

Fare calcoli con il caso: introduzione alla tecnica Monte Carlo

Franco Bagnoli



Dipartimento di Energetica e Centro per lo Studio di Dinamiche Complesse (CSDC)

Università di Firenze, Via S. Marta 3 50139 Firenze

franco.bagnoli@unifi.it <http://francobagnoli.complexworld.net>

Lezione tenuta a festivaletteratura

Mantova 8/9/2010 <http://www.festivaletteratura.it/>

Sommario

Questa lezione è composta di due parti. Nella prima, cercherò di spiegare come si può modellizzare un sistema fisico in modo da poterlo simulare su un computer, e come si “rappresenta” un processo di misura su tale sistema. Illustrerò poi il principale risultato della meccanica statistica: nel caso di sistemi all’equilibrio è possibile “separare” il problema in tanti sottoproblemi più semplici. Purtroppo però, nei casi reali resta sempre almeno un sottoproblema “difficile”, le cui proprietà vanno derivate per mezzo del calcolatore. La meccanica statistica ci assicura però che le proprietà di equilibrio di un sistema sono indipendenti dal tipo di dinamica, per cui possiamo trarre profitto dai vantaggi di una traiettoria “aleatoria” per effettuare i calcoli.

Nella seconda parte darò un esempio di come si possono effettuare queste misure con una dinamica casuale, calcolando l’area di un cerchio tirando numeri a caso.

Un avvertimento: non è la stessa cosa fare i calcoli *con il caso* o fare i calcoli *a caso*.

Dalla Dinamica alla statistica e ritorno

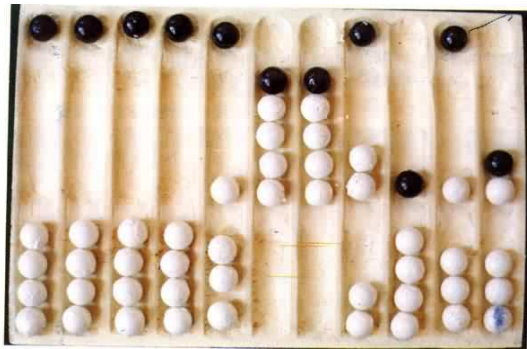
Una misura effettuata su un sistema fisico comporta necessariamente una operazione di *media*: né i nostri sensi né gli apparati sperimentali sono capaci di tenere traccia delle innumerevoli interazioni con gli atomi, le molecole e gli altri componenti microscopici della realtà. Se vogliamo ottenere per via teorica o computazionale il valore di una misura, bisogna essere capaci di effettuare tale operazione di media sui movimenti dei componenti microscopici.

In questa prima parte si cercherà di illustrare come si può “simulare” un sistema fisico su un computer (movimento delle particelle ed operazione di media), e come questo procedimento sia pesante anche per i moderni computer, figurarsi quando i calcolatori erano al più azionati a manovella. Ricordiamo che il tipico numero di componenti per un sistema macroscopico è il numero di Avogadro, 10^{23} . Per fortuna, in alcuni casi piuttosto comuni, si può supporre che il sistema “esplori” tutto lo spazio a disposizione, senza altra preferenza che (eventualmente) il valore dell’energia. Per esempio, le molecole di un liquido in un bicchiere esplorano lo spazio a disposizione (magari prendendo il tempo necessario): per convincersene si può lasciar cadere una goccia di inchiostro. Se il colorante è per esempio più leggero dell’acqua (si può usare del vino) l’energia gravitazionale concentra il colorante in alto, ma non gli fa preferire un lato piuttosto che un altro.

In questa maniera, si può sostituire alla simulazione della *dinamica*, la ricerca della *distribuzione di probabilità* corrispondente ad una osservazione su tempi lunghi. Questo approccio ha una serie di vantaggi matematici, in particolare quello che rende possibile *separare* gruppi di variabili, spezzando il problema originario in una serie di problemi più semplici. Resta però in genere un *nucleo duro*, un problema che non si può affrontare con carta e penna ma che va simulato al calcolatore.

A questo punto si fa una specie di “mossa del cavallo”, e si torna indietro dalla statistica (la distribuzione di probabilità) alla dinamica. Si cerca cioè una dinamica *che coinvolga solo le variabili del problema “duro”*, che porti alla stessa distribuzione di probabilità del problema originario. Nel fare ciò, cerchiamo anche il metodo per accelerare i calcoli, ovvero cerchiamo una dinamica che esplori più rapidamente possibile lo spazio a disposizione. In genere, si usa una dinamica *stocastica*, che utilizza *numeri casuali*: il metodo Monte Carlo.

Ricordiamo lo scopo di fondo di questa simulazione: calcolare delle medie, ovvero delle somme che in termini matematici, si chiamano *integrali* (ovvero delle aree, volumi, ecc.). Un esempio, il calcolo Monte Carlo dell’area di un cerchio, verrà illustrato nel secondo capitolo.



a



b

Figura 1.1: a) Un pallottoliera (abbaco), già utilizzato millenni fa dai babilonesi, egiziani e cinesi; b) Un pallottoliera russo. Notare l'incurvatura delle stanghette metalliche, per stabilizzare le configurazioni.

1.1 Introduzione

Calculus, in latino, era il nome del ciottolo. Come le fiche della roulette, i ciottoli si possono usare come gettoni per tener conto dell'andamento di un gioco: a tutti sarà capitato di giocare a sette e mezzo (o a poker) usando dei sassolini al posto delle monete. E ovviamente si possono usare i ciottoli per fare dei calcoli, dalla semplice corrispondenza 1 ciottolo = 1 oggetto (per esempio per contare le pecore di un gregge o le preghiere con un rosario), a corrispondenze più sofisticate come in un pallottoliera, vedi la figura 1.1-a. Lo scopo di usare ciottoli o altro come ausilio del calcolo è quello di "meccanizzare" il calcolo stesso, evitando di dover tenere a memoria risultati parziali, o di dover eseguire operazioni troppo difficili.

A pensarci bene, il grosso problema dell'utilizzo di strumenti ausiliari come le pietruzze è dato dal caso: una scossa al tavolo su cui stanno i sassi ben separati a seconda del loro significato (i soldi miei contro a quelli tuoi, oppure i sassi simboleggianti le centinaia contro quelli che simboleggiano le decine) può rovinare tutto il calcolo. In fondo il pallottoliera altro non è che un sistema per tenere ordine tra ciottoli usati per eseguire dei calcoli, e renderli insensibili a piccole scosse.

I calcolatori altro non sono che pallottolieri sofisticati, con elettroni al posto dei ciottoli, e anche in questo caso il caso (rumore termico) può rovinare tutto. Non a caso i transistor in un computer non vengono fatti lavorare nel regime lineare, dove ogni perturbazione verrebbe amplificata, ma in regime di saturazione, o binario. Un po' come nei pallottolieri russi (figura 1.1-b) dove le stanghette metalliche sono piegate per renderlo più insensibile ai disturbi.

È quindi ironico che nel 1945 Stanislaw Ulam e John von Neumann [1, 2], nell'ambito del segretissimo progetto Manhattan, abbiano pensato di utilizzare i primi computer per effettuare calcoli utilizzando numeri casuali, che a loro volta non sono facili da generare in un computer. L'idea è quella di realizzare un *processo stocastico* (già studiato ampiamente in maniera teorica nei primi anni del '900, da Einstein e Kolmogorov in particolare), ovvero un processo determinato dal caso, per *campionare* uno spazio e in definitiva calcolare un'area, ovvero in termini tecnici fare un *integrale*. Un esempio classico, che svilupperemo in seguito, è quello di calcolare l'area di un cerchio disegnato sul suolo contando quanti sassi gettati casualmente cadono

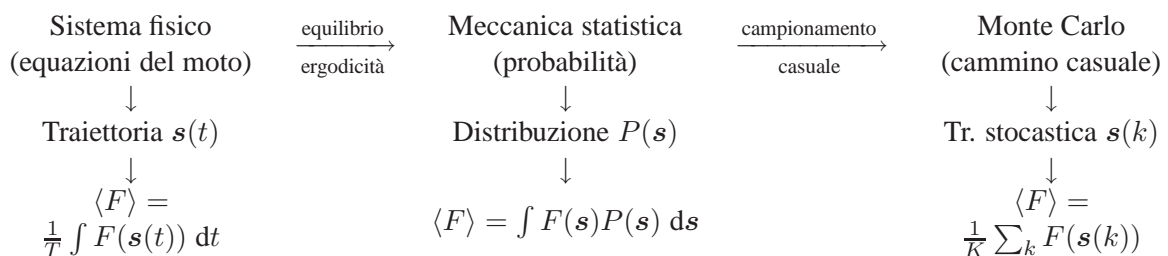


Figura 1.2: Il casinò di Monte Carlo.

nel suo interno. L'uso di tecniche casuali (lancio di oggetti) per effettuare calcoli era già stata utilizzata sperimentalmente alla fine dell'800, ma ovviamente non erano mai stati utilizzati dei computer, dato che non ancora esistevano.

Il nome *Monte Carlo* fu suggerito da von Neumann pensando alla roulette del famoso casinò, avendo saputo da Ulam che un suo zio era solito chiedere in prestito soldi con la scusa "devo andare a Montbla bla e Carlo".

Ma se è abbastanza intuitivo capire come si possano calcolare delle aree utilizzando dei sassi gettati in maniera casuale, non si apprezza a fondo il metodo Monte Carlo se non ci si rende conto che esiste un cammino logico che parte dalla *dinamica* di un sistema fisico qualsiasi e porta, attraverso la *meccanica statistica*, al calcolo di integrali complicati. A questo punto si inserisce il calcolo Monte Carlo, che riconduce il calcolo di integrali a misure effettuate su una traiettoria stocastica, ovvero su una *dinamica* fittizia. Lo schema di base è il seguente:



Esistono notevoli vantaggi computazionali nell'utilizzare il metodo Monte Carlo rispetto alla simulazione diretta di un sistema fisico, e non a caso solo nel 1953 Fermi, Pasta e Ulam provarono a utilizzare i calcolatori, ormai non più utilizzati per i calcoli della bomba atomica, per simulare un semplice sistema fisico [3], scoprendo tra l'altro che le loro ipotesi sulle condizioni per l'applicazione della meccanica statistica (su cui si basa il calcolo Monte Carlo) non erano per niente verificate...

