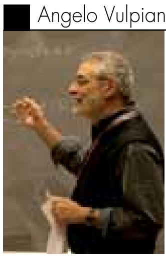
ANDREI ANDREEVICH MARKOV: UN MATEMATICO FURIOSO E LE SUE CATENE

di Angelo Vulpiani

Angelo Vulpiani



Docente di Fisica teorica, insegna Meccanica statistica e Meccanica statistica del non equilibrio presso il Dipartimento di Fisica dell'Università di Roma "La Sapienza". È stato visitatore presso diversi istituti di ricerca e Università in Francia, Belgio, Svezia, Danimarca e Stati Uniti. I suoi interessi scientifici riguardano il caos e la complessità nei sistemi dinamici, la Meccanica statistica di non equilibrio e dei sistemi disordinati, la turbolenza sviluppata, i fenomeni di trasporto e diffusione. Oltre ad articoli su riviste internazionali e libri specialistici in inglese, ha pubblicato: *Determinismo e Caos* (Nuova Italia Scientifica, 1994; Carocci, 2004) e *Caso, Probabilità e Complessità* (Ediesse, 2014).

Vita (agitata) ed opere

Da giovane Andrei Andreevich Markov (1856-1922) non fu il tipico studente brillante: i suoi risultati erano modesti in tutte le materie (con l'unica eccezione della Matematica); aveva inoltre un carattere decisamente ruvido e poco accomodante. Venne così soprannominato *Andrei reistovy* (il furioso), diventò poi *l'accademico militante* quando, da affermato scienziato, si impegnò costantemente contro l'autocrazia dello Zar, il servilismo delle istituzioni e l'oscurantismo della chiesa ortodossa.

Nel 1874 entrò all'Università di San Pietroburgo dove studiò Matematica sotto la guida di ottimi docenti come Korkin e Zolotarev e soprattutto con il più importante matematico russo dell'epoca, il grande Pafnuty Lvovich Chebyshev (1821- 1894), che fu suo mentore. Ebbe una carriera folgorante: ancora studente, nel 1877, le sue ricerche sull'uso delle frazioni continue per risolvere equazioni differenziali sono premiate con una medaglia d'oro; nel 1880 discute la tesi e nel 1884 consegue il dottorato;

ANDREI ANDREEVICH MARKOV (IMMAGINE DI PUBBLICO DOMINIO, TRATTA DA WWW.WIKIPEDIA.ORG)

a trenta anni è già professore presso l'Università di San Pietroburgo; pochi anni dopo è eletto socio della prestigiosa Accademia delle Scienze.

Markov con il suo carattere poco conciliante ebbe non pochi problemi ma per fortuna nei suoi anni giovanili poté usufruire della protezione di Chebyshev che aveva capito il suo valore.

Ecco solo un breve elenco delle sue attività antiautoritarie. Nel 1902 protestò contro il pavido comportamento dell'Accademia delle Scienze che, sotto la pressione dello Zar, non ratificò l'elezione come membro onorario dello scrittore M. Gorky ed accolse come soci alcuni membri dell'aristocrazia completamente privi di meriti culturali.

Quando il ministro dell'interno diede disposizioni che i docenti universitari venissero equiparati a funzionari di polizia e quindi tenuti a riferire di eventuali attività antigovernative degli studenti, Markov rispose che era un professore di teoria delle probabilità e non un poliziotto, in ogni caso non approvava queste disposizioni e "*non poteva cambiare opinione per ordine dei superiori*".

Nel 1905 si scagliò contro le disposizioni dell'Università di San Pietroburgo sulle quote di ammissione per studenti ebrei.

In seguito alla decisione del sinodo della chiesa ortodossa di scomunicare Tolstoj, Markov chiese, con una lettera formale, di essere scomunicato pure lui in quanto condivideva le posizioni del grande scrittore. La sua richiesta venne accolta, ma solo parzialmente in quanto non era stata inviata in modo appropriato; il sinodo rispose in modo formale: "*Markov si è allontanato dalla Chiesa di Dio e noi lo espelliamo dalla lista dei Credenti Ortodossi*".

La cosa interessante è che il radicalismo di Markov non rimase un fatto privato, ma fu importante anche nella sua attività scientifica e nella storia della teoria delle probabilità.

