

Mille mondi nel computer:

le simulazioni numeriche

di sistemi complessi

Enzo Marinari

(Roma *La Sapienza*)

Aula *La Ginestra*, 8 novembre 2006



NUOVO FUMETTO:
linus

Grazie agli organizzatori per l'invito.

Che cosa vi racconto?

La complessità.

Ma, in primo luogo, che cosa è la complessità?

Dare una definizione è una parte notevole della risoluzione del problema.

Fino a pochi anni fa un controllo quantitativo era impossibile.

Le cose sono cambiate grazie ai **computer**.

I computer suggeriscono e permettono nuove domande.

Che cosa vi racconto?

Una definizione di complessità.

Il ruolo importante della frustrazione.

Le simulazioni numeriche e il metodo Monte Carlo.

I sistemi complessi: i vetri di spin, i vetri strutturali, la
NP-completezza.

Le simulazioni numeriche di sistemi complessi.

Le barriere ci rallentano: le dinamiche lente.

I metodi Monte Carlo ottimizzati.

Mille mondi nei computer.

Google è davvero incredibile.

Ho cercato una definizione di **complexity**...

Ne ho trovate moltissime. Nelle due pagine seguenti solo una scelta selezionata.

Il gran numero di possibili definizioni e la loro enorme diversità ci dicono già chiaramente che scegliere una definizione utile è certamente non banale.

Ma guardando le definizioni scopriamo che Google è davvero incredibile, perchè ci dice, se lo sappiamo scorgere, tutto quel che ci interessa (Giorgio Parisi a questo punto citerebbe un oracolo greco, che non dice ma si limita a indicare).

define:complexity - Cerca con Google

file:///home/marinari/TALKS/FACOLTA/defcomp.html

[Accesso](#)

[Google](#)

Web [Immagini](#) [Gruppi](#) [News](#) [altro »](#)

define:complexity

Cerca

[Ricerca avanzata](#)
[Preferenze](#)

Cerca: il Web pagine in Italiano pagine provenienti da: Italia

Web

Frase correlate: [cyclomatic complexity](#) [computational complexity](#) [cognitive complexity](#) [complexity theory](#) [complexity class](#) [essential complexity](#) [accidental complexity](#) [time complexity](#) [date complexity metric](#) [complexity classes](#)

Definizione di **complexity** su Internet Inglese:

- the quality of being intricate and compounded; "he enjoyed the complexity of modern computers"
wordnet.princeton.edu/perl/webwn
- We can say there are two kinds of complexity. Detail Complexity is when there are many variables. Dynamic Complexity is situations where cause and effect are subtle, and where the effects over time of interventions are not obvious.
www.worldtrans.org/whole/wholedefs.html
- When a wine is at once rich and deep, yet balanced and showing finesse. No greater compliment can be paid a wine. A mature Chateau Latour, d'Yquem or La Tache Burgundy are prime examples of complex wines.
www.eosvintage.com/glossary.html
- Generally avoided as an overused and poorly defined word, except in specific systems.
ishi.lanl.gov/diversity/Glossary1_div.html
- Poor Terminology! Like `specificity`, the term `complexity` appears in many scientific papers, but it is not always well defined. (See however M. Li and P. Vitanyi, A Introduction to Kolmogorov Complexity and Its Applications, second edition, Springer-Verlag, New York, ISBN 0-387-94868-6, 1997) When one comes across a proposed use in the literature one can unveil this difficulty by asking: How would I measure this complexity? What are the units of complexity? ...
www-lmmb.ncifcrf.gov/~toms/glossary.html
- is a measure of the number of possible states a system can take on, ie, the condition of a system, situation, or organization that is integrated with some degree of order but has too many elements and relationships to understand in simple analytic or logical ways.
www.mountainquestinstitute.com/definitions.htm
- The term "low complexity sequence" may be thought of as synonymous with regions of locally biased amino acid composition. In these regions, the sequence composition deviates from the random model that underlies the calculation of the statistical significance (P-value) of an alignment. Such alignments among low complexity sequences are statistically but not biologically significant, ie, one cannot infer homology (common ancestry) or functional similarity.
falcon.roswellpark.org/labweb/glossary.html
- References to games as more or less complex or simple are almost invariably concerned with rules mechanisms, and usually embody some kind of subjective judgement. Complexity in game settings is almost universally considered desirable and probably inevitable if a campaign (qv) is to evolve and display any kind of depth or subtlety.
www.mud.co.uk/richard/ifan394.htm
- the perception of multiple flavors
www.coffeeuniverse.com/university_taste.html

1 of 2

11/01/2006 03:40 PM

define:complexity - Cerca con Google

file:///home/marinari/TALKS/FACOLTA/defcomp.html

- Short name for McCabe Cyclomatic Complexity.
www.mccabe.com/iq_research_iqgloss.htm
- The number of species at each trophic level and the number of trophic levels in a community.
highered.mcgraw-hill.com/sites/0070294267/student_view0/glossary_a-d.html
- In services marketing, the number of steps required to perform a service. p. 632
users.wbs.warwick.ac.uk/dibb_simkin/student/glossary/ch20.html
- A set of structure-based metrics that measure the attribute of the degree to which a system or component has a design or implementation that is difficult to understand and verify. IEEE96
www.hi.is/~oddur/spisland/ref/def.htm
- Not clean or flat , interesting to drink.
www.geocities.com/NapaValley/1824/winegloss.htm
- Interdependence means that power is shared among branches of government
fs.huntingdon.edu/~lewis/Syl/PA/306StillmanStudOuts.htm
- domain of emergent properties and non-linear relationships between factors; unlike chaos, which is inherently uncertain, may often create an illusion of predictability, especially where linear analysis is applied within a short-term, narrow set of assumptions
www.soul-dynamics.com/glossary
- A term often used to describe blended coffees but not restricted to them. If a coffee has good complexity it should have a harmonious multiplicity of taste sensations.
abyssinica.co.uk/glossary.htm

define:complexity Cerca

[Strumenti per le lingue](#) | [Suggerimenti per la ricerca](#) | [Non sei soddisfatto? Aiutaci a migliorare](#)