1.2 Dalla dinamica alla statistica

Il succo della prima parte della lezione è contenuto nella seguente serie di uguaglianze:

$$\langle F \rangle = \frac{1}{\tau} \int_0^\tau F(\mathbf{s}(t)) dt = \int F(\mathbf{s})P(\mathbf{s}) d\mathbf{s} = \frac{1}{K} \sum_k F(\mathbf{s}(k)).$$

Con il primo termine, $\langle F \rangle$, indico l'operazione di media su una grandezza macroscopica (detta osservabile) F , per esempio la misura della pressione di un gas. Infatti, le misure che facciamo su un sistema fisico, sono in effetti delle medie effettuate su tante interazioni microscopiche. Nel caso della pressione, quello che misuriamo è il risultato di un numero molto grande di collisioni microscopiche.

In molti casi, la fisica di Newton è sufficiente per descrivere, o *modellizzare* il sistema sotto esame, nel senso che conosciamo sia come rappresentare i componenti microscopici, per esempio le molecole, sia le forze che si stabiliscono tra questi. Con l'avvento del computer, è possibile calcolare l'evoluzione temporale del sistema. Dato che i sistemi sono vari, bisogna usare notazioni diverse quando si tratta di un gas, di un cristallo o di un altro sistema. Però, in ogni caso, lo *stato* del sistema è dato da una lista di numeri, per esempio dalle posizioni e velocità delle varie molecole. Indico questa lista di numeri con il simbolo \mathbf{s} . Le leggi di Newton (vedi approfondimento 4.1) ci dicono che è possibile calcolare lo stato futuro del sistema conoscendo lo stato attuale e la forma delle forze. Ovvero, possiamo calcolare la *traiettoria* dell'intero sistema, che indico con $\mathbf{s}(t)$. Si noti che non sto parlando della traiettoria di una determinata molecola, ma dell'insieme di tutte le traiettorie di tutte le molecole. Per esempio, se ho 1000 molecole, lo stato \mathbf{s} del sistema è dato da una lista di 6000 numeri, che corrispondono alla posizione spaziale (tre numeri) e alla velocità (altri tre numeri) per ogni molecola. Il tipico numero di molecole di un gas "reale" è il numero di Avogadro, 10^{23} . Nessun calcolatore riesce a calcolare la traiettoria di un tale sistema in un tempo "ragionevole". Si può dire che con un supercomputer si riesce a calcolare la traiettoria di un milione di atomi (10^6) per un tempo di qualche centinaio di nanosecondi, mettendoci magari una settimana di tempo "reale". Il problema di fondo è che il computer, per non fare errori troppo grossi, deve avanzare con passi temporali molto piccoli (approfondimento 4.1).

Per un calcolatore non è difficile, data la posizione istantanea $\mathbf{s}(t)$, calcolare il valore istantaneo dell'osservabile $F(\mathbf{s}(t))$, per esempio, la forza agente su una parete, da cui si ottiene la pressione. Però, questa sarà una quantità estremamente variabile nel tempo, dato che si tratta di tante collisioni ognuna delle quali avviene in un tempo molto corto. Se il nostro strumento di misura fosse un microfono molto sensibile, quello che sentiremmo è un rumore (tanti urti) molto fluttuante. Per gli usi pratici però siamo più interessati al suo valore medio: gli strumenti di misura di solito effettuano questa operazione di media *temporale* sia per come sono costruiti (per esempio perché hanno parti con grande inerzia che non possono seguire le fluttuazioni molecolari), sia eventualmente in maniera numerica. Simboleggiamo l'operazione di media con la prima equazione,

$$\langle F \rangle = \frac{1}{\tau} \int_0^\tau F(\mathbf{s}(t)) dt,$$

dove con il simbolo τ abbiamo indicato il tempo "caratteristico" dello strumento. Potrebbe essere per esempio dell'ordine del microsecondo o del decimo di secondo.

Il calcolo della traiettoria $\mathbf{s}(t)$ è però una operazione molto impegnativa anche per un computer, figuriamoci per un umano (anche se molti astronomi hanno calcolato le traiettorie dei pianeti a mano). Dato che solo da poco tempo esistono i calcolatori, i ricercatori si sono

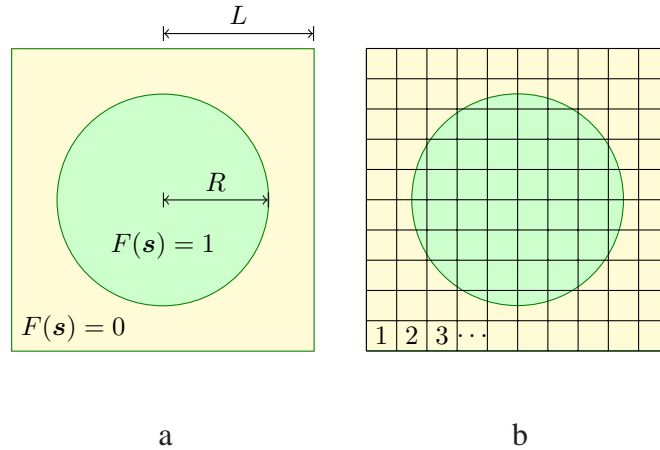


Figura 1.3: (a) Dominio di integrazione, la funzione $F(\mathbf{s})$ vale 1 nel cerchio e 0 al di fuori di questo; (b) partizione del dominio.

ingegnati per scoprire dei metodi che permettano di ottenere il valore delle osservabili dei modelli senza dover seguire la traiettoria del sistema.

Per rendere il discorso più intuitivo, pensiamo di poter rappresentare lo stato \mathbf{s} del sistema come un punto in un quadrato di lato $2L$, in due dimensioni. Il quadrato rappresenta il *dominio* di variabilità di \mathbf{s} . Si tenga però presente che nel caso reale abbiamo uno spazio con tante variabili, e che alcune di queste come le velocità possono in principio variare fino all'infinito (o meglio fino alla velocità della luce...).

Sempre per fare le cose semplici, supponiamo che la funzione $F(\mathbf{s})$ prenda il valore 1 all'interno di un cerchio di raggio R , centrato nel quadrato, e zero al di fuori, come mostrato in figura 1.3-a. Se la traiettoria passa molto tempo vicino al centro del quadrato, il valore di $\langle F \rangle$ sarà vicino ad 1, se invece la traiettoria starà molto tempo vicino ai bordi il valore di $\langle F \rangle$ sarà quasi zero.

Adesso dividiamo il dominio in quadretti \mathbf{s}_i , numerandoli con l'indice $i = 1, 2, \dots, M$, come mostrato in figura 1.3-b. Se il tempo di misura τ è molto più lungo del tempo di cui in media il sistema ha bisogno per "esplorare" il dominio, allora la frazione di tempo Δt_i che il sistema passa in ogni quadretto non dipenderà dalle condizioni iniziali. Possiamo quindi definire la distribuzione di probabilità $P(\mathbf{s}_i)$ che mi dà la probabilità che prendendo il sistema in un istante a caso, lo trovi nel quadretto i , ovvero

$$P(\mathbf{s}_i) = \frac{\Delta t_i}{\tau}.$$

Ma allora, invece di calcolare delle quantità sulla traiettoria, posso calcolare il valore delle osservabili sul dominio, facendo una somma sui vari quadretti,

$$\langle F \rangle = \frac{1}{\tau} \int_0^\tau F(\mathbf{s}(t)) dt = \sum_i F(\mathbf{s}_i) P(\mathbf{s}_i),$$

e nel limite in cui i quadretti diventano molto piccoli, eseguendo un integrale

$$\langle F \rangle = \frac{1}{\tau} \int_0^\tau F(\mathbf{s}(t)) dt = \int F(\mathbf{s}) P(\mathbf{s}) d\mathbf{s}.$$

In molti casi risulta molto più facile determinare la distribuzione $P(\mathbf{s})$ invece del calcolo della traiettoria $\mathbf{s}(t)$.

1.3 La distribuzione di probabilità

Ma come trovare la distribuzione $P(\mathbf{s})$? L'intuizione di Boltzmann e Gibbs, all'inizio del secolo, fu quella di presupporre che, all'equilibrio e in assenza di particolari disposizioni, la *tipica* traiettoria di un sistema chiuso trascorra in ogni porzione di spazio un tempo *proporzionale* all'area della porzione stessa, ovvero, $P(\mathbf{s}) = \text{costante}$. Come illustrato nell'approfondimento 4.2, questa ipotesi è quella "più plausibile", in assenza di dettagliate informazioni sul sistema.

Sempre sulla base della stessa impostazione di "plausibilità", risulta che invece la distribuzione di probabilità di un sistema che nello stato \mathbf{s} abbia energia $E(\mathbf{s})$, in contatto con un serbatoio di calore a temperatura T sia

$$P(\mathbf{s}) = \frac{\exp(-E(\mathbf{s})/T)}{Z},$$

Dove $Z = \int \exp(-E(\mathbf{s})/T) d\mathbf{s}$ è una costante necessaria a far sì che la probabilità abbia integrale uguale ad 1. Si noti che nel caso in cui la temperatura T è molto alta, anche il sistema con energia variabile arriva ad avere una distribuzione $P(x)$ "piatta": quando la temperatura è alta la traiettoria salta da tutte le parti con la stessa probabilità.