Markov fin dalla nascita ebbe problemi alle gambe, subì diverse operazioni chirurgiche e spesso fu costretto ad usare le stampelle. Gli ultimi anni della sua vita furono rattristati dal suicidio del collega ed amico Alexander Mikhailovich Lyapunov (1857-1918) e da crescenti problemi di salute; morì nel 1922 per i postumi di un'operazione alle gambe. Fu anche un giocatore di scacchi di livello internazionale: nel 1892 sostenne quattro partite con M. Chigorin (campione russo sfidante del campione del mondo in carica), ne vinse una e ne pareggiò un'altra. Uno dei figli di Markov, Andrei Andreevich junior (1903-1979), ha seguito le orme del padre ed è stato un matematico di grande livello, con contributi significativi all'Algebra ed alla Logica.

L'attività matematica (in particolare quella sul calcolo delle probabilità) di Markov a San Pietroburgo si inserisce in una consolidata tradizione che inizia nel 1725 con la nascita dell'Accademia delle Scienze voluta da Pietro il Grande. Tra i suoi membri l'Accademia ha annoverato matematici come Eulero e Daniel Bernoulli. Quest'ultimo nel 1738 inaugurò la probabilità in Russia con un articolo sui rischi e da allora la probabilità è stata una delle branche preferite dai matematici russi. Chebyshev, il maestro di Markov, fu colui che, intorno al 1850, cominciò a dare piena dignità matematica alla probabilità. Tra i contributi di Markov alla teoria della probabilità ricordiamo il perfezionamento della dimostrazione di Chebyshev del teorema del limite centrale (TLC). La prima dimostrazione generale del TLC per variabili indipendenti è dovuta invece ad un altro famoso allievo di Chebyshev, A.M. Lyapunov, che utilizzò i momenti e le funzioni caratteristiche.

Dopo la rivoluzione del 1917 l'indiscusso prestigio scientifico del nostro e le sue opinioni progressiste contribuirono alla decisione del nuovo governo di potenziare la scuola matematica di San Pietroburgo a discapito di quella di Mosca, che annoverava non pochi scienziati conservatori. Inutile dire che Andrei il furioso trovò il modo di entrare in conflitto anche con il nuovo regime.

Markov fu un insegnante entusiasta: era fermamente convinto che l'unico modo serio di apprendere la Matematica fosse risolvendo problemi; sempre disponibile per i suoi allievi anche per lezioni non ufficiali durante le vacanze, era amatissimo dai sui studenti; molti, dopo aver superato l'esame, seguivano una seconda volta i suoi corsi.

Nel 1913, quando lo Zar indisse un anno di celebrazioni per una ricorrenza della dinastia dei Romanov, Markov, poco entusiasta dell'iniziativa, cercò di contrastare i festeggiamenti ed organizzò una grande conferenza internazionale per il secondo centenario della pubblicazione dell'*Ars Conjectandi* di Jacob Bernoulli, dove appare la prima dimostrazione della legge dei grandi numeri si veda il Box A).

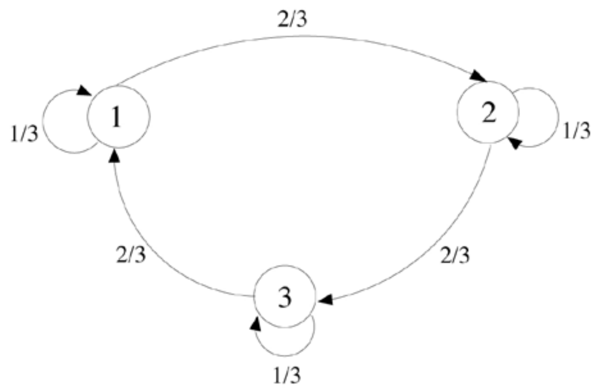


FIG. 1: SCHEMA GRAFICO DI UNA CATENA DI MARKOV CON 3 STATI, IN CORRISPONDENZA DI CIASCUNA FRECCIA È INDICATA LA PROBABILITÀ DI TRANSIZIONE