[Pagina principale di Google](#) - [Pubblicità](#) - [Soluzioni Aziendali](#) - [Tutto su Google](#)

©2006 Google

Ma che cosa è complesso, e cosa è realmente la complessità?

La teoria dell'informazione ha un ruolo importante nel costruire un primo approccio quantitativo.

Kolmogorov: la complessità algoritmica di un numero intero o di una stringa di caratteri di un alfabeto è la lunghezza in bit del messaggio più breve che può essere usato per identificare il numero.

Si può anche dire, in modo equivalente, che la complessità di Kolmogorov di un numero è la lunghezza del programma più breve che è in grado, dopo aver fatto tutto il lavoro necessario, di stampare questo numero.

Al peggio: copiamo la stringa e la stampiamo.

Questa sequenza è complessa?

F8885D34899B6DD7
9D401E83C432403
2B9ADDA82FD46D1F
841DCDF54A51C9EB
4A1BAEFB95FE0227
F628D3F125B91093
99A38A1C93EEFEEF
31796F294E571FB
53D04A429EEBA577
D9351CFEE4539823
6D103B14451A27BF
A5864643D26EDDOB
C8EA24F3565327C7
5E5C4041F97BCAB3
3CD10FF8076E378F
A0507F5EDF2E9B1B
6957A563342A5917
43FF8762BBC03C2B

Ma no!

```
main(){
  int i;
  unsigned long long int seed, noComplex, a = 1181783497276652981ULL;
  seed = 12473563ULL;
  noComplex = seed;
  for(i=0;i<20;i++){
    noComplex *= a;
    printf("%LX\n",noComplex);
  }
}
```

numeri in esadecimale (base 16).

Un codice che produca 10^{15} numeri ha la stessa lunghezza. È fra l'altro un “ottimo” generatore (inganna tutti...).

Generatore di valori pseudo-aleatori. (L'Ecuyer)

E questa sequenza è complessa?

E1D5F51953BE0967
3C5CFB245C6870D3
31A5A06455CD4C2F
8775C6BFEA2A503B
97363541100208B7
7FD9366D83FE0463
476C88216F53F0FF
199B8A6C7DEE674B
C8AD45C92EE12707
9AA7C2FAF9E382F3
8EA56CCE7F9ABCCF
ECD32EC16F19115B
38600D666B8D3457
38775E3702E0FC83
198BB7C94FE4FF9F
1A29EFE9C051DE6B
8A3164372FCF00A7
CE836DE08F818113

Ovviamente sí.

I dati riassumono un esperimento statistico sull'umore del direttore del nostro dipartimento.

Umore buono: 1, umore cattivo: 0.

4 giorni: 0110, e cioè un carattere esadecimale,
 $0, 1, \dots, 9, A, B, \dots, F$.

16 volte (cioè 64 giorni): in totale su una riga 64 bit, e cioè 16 caratteri esadecimali.

18 righe, e cioè un periodo di circa 3 anni (la durata di un ciclo elettorale).

Si tratta di un fenomeno assolutamente complesso, collegato a una miriade di fattori esterni.

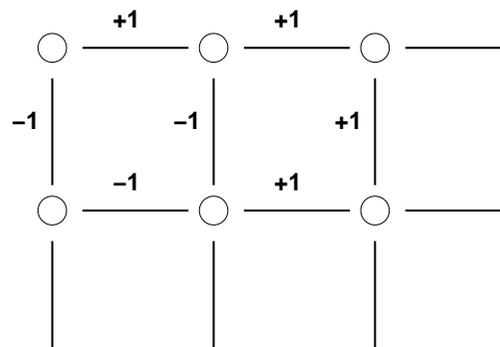
E questa sequenza è complessa?

27F2D011D155096F
58334FA707EB5E7B
B59A5B9D4C565BF7
1CA6EF519CA320A3
6194096AB2BC2A3F
5716F919EBF3218B
46CFDB09800A1647
8B9C884FEF06EB33
95E2B33B4D6AB20F
496B7C382141B79B
603779EB9E57FF97
9377CCE539E0F0C3
2837176F38FFF0DF
471FE6A47D8AB0AB
183E01C12484E7E7
8AFEF6EC0684153
EDD7F41ED63A36AF
FBA079B8AC6C9CBB

Ovviamente sí.

Questa è la soluzione di un problema di ottimo (complicato) (un problema importante, di cui parliamo meglio dopo).

Considero un reticolo in due dimensioni.



Sui siti ho delle variabili che possono valere ± 1 . Sui legami ho degli accoppiamenti: se due variabili sono collegate da un accoppiamento di valore $+1$ ho un premio se sono nello stesso stato, se invece la connessione è negativa ho un premio se sono in stati diversi (non si amano).