Il grosso vantaggio della formulazione di Boltzmann e Gibbs è matematico (si veda l'approfondimento 4.3). Se l'insieme delle variabili che abbiamo indicato con \mathbf{s} su può spezzare in due, per esempio le posizioni che indichiamo con \mathbf{x} e le velocità che indichiamo con \mathbf{v} , in maniera tale che l'energia $E(\mathbf{s})$ sia data dalla somma di una parte che dipende solo dalle \mathbf{x} e da una che dipende solo dalle \mathbf{v} ,

$$E(\mathbf{s}) = E_x(\mathbf{x}) + E_v(\mathbf{v}),$$

e quindi si ottiene che la distribuzione di probabilità si fattorizza,

$$P(\mathbf{s}) = P_x(\mathbf{x})P_v(\mathbf{v}).$$

Praticamente, il sistema si spezza in due problemi separati, legati solo dal fatto di avere la stessa temperatura. E, quando va bene, si può ripetere il processo più volte, fino ad avere tanti problemi semplici che si possono affrontare con carta e penna. Non sempre siamo così fortunati, di solito si riesce a separare un gran numero di variabili, ma resta comunque un problema "difficile", che va affrontato con il calcolatore. E' in ogni caso un bel risparmio.

Per capire come fare, cominciamo a esaminare il caso più semplice: quello "piatto". Abbiamo ridotto il nostro problema al calcolo di un'area. Studiamo per esempio il problema del calcolo dell'area del cerchio.

Indipendentemente (o quasi) dalle motivazioni fisiche, vediamo come si calcola un integrale (ovvero: un'area) utilizzando dei numeri casuali.

2.1 Integrali Monte-Carlo

Vogliamo calcolare l'area di un cerchio. Ovviamente non possiamo usare la formula che ci hanno insegnato a scuola, dato che vogliamo imparare una tecnica che ci permetta di calcolare aree arbitrarie, e anche volumi in molte dimensioni. La tecnica standard di tagliare il dominio in quadretti (o cubetti, o iper-cubetti) e contare quanti cadono dentro alla nostra figura è un procedimento che non si può applicare sempre. Se per esempio il dominio è molto esteso (come nel caso della variabile velocità) il procedimento rischia di durare tanto tempo e fino alla fine non abbiamo una idea di quanto valga questa area.

Possiamo rifarci a Galileo. Dato che non era molto ferrato in matematica, quello che faceva quando doveva calcolare un'area era di disegnarla su un cartoncino. Quindi pesava il cartoncino, ritagliava la figura in questione e quindi pesava quest'ultima. Il rapporto tra peso della figura e peso del dominio è uguale al rapporto tra le aree, *sempreché il cartoncino sia omogeneo*.

Questa non è una tecnica che si possa facilmente implementare su un computer. Però possiamo variarla in questa maniera: disegniamo la figura ed il dominio sul terreno, e lanciamo sopra di questo un gran numero di sassi, a caso. Dobbiamo controllare che la distribuzione dei sassi sia omogenea, il che equivale a controllare che lo spessore del cartoncino sia uniforme. Se lo è, allora basta contare quanti sassi sono caduti nella figura rispetto a quanti sono caduti nel dominio, ed abbiamo una stima del valore dell'area della figura, come si vede in figura 2.1.

Ovviamente, tanti più sassi tiriamo, tanto migliore sarà la stima, ma anche con pochi sassi possiamo avere un'idea grossolana di quanto vale il nostro integrale. La convergenza del valore dell'integrale al valore asintotico va come la radice quadrata del numero K di sassi, è quindi un metodo lento.

Purtroppo, questa soluzione è abbastanza inefficiente. In genere non possiamo semplicemente estrarre punti a caso nel dominio. Che fare per esempio se il dominio va all'infinito? Esiste però un'altra tecnica. Possiamo costruire delle traiettorie casuali, denominate *cammini a caso* o *random walk*, e usare queste traiettorie per "popolare" i nostri quadretti.

N= 1000 , pi= 3.1480000

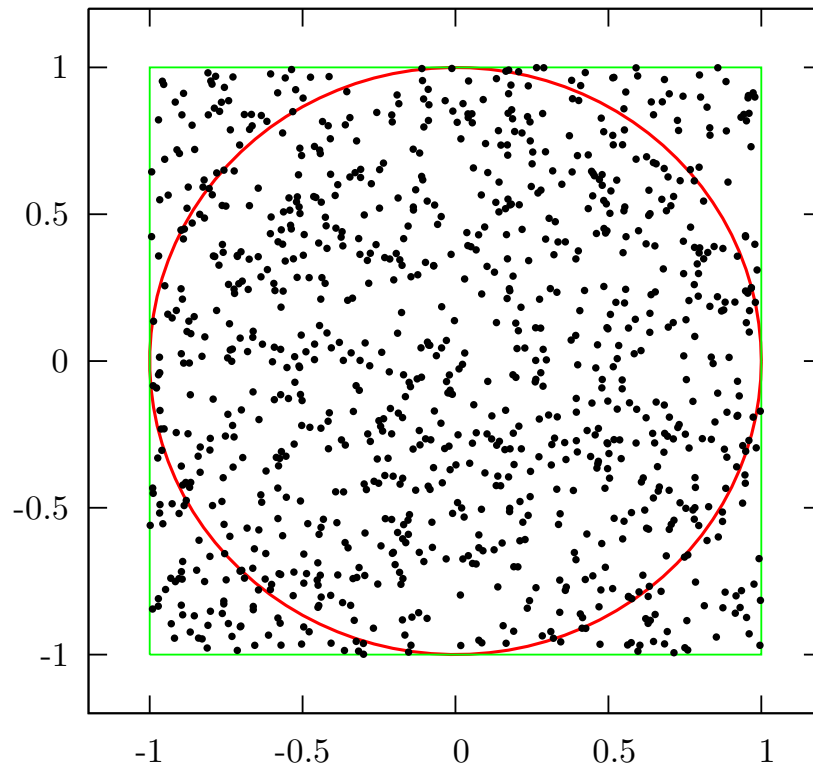


Figura 2.1: a) Campionamento diretto, 1000 punti.

2.2 Traiettorie stocastiche

L'idea è quella di trovare una dinamica *arbitraria*, che sia possibilmente molto facile e veloce da calcolare per un computer e che dia una traiettoria fittizia $\mathbf{s}(k)$ (sul "tempo" fittizio k), che però abbia la *stessa distribuzione di probabilità* del problema originale. Inoltre si desidera che questa traiettoria esplori rapidamente lo spazio a disposizione, così che dia subito delle stime del valore finale. Ovvero, vogliamo una procedura che mi permetta di ottenere rapidamente una sequenza $\mathbf{s}(1), \mathbf{s}(2), \dots, \mathbf{s}(k), \dots, \mathbf{s}(K)$ di stati o "fotografie" del sistema, su cui calcolare la media del valore dell'osservabile F ,

$$\langle F \rangle = \frac{1}{\tau} \int_0^\tau F(\mathbf{s}(t)) dt = \int F(\mathbf{s}) P(\mathbf{s}) d\mathbf{s} = \frac{1}{K} \sum_k F(\mathbf{s}(k)).$$

In fondo, se riuscissimo a fare delle fotografie sul sistema reale (e su alcuni sistemi si riesce a fare qualcosa di analogo), avremmo che per intervalli temporali piccoli, le fotografie cambierebbero di poco l'una dall'altra (e quindi darebbero poca informazione sul valore medio di una osservabile), mentre se le foto sono distanziate più del tempo "caratteristico" τ , avremmo una collezione di immagini ben assortita. Se possibile, vorremmo ottenere questa collezione nel tempo più corto possibile, adottando una dinamica che "saltelli" nello spazio appropriato.

Supponiamo che il dominio sia adesso molto grande, per esempio una pista circolare di atterraggio di un eliporto, delimitato da una recinzione quadrata. Invece di tirare dei sassi a caso nel dominio (sarebbe difficile avere una distribuzione uniforme), partiamo dal centro del cerchio e tiriamo un sasso a caso, in una direzione arbitraria. Raggiungiamo il sasso, e ne

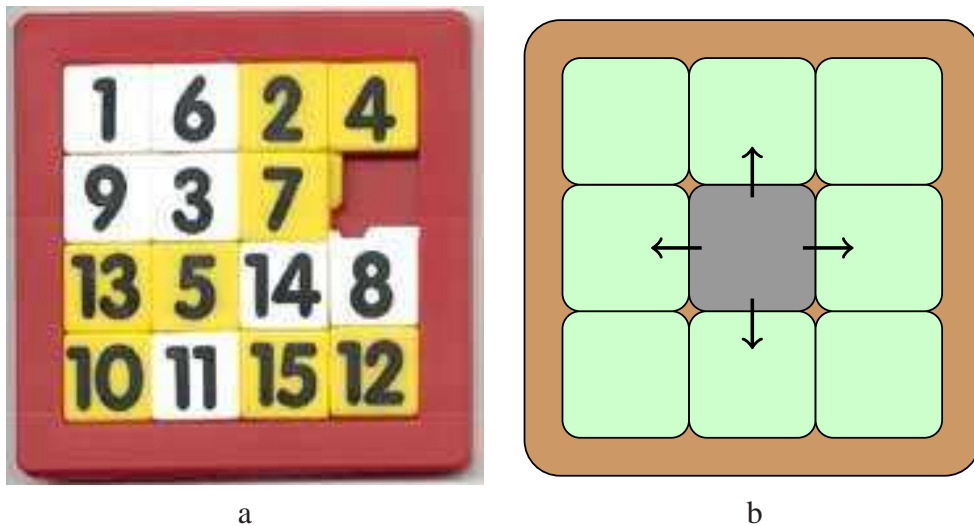


Figura 2.2: (a) Il gioco del 15; (b) il gioco del 9 con i movimenti possibili della tessera grigia nella posizione centrale.

tiriamo a caso un altro, sempre in una direzione arbitraria. Se per un momento supponiamo che il dominio coincida con la superficie di una ciambella, ovvero che uscendo da un lato della recinzione si rientri dall'altro, si vede come una tale prassi finirà per coprire il dominio di uno strato omogeneo di sassi.

Ma non viviamo su una ciambella. Prima o poi un sasso uscirà dal dominio, ovvero andrà a cadere fuori dalla recinzione. Che fare in questo caso? ci sono tre possibilità:

1. Si fa finta di nulla e si getta un altro sasso, sperando che cada nel dominio.
2. Si torna indietro all'ultimo sasso e se ne getta un altro.
3. Si fa qualcosa di diverso.