All'inizio del XX secolo Markov fu coinvolto in un animato dibattito con il matematico moscovita Nekrasov, le cui opinioni politiche e religiose erano opposte alle sue. Il controverso problema verteva sulla regolarità statistica dei comportamenti sociali (si veda il Box B). Markov sosteneva che tale regolarità fosse una mera conseguenza della legge dei grandi numeri, niente a che vedere con cose come libero arbitrio o convinzioni politiche e religiose. Nekrasov faceva notare che la legge dei grandi numeri non poteva essere sufficiente a spiegare la regolarità statistica, in quanto tale legge varrebbe solo sotto l'ipotesi di eventi indipendenti. In qualche modo l'osservazione di Nekrasov era sensata, infatti all'epoca l'unica legge dei grandi numeri dimostrata era quella di J. Bernoulli per eventi indipendenti. Markov, per controbattere a questa obiezione, dovette costruire una teoria di processi non indipendenti. Da una lettera ad un suo amico è ben evidente quanto fosse contento di aver messo in difficoltà il suo collega moscovita: *"L'unico ruolo di P.A. Nekrasov, a mio avviso, è stato nell'aver sollevato il problema* (...). *Sono arrivato alla costruzione di un sistema con un carattere così generale che P.A. Nekrasov non può neanche sognarsela. Ho considerato variabili connesse in una semplice catena, da questo viene l'idea della possibilità di estendere i teoremi limite anche a queste catene".*
Questo è il certificato di nascita di quei processi stocastici ora chiamati catene di Markov (CM), che trovano utilizzo in una vasta classe di problemi in Fisica, Chimica, Biologia, Economia ed addirittura nel motore di ricerca di Google (si veda il Box B).
Per dare un'idea delle catene di Markov e di come sia possibile avere la legge dei grandi numeri anche per eventi non indipendenti, consideriamo il seguente gioco in cui si hanno tre cerchi numerati con 1, 2 e 3; dentro ogni cerchio c'è una sorta di roulette con caratteristiche diverse per ogni cerchio. Un viaggiatore parte dal cerchio numero 1, la cui roulette è divisa in due settori, contrassegnati con il numero 1 (in un settore di 120 gradi) e 2 (in un settore di 240 gradi) e fa girare la roulette: se esce 1 rimane fermo, mentre se esce 2 salta nel cerchio 2. Abbiamo quindi che rimane in 1 con probabilità 1/3 oppure salta in 2 con probabilità 2/3. Poi il gioco si ripete. Avremo quindi che le regole dei salti sono definite dalle varie roulette, in termini matematici dalle probabilità $P_i \rightarrow j$ di saltare dal cerchio i al cerchio j. Le posizioni che si susseguono nel tempo non sono indipendenti, ad esempio nel caso mostrato in Figura 1 abbiamo che da 1 si ha 2 più facilmente che 1 e non si ha mai 3; da 2 si ha 3 più facilmente che 2 e non si può avere 1 ecc.
È possibile mostrare, usando la simmetria del problema, che la legge dei grandi numeri vale: nonostante gli eventi non siano indipendenti, se il gioco viene protratto a lungo il viaggiatore passa in ogni stato un terzo del tempo.
Originariamente Markov considerò il caso con variabili dicotomiche che possono assumere solo due valori, in seguito trattò il caso più generale con un numero finito di stati, dimostrando come la legge dei grandi numeri valga, sotto opportune condizioni (si veda il Box C), anche per variabili non indipendenti.
Pochi anni dopo Markov utilizzò le CM per un'analisi statistica di alcuni testi di Pushkin; il suo approccio, da un punto di vista linguistico, era piuttosto elementare in quanto trattava il testo come una successione di vocali e consonanti. Tuttavia questi lavori sono stati il punto di partenza dell'uso di tecniche probabilistiche per la linguistica: ancora oggi le CM sono usate per l'attribuzione dei testi.
Le catene di Markov che descrivono processi stocastici che assumono solo stati discreti ed evolvono con un tempo discreto costituiscono il caso più semplice non banale di processo stocastico. Lo sviluppo di questo campo ha avuto forti impulsi e motivazioni dalla Fisica, in particolare dallo studio del moto browniano all'inizio del XX secolo da parte di Einstein, Smoluchowski e Langevin.
Negli anni '30 del XX secolo Kolmogorov ha iniziato la formalizzazione dei processi stocastici markoviani. Con questo termine si intendono, oltre alle CM con un numero finito di stati, anche le CM con un numero infinito numerabile di stati ed inoltre i processi con stati discreti e tempo continuo (descritti dalle cosiddette *master equations*) e quelli con stati continui e tempo continuo (regolati dalle equazioni di Fokker- Planck). Ora i processi stocastici trovano applicazioni nei campi più svariati: in Fisica, Chimica, Biologia nonché in Economia e Finanza.