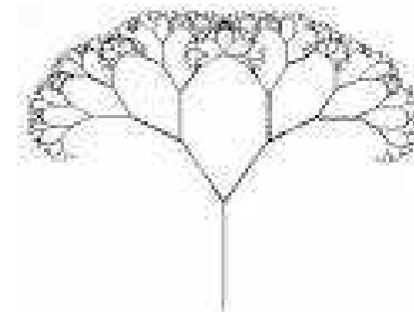
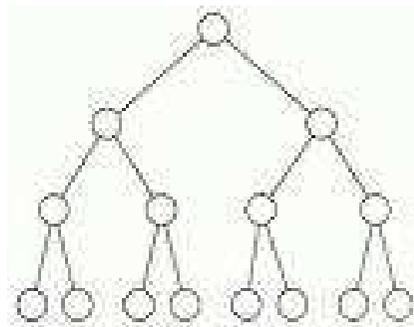
La sequenza di numeri vista prima è la soluzione per un reticolo di dimensioni $64 \cdot 18$.

Ho rappresentato lo stato -1 con il bit di valore zero, lo stato $+1$ con il bit di valore 1, ho costruito dei numeri (proprio come nel caso dell'umore giornaliero) e li ho mostrati.

Disclaimer etico: quello sull'umore era solo un esperimento simulato, creato con dei numeri random...

Complessità di un sistema fisico sarà qui in sostanza il fatto che **il sistema possa trovarsi in uno di tanti stati (stabili o metastabili) non legati da connessioni evidenti o facili da scoprire.**

Questi stati possono essere **organizzati in modo sofisticato**: ad esempio la soluzione della teoria di campo medio dei vetri di spin (un prototipo formidabile di sistema complesso) stabilisce che gli **stati vivono su un albero genealogico, e sono organizzati in una struttura ultrametrica.**





La complessità è possibilità di **organizzare una tassonomia**,
o, in altre parole, è **possibilità di conoscenza**.

Ovviamente sistemi evidentemente complicati è piú facile che esibiscano comportamenti complessi (nel senso che abbiamo definito):
gli esseri umani sono un esempio tipico.

Ci interessa (e possediamo) una **definizione di complessità stringente**:
questo è il ruolo ed il contributo delle **scienze esatte**, qui della **fisica**.

Un primo esempio. È spesso piacevole **leggere prima un libro** (grande spazio di fantasia e di complessità) e **guardare poi il film** (riduzione della complessità: scelta degli attori, degli scenari...): il contrario non funziona, perchè il film, riducendo la complessità, toglie piacere alla lettura del libro.

Un secondo esempio: **software open source in opposizione a uno sviluppo segreto, proprietario**.

Lavori recenti: siamo in grado di quantificare bene, e spesso di calcolare analiticamente, cosa voglia dire “molti stati”.

Entropia configurazionale.

Lavoro alla Sapienza, Dipartimento di Fisica e SMC-INFM:
Andrea Cavagna, Irene Giardina, Giorgio Parisi, Tommaso Rizzo.

In tutto un gruppo di sistemi di grande complessità il numero di “stati” presenti per una data scelta degli accoppiamenti o dei vincoli del sistema, cresce, quando il numero N di componenti del sistema cresce, esponenzialmente:

$$e^{N\Sigma}$$

e il valore tipico di Σ può essere calcolato matematicamente.

Le simulazioni numeriche e il metodo Monte Carlo

Come imparare numericamente qualcosa su un sistema fisico molto complicato?

Cosa voglio calcolare?

Probabilità di una certa configurazione del sistema.

E cioè?

Una somma su tutti i possibili valori delle variabili del sistema con un peso che è proprio la loro probabilità.

Una somma, e cioè un'integrale.

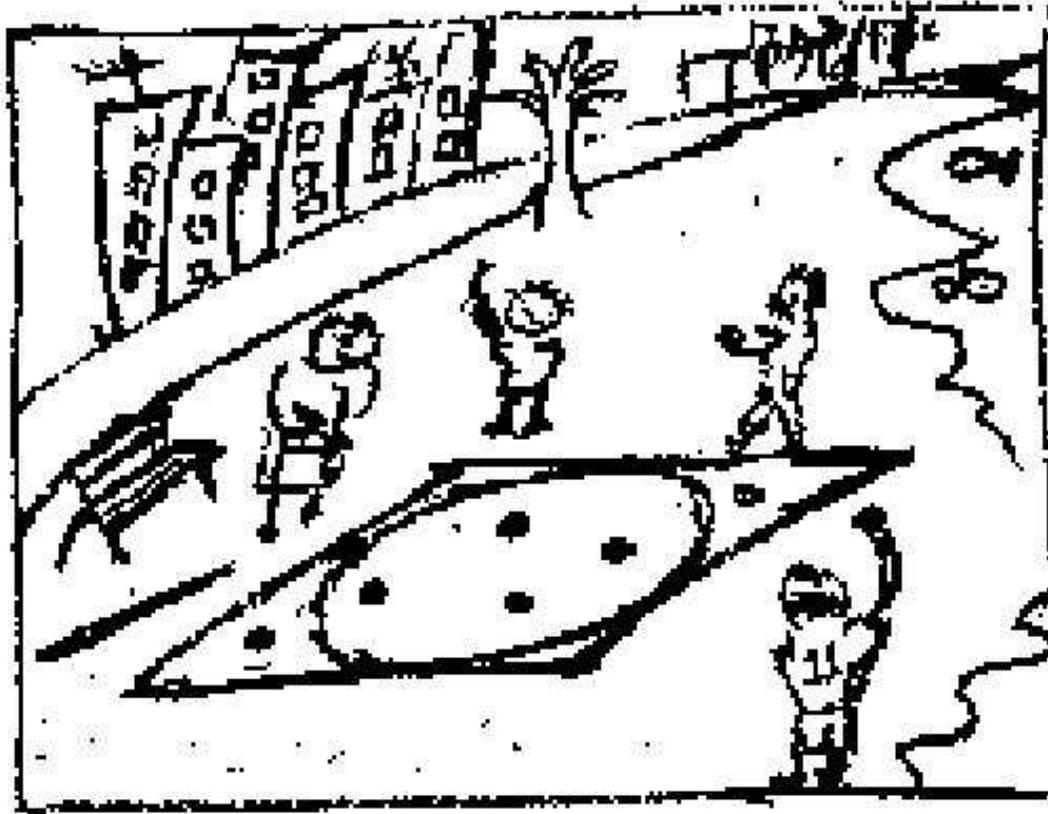
Come si calcola un'area (anche di una superficie multidimensionale)?