Ovviamente la risposta giusta è la numero tre. Però, per capire cosa bisogna fare, dobbiamo ancora una volta cambiare modello.

Prima di procedere, cerchiamo di enucleare l'idea che stiamo sviluppando. Vogliamo generare una distribuzione di probabilità uniforme attraverso un cammino casuale (random walk), i cui passi hanno orientazione a caso e lunghezza stabilita (o che segue una determinata distribuzione di probabilità).

Per capire cosa fare, prendiamo in esame un caso ideale: il gioco del 15, illustrato in figura 2.2-a, che però per semplicità riduciamo al gioco del 9, in cui abbiamo 8 tessere che si possono muovere in uno spazio composto da una griglia di 3×3 (9) quadrati. Dato che le tessere non possono sovrapporsi, si può pensare che a muoversi sia il "buco" (la tessera mancante, rappresentata come una tessera di colore diverso), come illustrato in figura 2.2-b. Normalmente le tessere vengono mosse con uno scopo, ma quando invece dobbiamo rimescolare il tutto, vogliamo che il "buco" si muova a caso. Cerchiamo di ottenere matematicamente questo risultato.

La probabilità $P(s_i, t)$ di trovarsi in un determinato quadrato i ad un dato tempo t è data dalla somma delle probabilità $P(s_j, t - 1)$ di trovarsi in uno dei quadrati j adiacenti al tempo precedente, per la probabilità di transizione $W(j \rightarrow i)$ di fare il "passo" verso il quadrato di

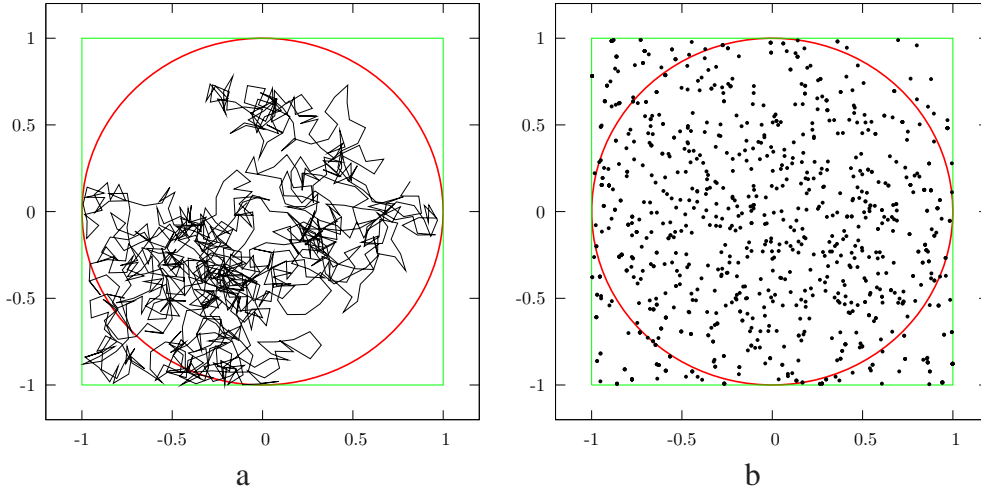


Figura 2.3: Traiettorie casuali per il calcolo dell'area del cerchio. (a) 1000 passi casuali di ampiezza $d = 0.1$, la traiettoria non copre ancora uniformemente il dominio. b) 1000 passi casuali di ampiezza $d = 0.4$. La distribuzione è molto più omogenea.

destinazione,

$$P(\mathbf{s}_i, t) = \sum_j P(\mathbf{s}_j, t-1)W(j \rightarrow i).$$

Possiamo chiamare $P(\mathbf{s}_j, t-1)W(j \rightarrow i)$ la probabilità di arrivare in i partendo da j in un passo.

Per chi ama il linguaggio matematico, quello che siamo facendo è cercare un processo stocastico a tempo discreto (la nostra traiettoria stocastica). Il processo deve raggiungere uno stato stazionario, e su tale stato le medie temporali (sulla traiettoria stocastica) devono essere uguali alle medie su tante diverse realizzazioni del processo (ovvero il processo dev'essere *ergodico*). In tali condizioni, la media sulle realizzazioni è descritta da una catena di Markov la cui matrice di transizione è $W_{ij} = W(i \rightarrow j)$. Vogliamo che la distribuzione asintotica di tale catena sia la nostra distribuzione di probabilità $P(\mathbf{s}_i)$ (che quindi non dipenderà più dal tempo).

Perché il processo sia stazionario, non dobbiamo avere “derive” nella probabilità. Questa condizione si può ottenere in tante maniere, per esempio attraverso cicli $i \rightarrow j \rightarrow k \rightarrow i$ che mantengono l'omogeneità del sistema. Tuttavia, la condizione più semplice è quella di imporre il *bilancio dettagliato*: la probabilità di arrivare in i da j in un passo sia sempre uguale a quella di arrivare in j da i :

$$P(\mathbf{s}_j)W(j \rightarrow i) = P(\mathbf{s}_i)W(i \rightarrow j).$$

Noi vogliamo che, mediando su un gran numero di mosse, la probabilità di trovare il buco in uno dei 9 quadrati sia costante ($P(\mathbf{s}_i) = 1/9$). Quindi otteniamo dall'equazione qui sopra che la probabilità $W(i \rightarrow j)$ di fare un passo da i a j sia la stessa che quella di fare un passo da j ad i , $W(j \rightarrow i)$.

Vediamo di trovare esplicitamente queste probabilità di transizione. Ovviamente, se i quadrati non sono adiacenti, questa probabilità è 0. Nella posizione centrale, se non vogliamo “sprecare tempo” in mosse inutili, imponiamo che la probabilità di stare fermo sia nulla, e quindi che il buco abbia probabilità 1/4 di muoversi verso i 4 quadretti vicini. Imponendo il bilancio dettagliato dobbiamo quindi avere che la probabilità di muoversi tra quadrati adiacenti è sempre uguale a 1/4. Ma i quadrati che stanno a metà di un lato hanno solo tre vicini, e quelli

che stanno sugli angoli ne hanno solo 2. Concludiamo che la probabilità di *restare fermi* sul quadrato centrale è 0, sui quadrati a metà dei lati è $1/4$, e sui quadrati di angolo è $1/2$, ovvero uguale alla probabilità di raggiungere un vicino “fuori” dall’area di interesse.

Quindi, abbiamo scoperto cosa dobbiamo fare quando un sasso “esce” dall’area di interesse: dobbiamo lasciar cadere un altro sasso nel punto in cui stiamo (per marcare il fatto che “stiamo fermi” un turno), e quindi tirare un altro sasso a caso. Dopo un po’ si formeranno dei mucchietti di sassi vicino alla recinzione, in particolar modo negli angoli di questa, ma questo servirà a compensare il fatto che gli angoli sono più difficili da raggiungere dato che hanno meno “vicini”. Questo metodo si chiama “rejection”.

Come si può vedere nella figura 2.3, anche con il cammino casuale si riesce a coprire il dominio con una distribuzione omogenea, ovviamente a patto di attendere un numero di passi sufficiente. Si può accelerare la copertura scegliendo accuratamente il passo del cammino casuale.

Vediamo adesso come applicare gli integrali stocastici a dei casi più realistici.

3.1 Traiettorie pesate

Per ora abbiamo trattato solo il caso di una distribuzione di probabilità uniforme in un certa area (il dominio), e zero fuori da questa. Ma abbiamo visto che nei sistemi a temperatura fissa la probabilità $P(\mathbf{s}_i)$ è proporzionale a $\exp(-E(\mathbf{s}_i)/T)$. Per fortuna, è possibile estendere le considerazioni fatte anche a questo caso. Riprendendo la formula per il bilancio dettagliato

$$P(\mathbf{s}_i)W(i \rightarrow j) = P(\mathbf{s}_j)W(j \rightarrow i),$$

si ottiene

$$\frac{W(i \rightarrow j)}{W(j \rightarrow i)} = \frac{P(\mathbf{s}_j)}{P(\mathbf{s}_i)} = \exp\left(\frac{E(\mathbf{s}_i) - E(\mathbf{s}_j)}{T}\right).$$

In parole povere la formula precedente dice che la probabilità di fare un passo da i a j è grande (rispetto alla probabilità di fare un passo da j a i) se l'energia di \mathbf{s}_i è maggiore di quella di \mathbf{s}_j , ovvero che in media “si scende” verso energie più basse. Il denominatore T ci dice che questa differenza si “misura” in unità della temperatura, per temperature alte la differenza conta meno che per temperature basse.

Ovviamente è facile generalizzare la procedura nel caso in cui lo stato \mathbf{s} vari in maniera continua, come nel caso del cerchio nel quadrato: dato un punto \mathbf{s} si effettua una variazione casuale, ottenendo \mathbf{s}' . La variazione o spostamento può essere di ampiezza costante o variabile. Se \mathbf{s}' cade fuori dal dominio, si impone $\mathbf{s}' = \mathbf{s}$. La collezione dei vari stati (alcuni dei quali ripetuti) ci dà la traiettoria fittizia $\mathbf{s}(k)$, su cui poi si effettuano le medie.

La formula precedente ci dà solo il rapporto tra le probabilità di transizione. Ci sono varie “ricette” compatibili con questo rapporto. La più famosa è probabilmente quella di Metropolis, che dice:

1. Calcola l'energia $E = E(\mathbf{s})$ della configurazione presente (\mathbf{s}).
2. Varia la configurazione in qualche maniera casuale, ottenendo \mathbf{s}' .
3. Calcola l'energia $E' = E(\mathbf{s}')$ della nuova configurazione.
4. Se $E' < E$, “accetta la mossa”, ovvero sostituisci \mathbf{s} con \mathbf{s}' .