Box A: La legge dei grandi numeri ed il teorema del limite centrale

Il calcolo delle probabilità esce dall'ambito frivolo (e limitato) dei giochi di carte e dadi, in cui era nato nel XVII secolo, con l'*Ars Conjectandi* di Jacob Bernoulli(1654-1705), pubblicata postuma nel 1713. In questo libro viene dimostrata la *legge dei grandi numeri*: dato un evento che accade con probabilità p, la frequenza $f(N)$ con cui questo evento accade in N prove indipendenti, nel limite di grandi N, "tende" a p:

$$f(N) \to p \ .$$

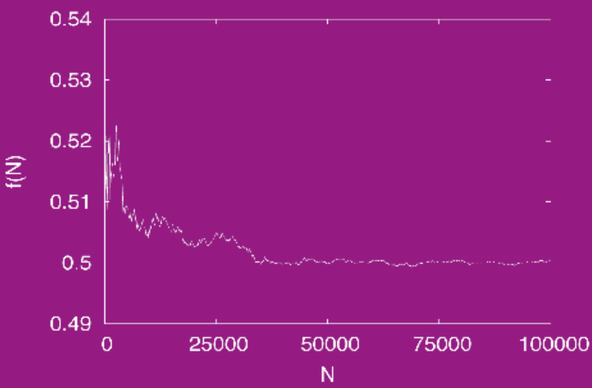


FIG. 2: UNA MONETA NON TRUCCATA È LANCIATA N VOLTE: È MOSTRATO L'ANDAMENTO DELLA FREQUENZA $f(N) = N^*/N$ DOVE N^* È IL NUMERO DI VOLTE CHE SI HA TESTA

In termini più precisi, per ogni $\epsilon > 0$ la probabilità che $f(N)$ si discosti più di ϵ da p diventa arbitrariamente piccola al crescere di N:

$$\lim_{N\to\infty} Prob\Big(\Big|\frac{f(N)}{N} - p\Big| > \epsilon\Big) = 0 \ . \qquad (1)$$

In modo analogo abbiamo che, data una successione di variabili indipendenti $x_1, x_2, \ldots, x_N$ la cui media è m e la varianza è finita, nel limite di N grande la probabilità che la "media empirica" $(x_1 + x_2 + \ldots + x_N)/N$ si discosti da m più di ϵ tende a zero per N tendente ad infinito:

$$\lim_{N\to\infty} Prob\Big(\Big|\frac{1}{N}\sum_{j=1}^{N} x_j - m\Big| > \epsilon\Big) = 0 \ . \qquad (2)$$

Il risultato (1) suggerisce un possibile modo di collegare il concetto di probabilità con il mondo reale: si può interpretare la probabilità di un evento come la sua frequenza nel limite di tante prove. Questa è l'essenza dell'interpretazione frequentistica che, in genere, è accettata dai fisici ma spesso non considerata appropriata nell'ambito delle scienze sociali ed economiche. Non è però questa la sede per una discussione delle interpretazioni della probabilità, tema interessante e delicato ma che appartiene più alla Filosofia della scienza che alla Matematica.

Nel 1716, nel libro *The Doctrine of Changes*, Abraham De Moivre (1667-1754) riuscì a dimostrare il primo caso di teorema del limite centrale: consideriamo N eventi indipendenti, ognuno dei quali avviene con probabilità p, indicando con $N^*(N)$ il numero di volte che l'evento accade; per N grande si ha:

$$Prob\Big(a \leq \frac{N^*(N) - pN}{\sqrt{p(1-p)N}} \leq b\Big) \simeq \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2} dx \ .$$

Questo risultato è un perfezionamento della legge dei grandi numeri. Infatti non solo possiamo dire che $N^*(N)/N$ è vicino a p, ma abbiamo anche la probabilità dello scostamento dal valor medio.

Come notato da Kac e Ulam, per qualche purista il risultato di De Moivre non sarebbe da considerare particolarmente importante in quanto si tratterebbe solo di un'applicazione piuttosto semplice di formule combinatorie elementari e dell'approssimazione di Stirling:

$$n! \simeq \sqrt{2\pi n} n^n e^{-n} .$$

Il risultato di De Moivre venne esteso da Laplace al caso di variabili indipendenti e discrete $\{x_i\}$ identicamente distribuite con media m e varianza σ^2; nel limite di N grande si ha:

$$Prob\Big(a \leq \frac{\sum_{j=1}^{N}(x_j - m)}{\sigma\sqrt{N}} \leq b\Big) \simeq \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2} dx \ . \quad (A.1)$$

La prima trattazione rigorosa del TLC per variabili identicamente distribuite, non necessariamente discrete e con varianza finita, è dovuta a Chebyshev, Markov e Lyapunov che, utilizzando il metodo delle funzioni caratteristiche e dei momenti, hanno dimostrato la (A.1) sotto ipotesi molto generali. La dimostrazione più generale del teorema del limite centrale per variabili indipendenti è dovuta al matematico svedese Lindeberg negli anni '20 del XX secolo.