Il metodo Monte Carlo si basa sull'uso di numeri random, e su un campionamento aleatorio dello spazio di integrazione.

Quando bisogna integrare su molte variabili, e cioè quando la dimensionalità del problema è alta, il metodo è **molto più efficiente** dei classici schemi di integrazione deterministici.

Calcolare degli **integrali** vuol dire per noi calcolare dei “**valori di aspettazione**” nel nostro sistema potenzialmente complesso: in un modello di mercato finanziario potremmo chiederci **quanti agenti in media vogliono vendere le loro azioni**, o in un materiale **qual'è la temperatura raggiunta in un certo momento**.

I ragazzi a Monte Carlo prediligono giocare nell'eliporto dei loro genitori.



(Grazie a Werner Krauth per l'esempio e per il contributo artistico)

Calcolare π non richiede, in realtà, il metodo Monte Carlo: ma i ragazzi di Monte Carlo sono spesso, e purtroppo, piú ricchi che studiosi.

Se invece voglio calcolare un'area piú complicata (quella di un laghetto) tirare le pietre è un ottimo metodo (la forma non conta).

Purtroppo però il metodo dei bambini di Monte Carlo (metodo Monte Carlo statico) funziona solo in casi molto semplici: appena il peso da usare è complicato diventa impossibile simulare in questo modo una corretta procedura di integrazione.

In questo caso si usa una **catena dinamica (un processo di Markov)**: si sposta lo stato del sistema con dei passetti nello spazio di integrazione, e si sceglie un modo di camminare che ci garantisce di ricostruire, dopo abbastanza passi, i corretti pesi, e cioè le proprietà di equilibrio del sistema.

Caratteristiche fondamentali per il processo dinamico che porta all'equilibrio.

Ergodicità: devo poter andare dovunque.

Proprietà del bilancio dettagliato (basterebbe meno).

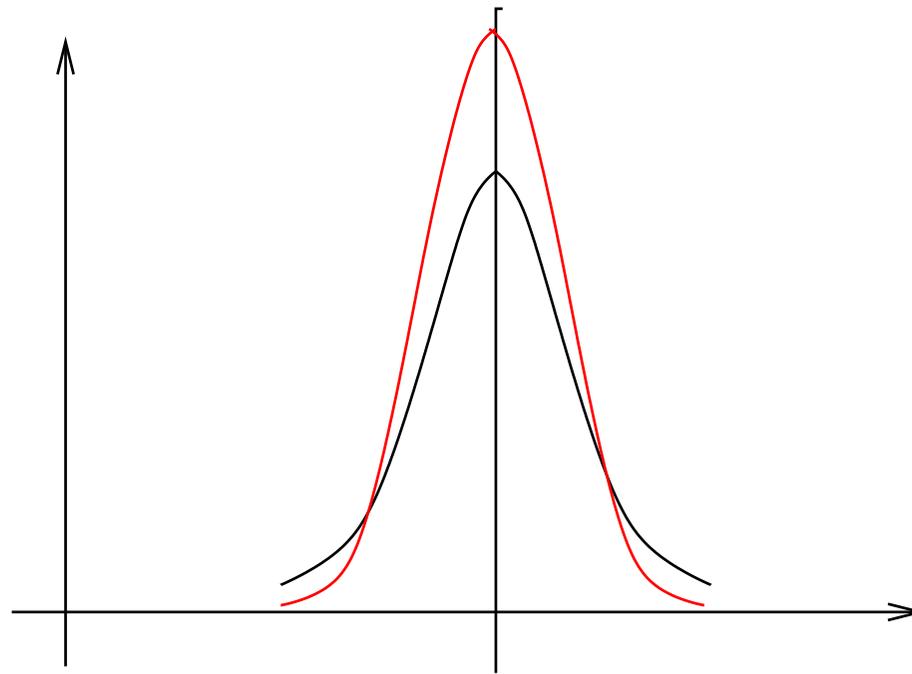
Il peso che deve avere all'equilibrio lo stato A moltiplicato per la probabilità di andare da A a B deve essere uguale al peso che deve avere all'equilibrio lo stato B moltiplicato per la probabilità di andare da B a A :

$$p(A)\pi(A \rightarrow B) = p(B)\pi(B \rightarrow A) .$$

Sotto queste condizioni la simulazione numerica arriva asintoticamente a stimare correttamente le grandezze interessanti.