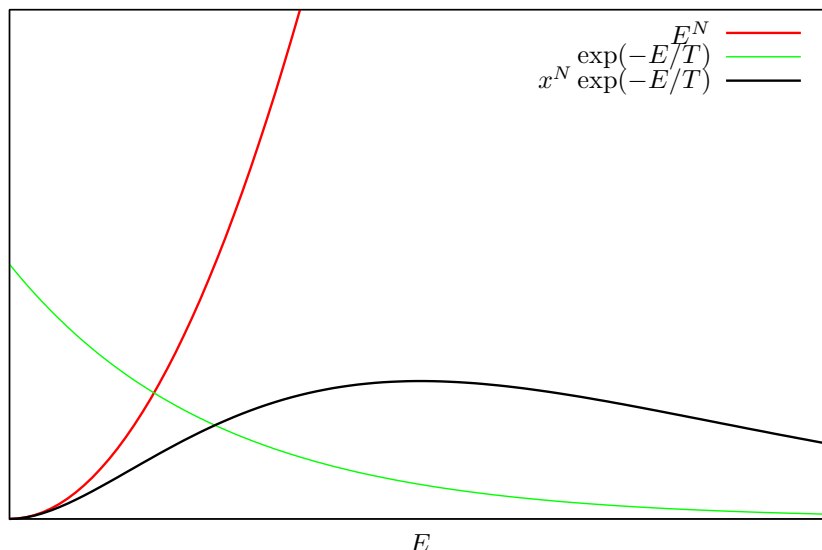


Figura 3.1: Il grafico qualitativo del numero di stati E^N , della loro probabilità $\exp(-E/T)$ e del prodotto di queste due quantità.

5. Altrimenti “accetta la mossa” con probabilità $\exp((E - E')/T)$.

Come si vede, le configurazioni ad energia “bassa” sono favorite rispetto a quelle a energia alta. Ci si potrebbe domandare come questa “discesa” costante dell’energia sia compatibile con il fatto che di solito i sistemi hanno una energia ben definita. Il fatto è che di solito il numero di configurazioni cresce con l’energia, così che c’è un bilancio tra probabilità di scegliere un “passo” che aumenta l’energia con la probabilità di compiere effettivamente questo passo. Per esempio, supponiamo che l’energia cresca come la distanza al quadrato rispetto all’origine degli assi (questa è forma dell’energia cinetica o dell’energia potenziale di una molla armonica). Se ci sono 2 variabili (per esempio la posizione x e la velocità v di un oscillatore armonico, abbiamo che il numero di punti nel nostro spazio degli stati corrispondenti con una energia E sia dato da tutti i punti a distanza \sqrt{E} dall’origine, quindi cresca come la distanza $r = E^{1/2}$ dall’origine (lunghezza del cerchio). Se abbiamo tre variabili, il numero degli stati ad energia costante corrisponderà alla superficie di una sfera di raggio $r = E^{1/2}$, ovvero crescerà come $r^2 \propto E$. Se prendiamo 4 variabili lo spazio degli stati sarà a 4 dimensioni, difficile da visualizzare, ma la “superficie” ad energia costante crescerà come r^3 , ovvero come $E^{3/2}$. Per un sistema con N gradi di libertà, il numero degli stati crescerà quindi come $E^{(N-1)/2}$, ovvero piuttosto rapidamente con E . D’altra parte, la probabilità di “salire” con l’energia fino a E decresce come $\exp(-E/T)$, garantendo da una parte che la traiettoria non “salga” mai all’infinito (l’esponenziale “batte” qualsiasi potenza) e fornendo un valore per l’energia media per grado di libertà (uguale a $T/2$), come mostrato in figura 3.1.

Infine, possiamo controllare che questa formula sia compatibile con quella per i sistemi chiusi: possiamo considerare il “dominio” precedente come una zona a energia costante, e lo spazio “fuori” dalla recinzione come uno spazio a energia infinita. La formula ci dice che tutti i passi “interni” al dominio sono equiprobabili, mentre quelli che oltrepassano la recinzione hanno probabilità nulla. Oppure, possiamo assegnare alla zona interna una energia nulla, e a quella esterna una energia positiva, crescente con la distanza dal dominio. A temperatura nulla, si ha che solo i passi “interni” sono accettati e quelli esterni sempre rifiutati. A temperature non infinite (positive) anche alcuni passi esterni sono accettati, anche se con probabilità bassa.

In questa maniera, utilizzando il peso dell’energia, possiamo estendere il dominio all’in-

finito (se l'energia cresce appunto all'infinito). Questo ci permette di fare calcoli anche su domini "disconnessi", dato che in questa maniera è possibile passare da una "valle" dell'energia all'altra. Questo è particolarmente utile quando il dominio o l'area di interesse è molto frastagliata, dato che altrimenti il sistema resterebbe intrappolato molto tempo in ogni "braccio".

Queste tecniche vengono sfruttate anche nel caso dell'ottimizzazione stocastica (simulated annealing), come illustrato negli approfondimenti.

Esaminiamo un po' più in dettaglio alcuni aspetti accennati in precedenza.

4.1 Dinamica molecolare

Lo studio dei sistemi gassosi del '700 e inizi '800 ci dice che tutti i gas, purché abbastanza diluiti ed a temperature abbastanza alte, si comportano nella stessa maniera. Questo vuol dire che il carattere “gassoso” non dipende molto dalle caratteristiche delle molecole. E in effetti, sappiamo che le molecole interagiscono fortemente solo a breve distanza, mentre quando sono ben distanziate non interagiscono (per questo si richiede che il gas sia diluito). Come mostrato in figura 4.1, l'interazione tra molecole non cariche è in genere di tipo fortemente repulsivo per una distanza r tra i centri minore della distanza interatomica R in un cristallo (come due palloni da calcio quando si cerca di “comprimerli” uno dentro l'altro) e debolmente attrattiva (a causa di effetti quantistici) quando le molecole sono separate da una distanza di pochi diametri.

Sappiamo inoltre che la temperatura è legata alla velocità media delle molecole. Se le molecole sono veloci, quando cozzano stanno in contatto per un tempo molto breve e sentono solo la parte repulsiva; se invece fossero lente, potrebbero stare molto tempo ad una distanza di pochi diametri, e questo potrebbe portare ad un effetto attrattivo (ovvero alla formazione di goccioline di liquido). Quindi, un gas diluito e a temperature moderatamente alte si comporta come una scatola piena di palline da ping-pong.

Cos'è la pressione? La pressione è la forza esercitata su una parete, divisa per l'area della parete stessa. La forza è ovviamente esercitata dalle palline che continuamente cozzano contro la parete. Ma durante ogni urto la forza è molto variabile. La forza è debolmente attrattiva (pressione verso l'interno della scatola) via via che le molecole si avvicinano, per poi diventare molto repulsiva durante l'urto vero e proprio, ritornando poi attrattiva quando la molecola si allontana. E questo si ripete ad ogni urto. Se usassimo uno strumento molto sensibile, sentiremmo un rumore molto variabile.

Ma i nostri sensi non sono in grado di seguire le collisioni delle molecole. Quello che misuriamo è una *media* (indicata da $\langle F \rangle$) della forza sulla parete, media calcolata su un intervallo di tempo che indico con τ (e che dipende dallo strumento di misura).

Ora, noi siamo in grado, usando un computer, di calcolare il moto di qualche milione di particelle che interagiscono attraverso la forza illustrata sopra. Sul sito web trovate alcuni programmi per fare queste simulazioni. Un gas “reale” ha un numero di particelle dell'or-

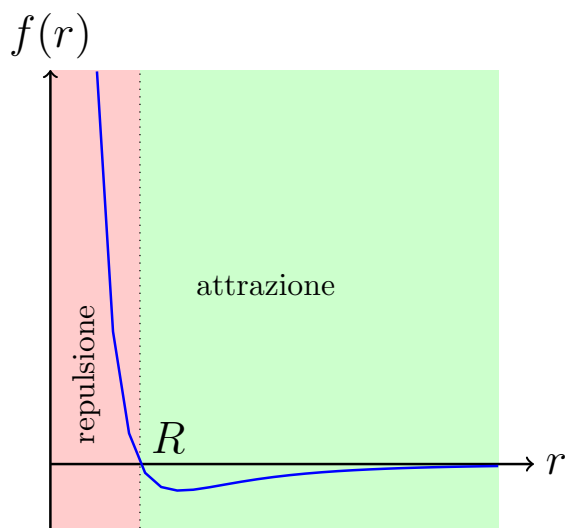


Figura 4.1: Diagramma della forza intermolecolare secondo lo schema di Lennard-Jones. Approssima bene l'andamento delle forze tra atomi simmetrici (gas nobili). Le forze sono repulsive a piccola distanza ed attrattive a distanze intermedie, e si annullano a grande distanza. La quantità R marca il punto in cui la forza si annulla e corrisponde alla distanza interatomica in un cristallo.

dine di 10^{23} , per cui il nostro misero 10^6 è una approssimazione un po' cruda, ma comunque funzionale.

Nell'approssimazione "classica", valida a temperature abbastanza alte per cui gli effetti quantistici sono trascurabili, è possibile calcolare la traiettoria (ovvero la sequenza di posizioni e velocità generate dal moto) di ognuna delle particelle.

Cominciamo con un caso semplice. L'integrazione numerica di una equazione differenziale del primo ordine per la variabile $x = x(t)$

$$\dot{x} = f(x),$$

dove, seguendo l'uso dei fisici, abbiamo indicato con un punto (\dot{x}) la derivata rispetto al tempo. Ricordando la definizione di derivata

$$\dot{x} = \lim_{\Delta t \rightarrow 0} \frac{x(t + \Delta t) - x(t)}{\Delta t},$$

e trascurando il limite per $\Delta t \rightarrow 0$, si può scrivere

$$x(t + \Delta t) = x(t) + f(x(t))\Delta t,$$

ovvero avanziamo di un passettino Δt nel tempo. Chiaramente, così facendo facciamo un errore, e ci sono schemi più sofisticati per ridurre questo errore, ma per il momento utilizziamo questo schema.

