d(c_1, c_2)$ tra i composti c_1 e c_2 in T ;
- ▲ la complessità di un composto c in T è la sua distanza dal composto semplice ottenibile con le stesse parti. Cioè $P(c)$ sono le parti di c , $S(P(c))$ è il composto semplice che ha le stesse parti di c , e $d(c, S(P(c)))$ è la complessità del composto c .

Si tratta di una definizione tipicamente matematica, che non si riferisce direttamente ad una specifica contingenza. Per applicarla è necessario un lavoro di costruzione di un modello matematico, nel senso che bisogna fare le opportune scelte per l'elenco degli oggetti O , l'elenco dei composti T , definire le due operazioni somma S e partizione P , definire una opportuna distanza.

Facciamo un esempio artificioso, ma utile per illustrare il meccanismo. Se i componenti sono insiemi con ripetizione delle cifre binarie 0 e 1, e i composti sono le parole binarie di lunghezza 5, possiamo usare come distanza il numero di cifre diverse. Con questa distanza i composti $c_1 = 01011$ e $c_2 = 01101$ sono a distanza $d(01011, 01101) = 2$ perché la terza e la quarta cifra sono diverse. Definiamo come parte di una parola l'insieme delle cifre che la compongono, $P(01011) = “0$ due volte e 1 tre volte”; definiamo come somma S la parola che si ottiene mettendo prima tutti gli zeri poi tutti gli uni, cioè $S(0$ due volte e 1 tre volte) = 00111. La distanza tra $c = 01011$ e $S(P(c)) = 00111$ è $d(01011, 00111) = 2$.

Più tecnicamente diciamo, per il lettore che conosce la statistica, che Nihat Ay considera ad esempio come composti le distribuzioni di probabilità congiunte in un modello statistico, come parti le distribuzioni marginali, come somma la distribuzione congiunta ottenuta con l'ipotesi di indipendenza, e come distanza una entropia relativa.

Questo tipo di teoria è suscettibile sia di sviluppi matematici che applicativi. La ricerca matematica si occuperà di mettere in relazione questa definizione di complessità con le altre disponibili e di studiarne le proprietà. Nelle applicazioni si tratta di valutare se questo indice è utilizzabile. Ad esempio, si potrebbero valutare diversi quadri clinici dal punto di vista di questa valutazione della complessità per vedere se da questo numero indice si possono ricavare criteri per la diagnosi e per la scelta delle terapie.

Questa breve illustrazione di alcuni criteri di complessità dal punto di vista della ricerca matematica e applicata mi permette di trarre alcune conclusioni. Primo, l'interesse attuale negli ambienti scientifici, tecnici, gestionali e politici per la complessità è giustificata dalla sua effettiva utilità per concettualizzare e modellare situazioni difficili e dalla potenziale disponibilità di strumenti scientifici e tecnici per operare utilmente. Secondo, gli strumenti matematici disponibili o in corso di sviluppi sono sia strumenti classici di metodologia consolidata, sia strumenti nuovi in corso di sviluppo. Dal punto di vista filosofico, la nozione di complessità è una delle direzioni nelle quali sono in corso di definitivo superamento le concezioni positivistiche del mondo, che erano basate su acquisizioni scientifiche di tipo meccanico, dunque non relative a sistemi complessi.

Note

- (1) Informazioni sulla rete ASSYST si trovano su Internet all'indirizzo <http://www.assystcomplexity.eu/>. Membro italiano della rete è l' istituto ISI diretto dal fisico Mario Rasetti, vedi all'indirizzo <http://www.isi.it/about>.
- (2) Per un orientamento ed illustrazioni grafiche si può consultare la voce *Grafo* di Wikipedia all'indirizzo <http://it.wikipedia.org/wiki/Grafo>.
- (3) Altri esempi ed illustrazioni si trovano alla voce *Complesso simpliciale* di Wikipedia all'indirizzo http://it.wikipedia.org/wiki/Complesso_simpliciale.
- (4) Su Boltzmann: http://it.wikipedia.org/wiki/Ludwig_Boltzmann. Su Gibbs: http://it.wikipedia.org/wiki/Willard_Gibbs.
- (5) Informazioni su Sloot, anche lui membro della rete ASSYST, si trovano all'indirizzo <http://staff.science.uva.nl/~sloot/>.
- (6) Informazioni su Campbell e sulle sue linee di ricerca di trovano all'indirizzo http://en.wikipedia.org/wiki/Donald_T._Campbell.
- (7) Informazioni sulle ricerche di Ay (membro di ASSYST) si trovano all'indirizzo <http://www.mis.mpg.de/ay/homepages/nay/>.

Sunto

La teoria della complessità ha suscitato recentemente grande interesse in vista delle sue applicazioni alla società, ad esempio ai sistemi sanitari. Vari istituti di ricerca sono attivi nello sviluppo di metodologie scientifiche in grado di affrontare questo tipo di applicazioni. I modelli matematici di complessità che vengono usati sono basati sia su strumenti classici della Statistica e della Fisica Statistica che su nuove idee che cercano di utilizzare stimoli che vengono dal moderno pensiero evoluzionistico. Questo articolo è basato su un breve intervento alla sessione di "Le parole della scienza" seguita al convegno Sefir "Complessità e Struttura", Città del Vaticano 25 novembre 2011.

Abstract

Complexity theory has recently raised considerable interest in view of its applications to social system such as health care. A number of research institution are active in developing scientific methods to deal with such applications. The mathematical models for complexity which are used are based both on classical tools from Statistics and Statistical Physics and on new ideas intended to tackle the modern evolutionary thinking. This paper is based on a short talk presented to the discussion session "Le parole della Scienza" following the Sefir meeting "Complessità e struttura", Vatican City Nov 25, 2011.