Il problema è quanto tempo ci vuole perchè questo succeda.

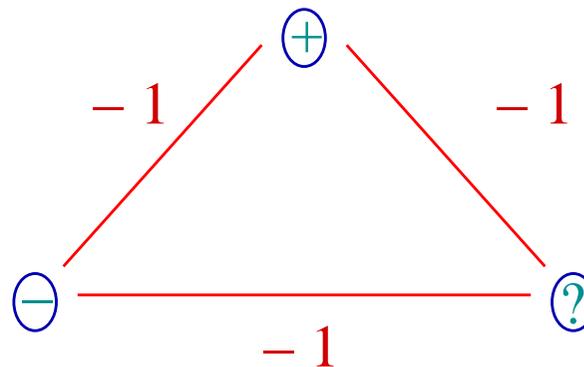
Il grande vantaggio nei tipici sistemi “di bassa complessità” è che basta ispezionare una porzione molto ristretta dei valori che le variabili possono prendere: in figura la probabilità di una grandezza misurabile (può essere falso nei sistemi complessi).



La frustrazione

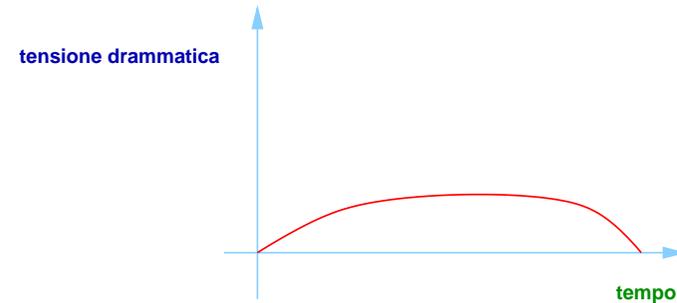
Un meccanismo fondamentale, alla base dell'apparire delle complessità nei sistemi fisici, è quello della **frustrazione**. In molte situazioni è impossibile per il sistema soddisfare tutti i vincoli che costruiscono l'interazione.

La frustrazione rende impossibile raggiungere una situazione ottimale e crea degenerazione. (degenerazione qui non è una cosa brutta: vuol dire molti stati...)

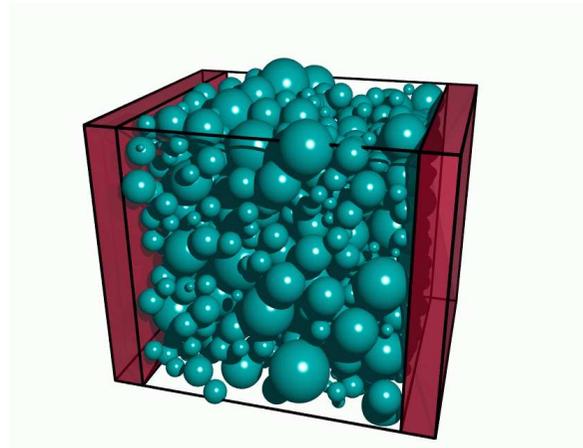


Sia in **materiali interessanti** (materiali diluiti in cui il segno dell'interazione cambia con la distanza)
che nella **vita sociale**.

Shakespeare: **tensione drammatica:**



Frustrazione statica o dinamica, con o senza disordine.



Le sfere dure:

Dal sito web di Nigel Wilding

Abbiamo già introdotto i **vetri di spin**: prototipo di materiale “complesso”.

Schematizzati da: **modello reticolare, interazione a primi vicini negativa o positiva (con uguale probabilità)**. Frustrazione aleatoria.

Questi modelli, nei limiti in cui possono venire risolti analiticamente, hanno **comportamenti (generati dalla presenza di frustrazione e quindi di complessità) del tutto nuovi** e molto diversi dai comportamenti di modelli fisici usuali.

- Stati organizzati su un albero ultrametrico.
- Alcune grandezze non sono automedianti: possono esistere stati di equilibrio che come configurazioni sono molto diversi l'uno dall'altro. La distribuzione di probabilità non diventa stretta per grandi volumi.
- Complessi effetti di memoria.

Stato di energia minima delle variabili di un vetro di spin: è anche uno dei prototipi di problema della classe “NP-completa”.

Teoria delle complessità computazionale.

Quanto costa risolvere (il caso peggiore di) un problema quando la sua taglia N cresce?

Classe P : poly in N (su una macchina di Turing deterministica).

Classe NP : poly su una macchina di Turing non deterministica (infiniti computer in parallelo). In altre parole: la soluzione può essere verificata in tempo poly su una macchina di Turing deterministica.

Insieme “ NP -completo”: insieme di problemi “equivalenti”. Si può passare in tempo polinomiale dall’algoritmo per risolverne uno a quello per risolverne un altro.

Vetro di spin interessante anche come problema di ottimizzazione.

Altro tipico NP : ricerca di circuiti Hamiltoniani.

I **vetri strutturali** (materiali interessanti e non compresi) hanno molto a che fare con il tipo di problemi che stiamo descrivendo, cioè i vetri di spin.

L'assenza di disordine congelato è compensato da una frustrazione (geometrica) auto-indotta dinamicamente.

Lo **stato amorfo**, non liquido né cristallo.