Come si fa, le equazioni di moto di Newton dicono che le accelerazioni sono proporzionali alle forze, ovvero la famosa $f = ma$ che noi scriviamo come

$$\ddot{x} = \frac{1}{m}f(x).$$

Introducendo la velocità $v = \dot{x}$ possiamo scrivere

$$\begin{aligned}\dot{x} &= v \\ \dot{v} &= \frac{1}{m}f(x)\end{aligned}$$

e da qui ottenere

$$\begin{aligned}x(t + \Delta t) &= x(t) + v(t)\Delta t \\ v(t + \Delta t) &= v(t) + \frac{\Delta t}{m}f(x(t)).\end{aligned}$$

Questa è il primo passo per la determinazione numerica delle traiettorie del moto di un sistema fisico. Ovviamente, per i casi pratici è molto meglio utilizzare algoritmi che non commettano errori così grandi come quello descritto, per non essere obbligati ad utilizzare intervallini Δt così piccoli che per avere uno spostamento significativo c'è bisogno di un "tempo macchina" enorme. Certo, questo è un problema di efficienza, ma anche l'efficienza può convertire un problema insolubile su scala pratica in un problema "affrontabile".

4.2 La distribuzione di equilibrio

Vogliamo determinare la distribuzione di equilibrio $P(s)$ *più probabile*, dati una serie di vincoli. Anche se questa affermazione suona un po' strana, l'idea è che una distribuzione di probabilità dà una serie di informazioni su un sistema. Pensiamo per esempio alla probabilità che esca una determinata faccia di un dado, $P(i)$ con $i = 1, 2, \dots, 6$. Se per esempio il dado è truccato così che $P(6) = 1$ e $P(1) = P(2) = \dots = P(5) = 0$, abbiamo la massima informazione (o la minima incertezza) del risultato (esce sempre il 6). Un "buon" dado è quello che ci dà la massima incertezza, ovvero tale che $P(i) = 1/6$.

Possiamo definire la quantità di incertezza S , o *entropia* della distribuzione $P(i)$, come

$$S = - \sum_i P(i) \log(P(i)),$$

con l'accortezza di definire $0 \log(0) = 0$ (che corrisponde al limite $\lim_{x \rightarrow 0} x \log(x)$). Come si può facilmente verificare, la certezza corrisponde a entropia $S = 0$ mentre la massima incertezza corrisponde a $S = \log(K)$, con K numero di stati.

Vogliamo adesso massimizzare S compatibilmente con certi vincoli. E' come trovare il punto di massima altezza di una strada: si massimizza la funzione "altezza", ma con il vincolo che il punto sia sulla strada (altrimenti ci darebbe la cima di una montagna). Guardando una cartina geografica, si vede facilmente che il punto di massima (o minima) altezza sulla strada si ha quando la strada stessa è *tangente* ad una curva di livello. Si tratta adesso di riscrivere questa condizione in termini matematici, ovvero di applicare la tecnica detta dei *moltiplicatori di Lagrange*. Per chi ha dimestichezza con l'analisi, l'estremo vincolato di una funzione $S(x_1, x_2, x_3, \dots)$, sottoposta ad un vincolo $G(x_1, x_2, x_3, \dots) = \text{costante}$, è dato dalla configurazioni di variabili x_i tali che

$$\frac{\partial}{\partial x_i}(S - \lambda G) = 0,$$

ovvero, in termini vettoriali

$$\nabla S - \lambda \nabla G = 0,$$

dove λ è una costante da determinare imponendo la condizione del vincolo alla soluzione. In essenza, ∇S è un vettore perpendicolare alle curve di livello di S (il cui modulo ci dice quando velocemente S sale o scende). Il vincolo corrisponde all'insieme dei valori delle variabili per cui G ha un certo valore, quindi si tratta di una curva di livello di G , e l'equazione vettoriale non dice altro che i vettori perpendicolari alle curve di livello delle due funzioni S e G devono essere proporzionali, ovvero che le curve di livello di S e G devono essere tangenti.

Nel nostro caso, vogliamo massimizzare S , che è una funzione delle varie componenti $P(i)$ della distribuzione di probabilità. Ma queste componenti non possono variare liberamente, dato che devono obbedire al vincolo $\sum_i P(i) = 1$ (normalizzazione della probabilità). L'equazione dei moltiplicatori di Lagrange in questo caso ci dice che

$$-\log(P(i)) - 1 - \lambda = 0,$$

ovvero che $P(i) = \text{costante} = 1/K$, risultato che corrisponde a quello che aspettavamo dall'esempio con i dadi (il valore di λ viene fissato dal vincolo).

Se adesso aggiungiamo il vincolo che l'energia media assuma un certo valore, ovvero

$$\sum_i E(i)P(i) = \langle E \rangle = \text{costante},$$

dobbiamo massimizzare

$$-\sum_i P(i) \log(P(i)) - \lambda_1 \sum_i P(i) - \lambda_2 \sum_i E(i)P(i).$$

Si ottiene

$$P(i) = \frac{1}{Z} \exp(-\beta E(i)),$$

in cui $Z = \sum_i \exp(-\beta E(i))$ è, come nell'altro caso, una costante che permette di normalizzare la distribuzione (primo vincolo). L'altra costante, β , deve essere determinata dal secondo vincolo.

Consideriamo due sistemi in contatto termico (possono scambiare energia ma non cambiare volume o scambiare particelle), ovvero che hanno la stessa temperatura T . Essi avranno anche la stessa energia media per particella (o meglio: per grado di libertà), il che ci dà l'indicazione che questa energia media sia collegata alla temperatura.

Supponiamo inoltre che i sistemi siano poco interagenti, così che l'energia E totale del sistema si possa scrivere come somma dell'energia delle due parti, $E = E_1 + E_2$. Come illustrato nell'approfondimento 4.3, in questo caso la distribuzione di probabilità si fattorizza, $P = P_1 P_2$ e l'entropia si somma $S = S_1 + S_2$.

Supponiamo adesso che ci sia il passaggio di una piccola porzione dell'energia, δE , dal sistema 1 al sistema 2 o viceversa. Dato che l'energia totale si conserva, la variazione dell'energia ΔE dei due sottosistemi dev'essere tale che $\Delta E_2 = -\Delta E_1$. Se però il sistema è all'equilibrio, l'entropia è massima e quindi la variazione dell'entropia dev'essere nulla al primo ordine in ΔE . Quindi

$$\Delta S = \frac{\partial S_1}{\partial E_1} \Delta E_1 + \frac{\partial S_2}{\partial E_2} \Delta E_2 = \left(\frac{\partial S_1}{\partial E_1} - \frac{\partial S_2}{\partial E_2} \right) \Delta E_1 = 0.$$

e quindi

$$\frac{\partial S_1}{\partial E_1} = \frac{\partial S_2}{\partial E_2},$$

che deve corrispondere all'equivalenza della temperatura tra i due sistemi.

Una condizione formalmente uguale si ottiene in termodinamica, in cui l'entropia ha una definizione almeno apparentemente molto diversa, e in cui si ottiene

$$\frac{\partial S}{\partial E} = \frac{1}{T},$$

con T temperatura termodinamica. Si può poi dimostrare che questa temperatura è la stessa di quella definita empiricamente attraverso i termometri a gas (ovvero dall'equazione del gas perfetto. Se adesso identifichiamo l'entropia da noi definita (entropia di informazione) con quella termodinamica, si scopre che tutto torna, e riusciamo a trovare un "nome" per la costante β , ovvero otteniamo la *distribuzione di Boltzmann*

$$P(\mathbf{s}) = \frac{1}{Z} \exp\left(-\frac{E(\mathbf{s})}{T}\right),$$

con

$$Z = \int \exp\left(-\frac{E(\mathbf{s})}{T}\right) d\mathbf{s}.$$

4.3 La "partizione"

Se si possono dividere le variabili \mathbf{s} in due insiemi (indipendenti) \mathbf{x} e \mathbf{y} in maniera tale che l'energia $E(\mathbf{s})$ sia $E(\mathbf{s}) = E(\mathbf{x}) + E(\mathbf{y})$, allora

$$\begin{aligned} Z &= \int d\mathbf{s} \exp\left(-\frac{E(\mathbf{s})}{T}\right) \\ &= \iint d\mathbf{x} d\mathbf{y} \exp\left(-\frac{E(\mathbf{x}) + E(\mathbf{y})}{T}\right) \\ &= \int d\mathbf{x} \exp\left(-\frac{E(\mathbf{x})}{T}\right) \int d\mathbf{y} \exp\left(-\frac{E(\mathbf{y})}{T}\right) \\ &= Z_x Z_y, \end{aligned}$$

dove si è sfruttato il fatto che $d\mathbf{s} = d\mathbf{x} d\mathbf{y}$. Da qui

$$\begin{aligned} P(\mathbf{s}) &= \frac{1}{Z} \exp\left(-\frac{E(\mathbf{s})}{T}\right) \\ &= \frac{1}{Z_x Z_y} \exp\left(-\frac{E(\mathbf{x}) + E(\mathbf{y})}{T}\right) \\ &= \frac{1}{Z_x} \exp\left(-\frac{E(\mathbf{x})}{T}\right) \frac{1}{Z_y} \exp\left(-\frac{E(\mathbf{y})}{T}\right) \\ &= P_x(\mathbf{x}) P_y(\mathbf{y}). \end{aligned}$$

In un gas, nel limite in cui possiamo considerare le molecole come puntiformi con interazioni istantanee, abbiamo

$$E = \frac{1}{2} \sum_i \frac{m_i}{2} ((v_x)_i^2 + (v_y)_i^2 + (v_z)_i^2),$$

e quindi abbiamo che tutte le variabili si separano, le coordinate x , y e z semplicemente non compaiono nell'energia, mentre le velocità compaiono quadraticamente. Si ottiene che la

distribuzione di probabilità per le coordinate è piatta (la posizione è equiprobabile dentro il volume del contenitore), mentre la distribuzione delle velocità è gaussiana (distribuzione di Maxwell)

$$P(v) = \frac{1}{\sqrt{2\pi T/m}} \exp\left(-\frac{mv^2}{2T}\right).$$

Il sistema che originalmente aveva $6M$ variabili (tre coordinate posizionali e tre velocità per ognuna delle M molecole, si spezza in $6M$ problemi indipendenti, ognuno dei quali molto semplice da risolvere. Si ottiene così facilmente la “legge” del gas perfetto $PV = NT$.