Per variabili non indipendenti ci si aspetta che, se le $\{x_n\}$ sono "debolmente dipendenti", allora il TLC valga ancora se la varianza σ^2 viene sostituita con una varianza efficace σ^2_{eff} che tenga conto delle correlazioni. Nel caso di processi stazionari si dimostra che, se la funzione di correlazione:

$$c(k) = \langle (x_t - m)(x_{t+k} - m) \rangle$$

decade a zero abbastanza velocemente al crescere di k (cioè più rapidamente di $1/k$), allora il TLC vale, rimpiazzando σ^2 nella (A.1) con la varianza efficace σ^2_{eff}:

$$\sigma^2_{eff} = \sigma^2 + 2\sum_{k=1}^{\infty} c(k) \ .$$

Box B: La strana nascita delle catene di Markov

Davvero poco nota la storia della nascita delle catene di Markov: non sono state introdotte per ragioni tecniche inerenti alla teoria della probabilità e neppure per scopi applicativi nelle scienze o nella tecnologia, bensì per una diatriba su un argomento di natura filosofica, il libero arbitrio.
Tutto iniziò con un acceso contrasto tra Markov e Nekrasov, un matematico moscovita di opinioni conservatrici ed oscurantiste. La divergenza riguardava l'interpretazione della regolarità statistica dei comportamenti sociali.
Grazie all'attività di Quetelet, uno dei fondatori della Statistica moderna, nella dominante cultura positivistica ottocentesca si era sviluppata la cosiddetta Fisica sociale che interpretava le osservate regolarità nei fenomeni sociali su grande scala come qualcosa di analogo alle leggi fisiche: "*Se ci diamo la pena d'esaminare e mettere insieme osservazioni accurate e sufficientemente numerose, troveremo che ciò che credevamo fosse un effetto del caso è soggetto a principi stabili* (...). *Tutto è previsto, tutto è legge: solo la nostra ignoranza ci porta a supporre che tutto sia soggetto ai capricci del caso*". Nonostante qualche riserva, ad esempio Quetelet venne da alcuni accusato di fatalismo e di tentare di applicare in modo acritico ai fenomeni sociali il determinismo di Laplace, nella seconda metà del XIX secolo la Fisica sociale ebbe grande importanza in Sociologia e Filosofia, influenzando personaggi come Marx e Durkheim.
Nekrasov vedeva la Fisica sociale come l'anticamera del materialismo e, peggio ancora, dell'ateismo e del marxismo e non riconosceva alle regolarità statistiche dei fenomeni sociali lo *status* di vere leggi, in quanto, in accordo con la tradizione della religione ortodossa, i comportamenti umani sarebbero conseguenza del libero arbitrio.
Markov al contrario sosteneva che tali regolarità fossero sostanzialmente dovute alla legge dei grandi numeri. Nekrasov fece notare che la LGN non sarebbe stata sufficiente a spiegare la regolarità statistica, in quanto tale legge avrebbe avuto valore solo sotto l'ipotesi di eventi indipendenti. Nekrasov diede anche un (piccolo) contributo tecnico: mostrò che per la LGN è sufficiente l'indipendenza a coppie delle variabili $\{x_i\}$, condizione questa più debole della mutua indipendenza; tuttavia concluse erroneamente che l'indipendenza a coppie è anche condizione necessaria. Da questa sua erronea conclusione sostenne che la presenza di regolarità statistiche nei fenomeni sociali avrebbe dimostrato l'esistenza del libero arbitrio.
Nekrasov nelle sue oscure combinazioni di religione, Filosofia e Matematica aveva addirittura cercato di farsi forte dell'autorità dello scomparso Chebyshev. Per Andrei il furioso questo era davvero troppo: non solo un mediocre matematico sproloquiava di probabilità ma addirittura osava coinvolgere il suo grande maestro. Per annientare completamente il nemico Markov costruì una teoria di processi non indipendenti, introducendo quelle che ora sono chiamate CM, mostrando che si può avere la LGN anche senza l'indipendenza.
L'attrito tra Markov e Nekrasov era solo in parte dovuto alle loro divergenze personali, si inseriva in realtà in un più generale scontro accademico (e politico in senso lato) tra la scuola matematica moscovita e quella di San Pietroburgo, filogovernativa e religiosa la prima, progressista e materialista la seconda.
Nekrasov non ha lasciato una grande ereditità matematica, di fatto il suo unico contributo veramente importante è stato nell'aver provocato Markov, costringendolo ad introdurre le CM. Sorprendentemente Nekrasov non ebbe particolari problemi dopo la rivoluzione d'ottobre e riuscì a trovare un ragionevole compromesso con il nuovo governo. Addirittura alla sua morte nel 1924 il giornale ufficiale *Izvestia* pubblicò un necrologio elogiativo in cui veniva esaltato il suo contributo di scienziato al servizio della causa proletaria. Solo qualche anno dopo, nel 1933, nel mezzo delle grandi purghe staliniste, venne invece bollato come esempio di scienziato reazionario al servizio della borghesia e dell'oscurantismo della chiesa ortodossa.