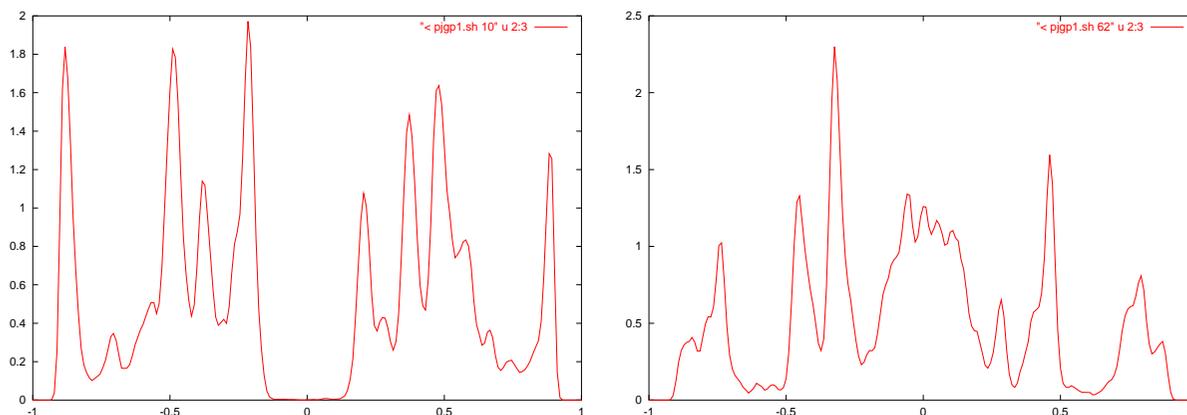
L'applicazione delle teoria della complessità ha quindi una rilevanza ancora maggiore per problemi fisici di grande interesse.

Simulazione numerica: cosa misuriamo?

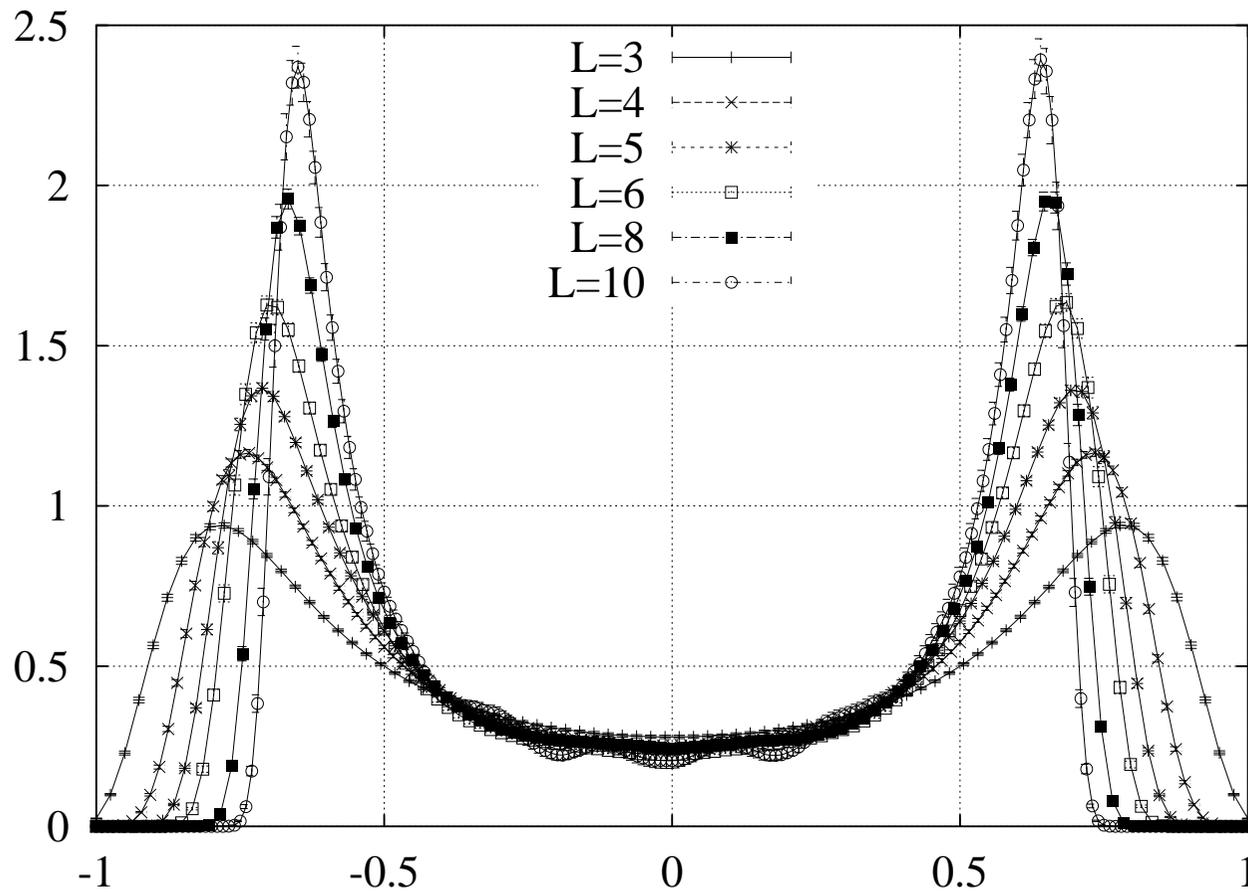
In questi sistemi ha un ruolo fondamentale q , l'overlap (o meglio, la somiglianza) fra due configurazioni di equilibrio.