Non sempre il problema è così facile. Se invece di molecole puntiformi usiamo la forma della forza illustrata nell’approfondimento 4.1, non siamo in grado di separare completamente il problema: possiamo fattorizzare le $3M$ velocità, ma le posizioni rimangono accoppiate.

In particolare, possiamo utilizzare la separabilità dell’energia per studiare separatamente i vari gruppi di variabili. Spesso, la distribuzione di probabilità delle velocità è facile da ottenere anche a mano, mentre quella delle posizioni va calcolata con il computer.

4.4 Realizzazione di un processo stocastico

L’analisi che abbiamo fatto ci ha portato a trovare una catena di Markov $W_{ij} = W(i \rightarrow j)$ che ci dà la probabilità di transizione tra lo stato s_i e lo stato s_j . Potremmo iterare semplicemente la distribuzione di probabilità $P(s_i, t)$ secondo l’equazione

$$P(s_i, t + 1) = \sum_j W_{ji} P(s_j, t),$$

e da qui ricavare il valore delle osservabili

$$\langle F \rangle(t) = \sum_i f(s_i) P(s_i, t).$$

Però, in questa maniera dobbiamo iterare un vettore ($P(s)$) che ha moltissime componenti. Supponiamo per esempio che si voglia studiare un gas composto da 10 particelle in una scatola di lato L . Dobbiamo *discretizzare* il sistema in quadretti. Pensiamo di dividere il lato della scatola in 10 intervalli, ed anche di dividere i possibili valori della velocità in 10 intervalli (una discretizzazione molto rozza). Lo stato di una particella è dato dalla sua posizione (x, y, z) e dalla sua velocità (v_x, v_y, v_z) , ovvero da 6 numeri. quindi i possibili stati di una particella sono 10^6 , ovvero un milione. Gli stati possibili di 10 particelle sono 10^{60} , che è un numero astronomico. E dato che $P(s)$ mi deve dare la probabilità di trovare il sistema in ognuno degli stati possibili, vuol dire che P ha 10^{60} componenti (e W ne ha 10^{120}). Non c’è nessuna speranza di poter comprare un computer con questa memoria, e comunque il numero di operazioni da fare sarebbe tremendo. Per fortuna, queste configurazioni sono molto ripetitive, per esempio, se le molecole sono uguali, tutte quelle che si ottengono scambiando tra loro due molecole hanno le stesse proprietà. Ma tenere traccia di tutte le simmetrie è molto difficile. E’ molto più facile cercare un processo stocastico (una traiettoria) ergodico che sia la distribuzione di probabilità cercata come media.

Come si fa a ottenere questo processo? L’idea di base è semplice: data una configurazione $s_i = s(t)$ ad un determinato tempo t , si deve scegliere un’altra configurazione s_j e fare il passo con probabilità W_{ij} . Cosa vuol dire fare “un passo con una determinata probabilità”? quello che si fa di solito è utilizzare una funzione (supponiamo che si chiami `random()`),

di cui parleremo tra poco, che ogni volta che viene chiamata ritorna un numero casuale tra 0 (incluso) ed 1 (escluso) con distribuzione di probabilità piatta. La probabilità che questo numero sia minore di una quantità p (tra zero ed uno) è proprio p , quindi basta scrivere qualcosa del tipo

```
r = random()
if (r < p) then ...
```

con $p = W_{ij}$. Se il test $r < p$ viene passato allora $s(t+1) = s'$.

Il problema è cosa fare se il test fallisce. Se, come nel caso del bilancio dettagliato, abbiamo la libertà di scegliere la probabilità da assegnare alla mossa nulla, ovvero alla “diagonale” W_{ii} , allora basta adottare come strategia che se il test $r < p$ fallisce non facciamo nulla e $s(t+1) = s(t)$ e semplicemente abbiamo sprecato un passo temporale (ovvero l’algoritmo di campionamento sarà lento. Per esempio, se prendiamo il gioco del 9, abbiamo visto che la tessera centrale ha 1/4 di probabilità di muoversi nelle quattro direzioni. Questo però è vero solo se scegliamo come mosse possibili lo spostamento nelle quattro direzioni. Potremmo invece scegliere a caso un’altra posizione tra le 9 disponibili, sempre avendo che W_{ij} è zero tranne che per gli spostamenti sui primi vicini. Ma allora, avremmo che lo spostamento verso uno dei primi vicini avviene con probabilità 1/9, e quindi la tessera deve rimanere ferma con probabilità $1 - 4/9 = 5/9$. L’algoritmo è chiaramente meno efficiente.

Questo esempio sottolinea l’importanza di proporre come “mossa” una configurazione tale che la transizione abbia una probabilità non nulla di avvenire. Non vale quindi la pena di scegliere, come configurazione s' , una configurazione a caso: abbiamo visto come il numero di configurazioni aumenti molto velocemente con l’energia, e quindi prendendo una configurazione a caso sarà con grande probabilità una configurazione di alta energia, e di conseguenza la probabilità associata alla mossa sarà piccola, e l’algoritmo molto inefficiente. Viceversa, conviene in genere prendere come configurazione s' una piccola variazione di s , così che la transizione avvenga con una probabilità alta. Certo, le due configurazioni $s(t)$ e $s(t+1)$ saranno molto *correlate*, e quindi non è corretto effettuare la misura di F su ogni configurazione (altrimenti si finisce per sottostimare le fluttuazioni della F). La regola d’oro del calcolo Monte-Carlo è che l’unità di “tempo fittizio” corrisponde a N passi elementari, dove N è il numero di variabili del sistema (ovvero dopo che ogni variabile ha avuto, in media, una occasione di cambiare).

Esistono comunque algoritmi che possono aumentare la probabilità di accettazione della mossa, o che possono “scorrelare” rapidamente le configurazioni. Per esempio, se il modello sotto esame è “vicino” ad uno risolubile analiticamente (piccole correzioni all’energia), si può sfruttare questa informazione per “preparare” la mossa, oppure si possono fare molti cambiamenti contemporaneamente, come nell’algoritmo di Swendsen-Wang. Per ulteriori informazioni, vedere i lavori di Werner Krauth [4].

Avevamo lasciato in sospeso cosa fare nel caso in cui non si abbia la libertà di scegliere W_{ii} . In questo caso, la cosa più efficiente (non sempre possibile) è quella di preparare tutte le possibili mosse, con le corrispondenti probabilità $p_1 = W_{i,1}, p_2 = W_{i,2}, \dots$. Ovviamente $\sum_j p_j = \sum_j W_{ij} = 1$. Dato un numero casuale r (tra zero ed uno), si comincia a confrontarlo con la prima possibilità, e se fallisce sottraiamo p_1 e lo confrontiamo con la seconda, e così via finché non troviamo l’azione corrispondente:

```
r = random()
i = 1
while (r > p(i)) then
```

```

r = r - p(i)
i = i + 1
end while

```

Ovviamente è molto meglio se le prime mosse proposte sono quelle che corrispondono alle probabilità più grandi.

Per terminare questa discussione, possiamo domandarci come fa un computer, che è una macchina deterministica (fa la stessa azione ogni volta che riparte dalla stessa configurazione, anche se a molti utenti questo non sembra vero), a generare dei numeri casuali.

Ma cosa intendiamo per casuali? per ora abbiamo solo parlato della distribuzione di probabilità, ma ovviamente questa proprietà non esaurisce l'argomento. Anche una sequenza ordinata di numeri potrebbe "coprire" un intervallo in maniera "piatta" (equispaziata) ma in maniera molto ordinata. In realtà, quello che vogliamo è che il processo sia imprevedibile: che il prossimo numero non si possa prevedere per quante informazioni io abbia sul sistema. Prendiamo il lancio di una moneta, o il "rosso/nero" alla roulette (o anche la sequenza dei numeri del lotto). Se i processi sono veramente aleatori, allora non mi servirà a niente conoscere la serie dei risultati passati (o il "ritardo" del numero del lotto), perché ogni nuova estrazione sarà scorrelata dalla precedente.

Questa imprevedibilità ha dei risvolti pratici molto importanti: quando i computer vogliono trasmettersi dei dati in forma criptata, usano dei numeri che non siano facilmente prevedibili (senza conoscere delle informazioni che i computer si sono scambiati in precedenza) per "offuscare" il messaggio.

I computer non possono generare numeri casuali. Però, possono sfruttare dei processi deterministici caotici per generare delle sequenze con dei tempi di correlazione molto lunghi, così che per trovare delle regolarità si debba osservare una sequenza astronomicamente lunga di passate estrazioni. I buoni sistemi operativi inoltre forniscono dei numeri "imprevedibili" ottenuti sfruttando varie sorgenti di imprevedibilità del mondo in contatto con il computer: l'intervallo temporale tra eventi quali la pressione di un tasto o il movimento del mouse, i pacchetti che arrivano dalla rete ecc.

4.5 Pregi e difetti del metodo Monte-Carlo

Il pregio più grande del metodo Monte-Carlo è la sua semplicità. In fondo si tratta solo di proporre una nuova configurazione e di calcolare la variazione dell'energia corrispondente.

La cosa che forse disturba più ingegneri e matematici è che si tratta di un metodo stocastico: non avremo mai l'area esatta della figura, ma solo una approssimazione che (se tutto va bene) convergerà piano piano al valore "vero". La teoria dei processi stocastici ci dice che questa convergenza va come la radice quadrata del numero di mosse, quindi una convergenza piuttosto lenta: per dimezzare l'errore dobbiamo quadruplicare il numero di mosse. D'altra parte, rispetto al metodo della suddivisione in quadrati, il metodo Monte Carlo ha il grande vantaggio che abbiamo rapidamente una stima del valore asintotico, mentre i dati provvisori dei metodi "esaustivi" di solito non danno nessuna informazione sul risultato finale.