Box C: Le catene di Markov in due parole

Un processo stocastico x_t che al tempo discreto t può assumere M possibili stati (che possiamo indicare con numeri interi 1, 2, …, M) è una catena di Markov se lo stato futuro dipende solo da quello presente, in formule:

$$Prob(x_t = i_t | x_{t-1} = i_{t-1}, ..., x_{t-n} = i_{t-n})$$
$$= Prob(x_t = i_t | x_{t-1} = i_{t-1})$$

dove i_t può valere 1, …, M. L'aspetto fondamentale delle CM è che la transizione allo stato $x_t = j$ condizionato al fatto che $x_{t-1} = i$, $x_{t-2} = k$, …, avviene con probabilità:

$$Prob(x_t = j | x_{t-1} = i) = P_{i \to j} = W_{ji}$$

indipendentemente dallo stato al tempo $t - 2$, quello al tempo $t - 3$, ecc.
Ci si potrebbe domandare perché non considerare anche casi più generali in cui il futuro al tempo t dipende dagli stati al tempo t-1, quello a t - 2, fino al tempo t - r. È facile convincersi che questo tipo di processi è riconducibile al caso r = 1; basta considerare una nuova catena in cui lo stato y_t è il vettore $(x_t, x_{t-1}, \ldots, x_{t-r+1})$.
Il caso più semplice è quello delle catene temporalmente omogenee, in cui le probabilità di transizione non dipen-

dono dal tempo t. Gli elementi di matrice W_{ji} non possono essere completamente arbitrari e devono soddisfare le seguenti relazioni:

$$W_{ij} \geq 0 \ , \ \sum_{i=1}^{M} W_{ij} = 1 \ .$$

La matrice di transizione $\{W_{ij}\}$ è, in qualche modo, il DNA della catena di Markov, infatti ne determina tutte le proprietà. Ad esempio per l'evoluzione temporale del vettore $\mathbf{P}(t) = (P_1(t), P_2(t), \ldots, P_M(t))$, dove $P_i(t)$ è la probabilità di essere al tempo t nello stato i, è immediato scrivere:

$$P_j(t+1) = \sum_{i=1}^{M} W_{ji} P_i(t) \qquad (C.1)$$

e quindi:

$$\mathbf{P}(t) = \hat{W}^t \mathbf{P}(0) \ .$$

È naturale domandarsi se al crescere di t, $\mathbf{P}(t)$ converge ad un vettore limite $\Pi = (\Pi_1, \Pi_2, \ldots, \Pi_M)$,. Se accade, allora:

$$\Pi_j = \sum_{i=1}^{M} W_{ji} \Pi_i \ , \qquad (C.2)$$

e le $\{\Pi_i\}$ sono chiamate probabilità invarianti (o stazionarie). Vale il seguente fondamentale teorema:

Se esiste un intero n tale che per ogni coppia (i, j) si ha una probabilità non nulla di andare da j a i in n passi, cioè:

$$[W^n]_{ji} > 0 \ ,$$

allora esiste un'unica probabilità invariante Π *e la convergenza è esponenzialmente veloce:*

$$\mathbf{P}(t) = \hat{W}^t \mathbf{P}(0) = \mathbf{\Pi} + O(e^{-t/\tau_c}) \ \rightarrow \ \mathbf{\Pi} \ .$$

Il tempo caratteristico τ_c è:

$$\tau_c = \frac{1}{|\ln(|\alpha_2|)|} \ ,$$

dove α_2 è il secondo autovalore della matrice $\hat{W}$. In questo tipo di matrice, per un teorema di Algebra lineare dovuto a Perron e Frobenius, il primo autovalore non è mai degenere: $\alpha_1 = 1 > |\alpha_2|$, quindi $|\alpha_2| < 1$ e τ_c è finito.
Tali CM sono ergodiche, cioè la media temporale lungo un *random walk* effettuato con le probabilità di transizione $\{P_{i \rightarrow j}\}$ è asintoticamente uguale alla media con le probabilità $\{\Pi_i\}$.