Simulazione di due copie del sistema:

- le porto all'equilibrio;
- le confronto tempo per tempo;
- faccio l'istogramma e ottengo una stima della probabilità (notate che l'asimmetria fornisce una stima dell'errore).



Media su varie istanze del disordine (non diventa un picco stretto per grandi volumi):



Le simulazioni numeriche sono necessarie per studiare sistemi complessi, dove è faticoso, anche se a volte possibile, ottenere risultati analitici forti.

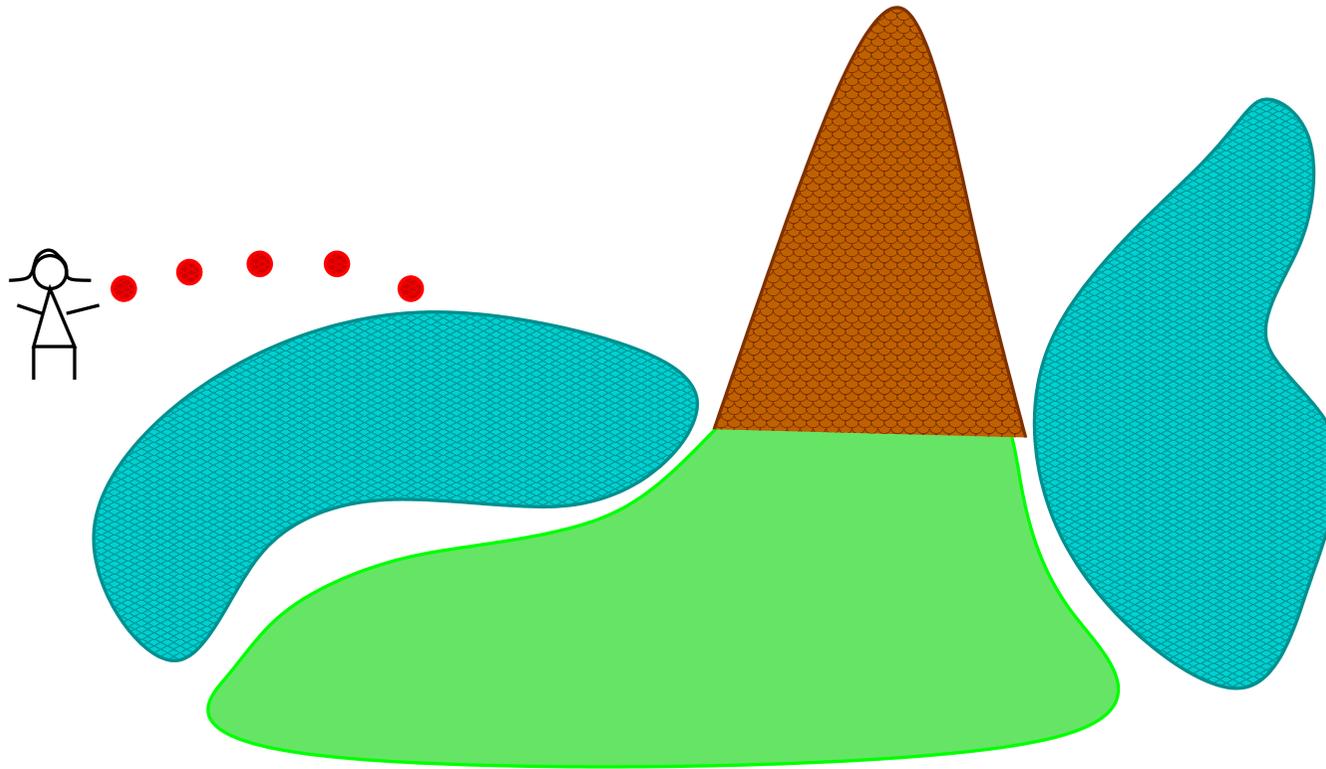
Ma sono difficili a farsi.

Comunque: sinergia fra approccio analitico, numerico e algoritmico.

Nei sistemi complessi il metodo Monte Carlo usuale non funziona proprio, nemmeno per piccoli campioni. Vedremo che questo è collegato all'esistenza di alte barriere energetiche.

La dinamica del metodo Monte Carlo può intrappolarsi in delle buche.

Torniamo al nostro eliporto. I bambini piú ambiziosi della verde Monte Carlo vogliono tirare pietre per misurare l'area di due laghetti che sono nel campo da golf di papà.



Ma ci sono delle barriere, e ci vuole molta, troppa forza per campionare allo stesso tempo i due laghetti.

Ci sono barriere energetiche.

Nei termini del nostro problema di ottimizzazione **esistono soluzioni molto convenienti di tipo molto diverso** (ricordate la distribuzione di probabilità della somiglianza che abbiamo visto prima), e **per passare dall'una all'altra con continuità devo passare attraverso configurazioni molto cattive**.

Dinamica lenta. Convergenza molto più lenta degli usuali decadimenti esponenziali.