Cerchiamo adesso di mettere in evidenza altri punti deboli del metodo. Abbiamo supposto di poter racchiudere la figura di cui vogliamo calcolare l'area dentro un dominio semplice. Ma spesso non conosciamo a priori i limiti della figura. Possiamo ovviamente "esagerare" e prendere un dominio molto più grande del necessario, ma questo pone altri problemi: supponiamo di racchiudere il nostro cerchio di due metri di diametro in un quadrato di lato un chilometro, e di usare un elicottero per lanciare i sassi. Avremo bisogno di un numero spropositato di sassi.

Troviamo quindi una “regola aurea” per il metodo Monte Carlo: l’ottimo è avere un rapporto tra sassi “buoni” e sassi “sprecati” intono al 50%.

Possiamo adattare le raccomandazioni fatte prima al metodo del random walk: il rapporto tra i sassi che cadono nella zona di interesse e quelli che cadono fuori da questa deve essere del 50% per avere la migliore convergenza. Ovvero: se facciamo lanci troppo piccoli, restiamo quasi sempre o dentro il cerchio o fuori da questo e la convergenza sarà troppo lenta. Se facciamo lanci troppo lunghi di nuovo abbiamo una statistica cattiva: quasi tutti i sassi usciranno dal recinto e non aumenteremo mai la precisione.

Come si fa a scoprire quale è la distanza ottimale di lancio? Basta tenere d’occhio la frazione di sassi “buoni” e sassi “sprecati”.

4.6 L’ottimizzazione stocastica: il simulated annealing

Una delle applicazioni più creative del metodo Monte-Carlo sta nell’utilizzare l’analogia tra energia ed costo di un processo per sviluppare una tecnica di ottimizzazione stocastica: il simulated annealing (ricottura simulata).

Tutti i processi di ottimizzazione mirano a trovare il minimo di una funzione. Nei casi reali non banali, le funzioni hanno molti minimi locali, uno dei quali è il minimo più minimo di tutti: il minimo assoluto. Se pensiamo alla funzione come ad un costo, il minimo assoluto corrisponde alla soluzione di minor costo. Se pensiamo ad una carta geografica di un paesaggio alpino, i minimi corrispondono ai punti più bassi delle valli, di solito sul fondo di un lago. Alcune valli (senza lago) non hanno minimo, ma una gola che porta ad altre valli. Ovviamente, il minimo assoluto sulla faccia della terra corrisponde alla fossa delle Marianne.

Se eliminiamo l’acqua sulla Terra (o lavoriamo su Marte o sulla Luna), possiamo associare ad ogni minimo un “bacino” di pertinenza: sono tutti i punti in cui, lasciata andare una pallina (che si muova di moto viscoso), si arriva al minimo in questione.

Ovviamente si possono presentare molti casi: il minimo assoluto potrebbe essere ben distinguibile, con un bacino molto ampio mentre i minimi “spuri” potrebbero essere molti meno profondi ed avere bacini piccoli. Questa è una situazione facile da risolvere: una volta che siamo arrivati nel bacino del minimo assoluto, basta continuare a scendere e troviamo il fondo. Questi sono i metodi deterministici (che ovviamente possono essere anche molto sofisticati, per aumentare l’efficienza).

Oppure ci potrebbe essere il caso opposto: molti minimi, alcuni dei quali profondi, distribuiti a caso e con bacini piccoli. Questo è il caso “impossibile”: non c’è molto da fare che cercare esaustivamente il minimo assoluto.

Per fortuna, sembra che molti dei problemi “mediamente difficili” abbiano una struttura ad imbuto (funnel): se visti “a grana grossa” mostrano una struttura semplice, con una valle larga ed un minimo chiaro, mentre su scala più piccola cominciano ad apparire dei minimi locali, con bacini abbastanza piccoli. Questo del resto è anche il caso delle valli alpine: se noi lasciamo andare giù una palla da ping-pong probabilmente questa si bloccherà tra due sassi (un minimo locale), se lanciamo una pallone del diametro di qualche metro (o aspettiamo che la valle sia coperta di neve) probabilmente arriviamo molto più in basso.

Il metodo del simulated annealing è adatto a questi casi mediamente difficili. Esistono problemi “difficili” (anche se non impossibili) in cui ci sono molti minimi, e per questi casi sono state sviluppate delle varianti del simulated annealing (come il simulated tempering).

Certo, non sempre siamo interessati al minimo assoluto. Spesso ci accontentiamo di un “buon” minimo, magari vicino al punto di partenza (come fa l’evoluzione naturale), oppure

siamo interessati a sapere qual'è la statistica dei minimi, in modo da sapere quanto tempo occorrerà in media per trovare delle soluzioni migliori di quelle già trovate, e se il gioco vale la candela.

Esempi di problemi “mediamente difficili” sono quelli in cui il numero di soluzioni cresce in maniera combinatoria. Un esempio classico è il problema del commesso viaggiatore: dato un insieme di città disposte su una carta geografica, il nostro commesso deve tracciare l'itinerario di lunghezza minima che passi da tutte le città almeno una volta. Una soluzione è rappresentata dalla lista delle città da visitare in ordine, ma ovviamente ci sono molte possibili itinerari. Un problema simile potrebbe essere quello della costruzione di un orario scolastico, che tenga conto del numero di ore di ogni materia, delle indisponibilità dei docenti, del fatto che non si possono avere più di tre ore consecutive di una materia, cercando di non avere “buchi”, ecc.

Queste soluzioni sono caratterizzate da una funzione costo, che va minimizzata, e da una serie di vincoli che non possono essere violati.

Data una soluzione proposta, l'approccio intuitivo è quello di variarla in maniera compatibile con i vincoli, tenendo traccia di quelle variazioni che abbassano il costo, e ripartendo poi da queste. Così facendo, resteremo però intrappolati in un minimo locale, come la pallina da ping-pong che resta incastrata tra i sassi.

Se ci immedesimiamo in Poseidone, dio dei terremoti, possiamo trovare la soluzione al problema della pallina: agitiamo un po' la Terra così che la pallina rimbalzi, sperando che prosegua la sua discesa verso il fondovalle.

L'idea del simulated annealing è simile. Se immedesimiamo la funzione costo con l'energia di un sistema fisico, possiamo effettuare una simulazione Monte-Carlo ad una determinata temperatura T . Quando la temperatura è alta, il punto rappresentativo (la soluzione) può “saltellare” qua e là, quando la temperatura è bassa invece effettuerà di preferenza dei movimenti che abbassano l'energia.

Sappiamo inoltre che, se aspettiamo un tempo lungo, la distribuzione di probabilità della traiettoria Monte-Carlo approssimerà la distribuzione di Boltzmann

$$P(\mathbf{s}) = \frac{1}{Z} \exp\left(-\frac{E(\mathbf{s})}{T}\right),$$

ovvero sarà concentrata in corrispondenza delle configurazioni corrispondenti alle energie più basse, con una “dispersione” data dalla temperatura T . La temperatura permette al punto rappresentativo di “saltare fuori” dai minimi locali, in effetti la dinamica “non vede” tutti i minimi che hanno una profondità (rispetto ai “passi”) che sia ordine T . Se la nostra funzione costo ha una struttura ad imbuto, possiamo supporre che, dopo un certo tempo, la distribuzione di probabilità sia concentrata in corrispondenza del fondo dell'imbuto. A questo punto possiamo abbassare la temperatura, concentrando la probabilità in corrispondenza dei minimi. Se “raffreddiamo” molto lentamente il sistema, riusciamo alla fine a raggiungere il minimo assoluto.

Si tratta, anche più del Monte-Carlo, di una procedura facile, ma lenta. Il punto più difficile è quello di generare nuove configurazioni che siano compatibili con i vincoli. Per facilitare il compito è possibile “addolcire” i vincoli, inserendoli nella funzione costo con un peso alto, così che sia più facile generare nuove configurazioni e si possa eventualmente violare un vincolo per qualche passo. Per esempio, nella costruzione di un orario scolastico si potrebbe “mettere da parte” per qualche tempo una materia, salvo reinserirla alla fine.

Il simulated annealing non esaurisce il problema dell'ottimizzazione stocastica. Wikipedia è un buon punto di partenza per approfondire l'argomento.

Bibliografia

- [1] N. Metropolis, *The Beginning of the Monte Carlo Method*, Los Alamos Science Special Issue 1987, 125 (1987), <http://library.lanl.gov/cgi-bin/getfile?00326866.pdf>
- [2] R. Eckhardt, *Stan Ulam, John von Neumann and the Monte Carlo Method*, Los Alamos Science Special Issue 1987, 131 (1987), http://library.sciencemadness.org/lanl1_a/lib-www/pubs/00326867.pdf
- [3] *Fermi Collected Papers* (1965) p. 977 <http://www.physics.utah.edu/~detar/phy6720/handouts/fpu/FermiCollectedPapers1965.pdf>
- [4] Werner Krauth *Statistical Mechanics: Algorithms and Computations* (Oxford University Press 2006); Introduction to Monte Carlo Algorithms by W. Krauth in “Advances in Computer Simulation” (J. Kertesz and I. Kondor, eds) Lecture Notes in Physics (Springer Verlag, 1998) arXiv:cond-mat/9612186
- [5] S. Kirkpatrick, C.D. Gelatt Jr., M.P. Vecchi, *Optimization by simulated annealing*, *Science* **220**, 671 (1983)<http://www.fisica.uniud.it/~ercolessi/MC/kgv1983.pdf>

1	Dalla Dinamica alla statistica e ritorno	1
1.1	Introduzione	2
1.2	Dalla dinamica alla statistica	4
1.3	La distribuzione di probabilità	6
2	Integrali stocastici	7
2.1	Integrali Monte-Carlo	7
2.2	Traiettorie stocastiche	8
3	Back to physics	12
3.1	Traiettorie pesate	12
4	Approfondimenti	15
4.1	Dinamica molecolare	15
4.2	La distribuzione di equilibrio	17
4.3	La “partizione”	19
4.4	Realizzazione di un processo stocastico	20
4.5	Pregi e difetti del metodo Monte-Carlo	22
4.6	L’ottimizzazione stocastica: il simulated annealing	23