Box D: Due applicazioni delle catene di Markov

Le catene di Markov hanno un ruolo fondamentale in molte applicazioni in Fisica, Astrofisica, Chimica, Biomatematica, Genetica, Geofisica, Ingegneria e comunicazioni. Tra i tanti, discuteremo brevemente due esempi molto importanti di utilizzo delle CM.

Il metodo Monte Carlo
Consideriamo il seguente problema: calcolare la media

$$\langle Q \rangle = \frac{1}{M} \sum_{j=1}^{M} Q_j \Pi_j \qquad (D.1)$$

dove Π_j è la probabilità dell'evento j; se M è molto grande, il calcolo può essere proibitivo (ad esempio, in Meccanica statistica il valore di M è enorme anche in sistemi piccoli). Per dare un'idea consideriamo il modello di Ising sul reticolo (una descrizione molto semplificata del magnetismo) in cui per ogni sito si hanno due possibilità: già con solo 100 siti il numero di stati permessi è enorme, $M = 2^{100} \simeq 10^{30}$. L'idea del metodo Monte Carlo è di sfruttare l'ergodicità e quindi rimpiazzare la (D.1) con una media temporale ottenuta seguendo un "viaggiatore" che salta tra i vari stati seguendo le regole di una catena di Markov che ha probabilità invarianti $\{\Pi_j\}$ e calcolare la media temporale:

$$\frac{1}{T} \sum_{t=1}^{T} Q_{j_t}$$

dove $j_1, j_2, \ldots, j_T$ indica la successione delle posizioni del viaggiatore. Nel caso la catena di Markov sia ergodica (condizione che si può accertare dalle probabilità di transizioni $\{P_{i \rightarrow j}\}$) allora per T grande si ha che la media temporale "tende" a $\langle Q \rangle$ (nel senso dell'equazione (2)).
Notiamo che nel metodo Monte Carlo c'è un'ampia arbitrarietà; infatti non ci sono vincoli particolari per la scelta delle probabilità di transizione purché la catena di Markov abbia come probabilità invariati le $\{\Pi_j\}$ e sia ergodica. Il metodo Monte Carlo si è rivelato estremamente potente ed utile; ad esempio, permette di determinare le proprietà termodinamiche di sistemi non banali (come i liquidi) a partire dalla conoscenza delle loro interazioni microscopiche (il potenziale tra coppie di molecole).
Ovviamente, nel caso il numero dei possibili stati sia molto grande (come accade praticamente sempre nei problemi di Meccanica statistica), la traiettoria non può essere tanto lunga da visitarli tutti. Ci si potrebbe quindi domandare quale sia il segreto dell'efficacia del metodo. In poche parole possiamo dire che il suo punto di forza è nel fatto che tipicamente vengono calcolate medie di quantità "non troppo strane" ed inoltre che il viaggiatore non perde tempo ad avventurarsi in stati con probabilità troppo bassa.

Google e le catene di Markov
Ogni volta che utilizziamo Google (od altri motori di ricerca su Internet), senza saperlo usiamo le catene di Markov. Quando

si inseriscono delle parole su Google (ad esempio *Markov chains applications*), il motore di ricerca, utilizzando un algoritmo basato sulle catene di Markov, fornisce un elenco, in ordine di importanza, delle pagine che contengono le parole desiderate.

Ecco un cenno alla procedura usata per decidere l'ordine:

a) si individuano le N pagine che contengono le parole e ad ogni pagina è associato un numero 1, 2, ..., N;
b) si determina il numero di *link* L_k, cioè il numero di pagine che puntano sulla k-ma pagina;
c) a partire da N e dal numero dei *link* $\{L_j\}$, con opportune regole si costruisce la matrice delle probabilità di transizione $\{P_i \to j\}$, in modo tale da avere una catena di Markov ergodica;
d) si calcolano le probabilità invarianti $\prod_1, \prod_2, \ldots, \prod_N$;
e) si crea il *ranking*: la prima pagina è quella con la probabilità più alta e così via.