La funzione esponenziale ha un ottimo ufficio stampa: tutti ne parlano per la crescita, pensando al furbo contadino che, senza vergogna, appoggia chicchi di grano su una scacchiera. Ma quando si prende l'inverso l'esponenziale decresce **molto** rapidamente: **in TV la ghigliottina, 100.000E, zac, 50.000, zaczac, 25.000, e con quattro soli errori ne restano poco più di 6000, non certo una cifra enorme.**



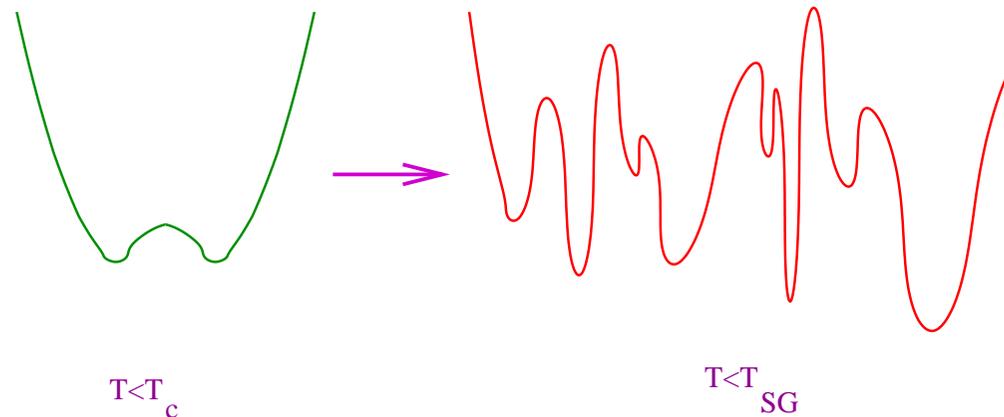
Questa severità della funzione esponenziale è alla base della località temporale in fisica: il nostro respiro di oggi non ha grandi effetti sul mondo fisico dell'anno 2206. **Nei sistemi complessi per colpa delle barriere energetiche abbiamo spesso convergenze lente, dove le correlazioni non decadono così velocemente come un'esponenziale.**

Metodi MC ottimizzati per situazioni difficili, con barriere alte.

Il tempering (E.M. e G. Parisi) e molte altre idee (multicanonical, umbrella sampling, parallel tempering).

L'altezza delle barriere dipende dalla temperatura del sistema. A noi interessa il comportamento a temperature basse, dove il sistema ha, appunto, barriere alte (ed è congelato). Si tratta di una legge di Murphy.

Sistema complesso: molte barriere alte.



Il Monte Carlo usuale si intrappola nelle buche.

Permettiamo dunque al sistema di cambiare la sua temperatura.

Invece di avere una dinamica a temperatura fissa esploriamo lo spazio delle temperature.

Le figure viste prima per la probabilità delle somiglianze sono ottenute con questo metodo.

La regola dinamica va scritta in modo tale che raccogliendo i dati ottenuti ad un dato valore della temperatura si ottenga il giusto risultato di equilibrio.

T alte: le barriere spariscono. Ritornano quando scendo a T basse.

Questi metodi funzionano in modo eccellente e permettono di imparare molto.

È vero quindi, nel nostro computer ci sono mille mondi.

Abbiamo parlato abbastanza vetri di spin e materiali amorfi.

- minority games
- ottimizzazione e SAT, problemi di soddisfacimento
- internet
- dinamiche sociali
- dinamiche di popolazione e di linguaggi
- reti genetiche
- sistema immunitario
- RNA folding
- etc...

I minority games

Una sera, in un bar a Santa Fe: il bar di El Farol (Brian Arthur).

Bella musica, buoni tacos e burritos.



Però spesso c'è troppa gente, e non è più divertente.

Se ci sono cento potenziali clienti: scegliere una strategia che ti ci faccia andare quando ci sono meno di 60 clienti.

Se sei in minoranza, quindi, vinci (è lo stesso se vuoi vendere o comprare azioni).

Molte generalizzazioni, grandi prospettive: per esempio fondamentalisti e speculatori, e simili (alla Sapienza ad esempio A. De Martino, I. Giardina, A. Tedeschi, M. Virasoro)

I problemi di soddisfacimento, SAT



3-SAT: N variabili booleane, che possono valere 0 (falso) o 1 (vero).

M clausole: vincoli sul sistema

Ogni clausola richiede che almeno una di un tripla di variabili (o la sua negazione) sia vera.

(condizione 1) O (condizione 2) O NOT(condizione 3)

I NOT sono la negazione e possono esserci o no.

Problema di ottimizzazione: determinare una configurazione delle variabili che minimizzi il numero di clausole false (e che possibilmente le porti a zero).

Il problema può essere scritto in modo da somigliare molto a un vetro di spin, e da essere trattato in modo conveniente mediante simulazione numerica (anche con iterazioni di procedure di tipo “message passing”).

Un ottimo algoritmo, utile per lo studio di grandi problemi concreti, proviene proprio dallo studio della meccanica statistica di sistemi disordinati e complessi.

Internet è una rete complessa sulla quale succedono miriadi di cose.

Uno dei problemi tipici: individuare gruppi, cluster e comunità, fisiche e semantiche.

Simulare numericamente questo mondo vuol dire cercare di comprenderne le caratteristiche fondamentali (fare esperimenti veri non è semplice). Robustezza, affidabilità, direzioni di espansione da perseguire.

Ad esempio: individuare i circuiti chiusi sulla rete (loops). Anche qui possono essere introdotti metodi numerici potenti e molto utili, message passing, Monte Carlo (E.M., R. Monasson, G. Semerjian, V. Van Kerrebroeck).