La matrice di transizione è così determinata: se esiste un *link* dalla pagina i alla pagina j si ha:

$$P_{j\to i} = \frac{\alpha}{L_j} + \frac{(1-\alpha)}{N} \quad ,$$

se invece non esiste allora:

$$P_{j\to i} = \frac{(1-\alpha)}{N}$$

dove il valore di α è compreso tra 0 ed 1 (tipicamente si usa $\alpha = 0.85$).
Questa scelta ha un'interpretazione semplice ed intuitiva: per una percentuale di tempo $1-\alpha$ un (ipotetico) navigatore nella rete partendo dalla pagina j pesca un sito a caso tra gli N possibili, oppure con probabilità α salta sul sito i in base alle connessioni. Il motivo tecnico per usare un valore $\alpha \neq 0$ è che in questo modo la CM è sicuramente ergodica.
Le probabilità invarianti sono determinate da un sistema di equazioni lineari:

$$\Pi_i = \sum_{j=1}^{N} \Pi_j P_{j\to i} \quad i = 1, 2, ..., N.$$

Naturalmente, se N è molto grande, non è possibile trovare la soluzione esatta e in genere si ricorre ad un metodo iterativo. Partendo ad esempio da $P_i(0) = 1/N$ si usa ripetutamente la (C.1):

$$P_i(n+1) = \sum_{j=1}^{N} P_j(n) P_{j\to i}\,.$$

Poiché la CM è ergodica, al crescere di n si ha la convergenza (in modo esponenzialmente veloce) alle probabilità invarianti. Alternativamente si può usare il metodo Monte Carlo: per ogni i la frequenza $f_i(T)$ di visita dello stato i su un lungo intervallo temporale $[0, T]$ sarà vicina a $\prod_i$.
Per concludere è opportuno precisare che il metodo precedentemente discusso è quello "asettico" (di base), nella realtà Google usa un algoritmo personalizzato diverso da utente ad utente che privilegia le pagine ed i siti visitati in passato. ■

Bibliografia

Sulla vita e l'opera scientifica di Markov:
Langville A.N. e Stewart W.J. (eds.), *Markov Anniversary Meeting*, Boson Books, 2006.
Basharin G.P., Langville A.N. e Naumov V.A., "The life and work of A.A. Markov", *Linear Algebra and its Applications* 386, 3 (2004).
Seneta E., "Markov and the Birth of Chain Dependence Theory", *International Statistical Review* 64, 255 (1966).
Hayes B., "First Links in the Markov Chain", *American Scientist* 101, 92 (2013).
Sheynin O.B., "A.A. Markov's Work on Probability", *Arch. Hist. Exact Sci.* 39, 337 (1988).

Su Nekrasov e la Statistica:
Seneta E., "Statistical regularity and free will: L.A.J. Quetelet and P.A. Nekrasov", *International Statistical Review* 71, 319 (2003).

Sulla Statistica e la probabiltà nelle scienze sociali:
Addeo F., *È normale la curva normale?*, Bonanno Editore, 2008.

Per una presentazione introduttiva della probabilità e la sua connessione con la teoria della misura:
Kac M. e Ulam S., *Mathematics and Logic*, Dover Publications, 1992.

Per un'introduzione alle varie interpretazioni della probabilità:
Hájek A., "Interpretations of Probability", in *Stanford Encyclopedia of Phylosophy*, consultabile su:
http://www.seop.leeds.ac.uk/entries/probability-interpret.

Due buoni libri sui processi di Markov che discutono con particolare attenzione le applicazioni in Fisica, Chimica e Biologia:
Haggström O., *Finite Markov Chains and Algorithmic Applications*, Cambridge University Press, 2002.
Bharucha-Reid A.T., *Elements of the Theory of Markov Processes and Their Applications*, Dover Publications, 2010.

Per l'uso delle catene di Markov in Linguistica:
Khmelev D.V. e Tweedie F.J., "Using Markov chains for identification of writers", *Literary an Linguistic Computing* 16, 299 (2001).
Khmelev D.V., "Disputed authorship resolution through using relative empirical entropy for Markov chains of letters in human language texts", *Journal of Quantitative Linguistics* 7, 201 (2000).

Per una rassegna, in parte anche storica, sui processi stocastici in Fisica (a partire dal moto browniano):
Chandrasekhar S., "Stochastic problems in physics and astronomy", *Rev. Mod. Phys.* 15, 1 (1943).

Un interessante libro sull'influenza (a volte insospettata) della religione sullo sviluppo della Matematica:
Graham L. e Kantor J.-M., *Naming Infinity*, Harvard University Press, 2009.

Sulla storia del teorema del limite centrale:
Adams W.J., *The life and times of the central limit theorem*, American Mathematical Society, 2008.
Fischer H., *A History of the Central Limit Theorem*, Springer Verlag, Berlin, 2010.