

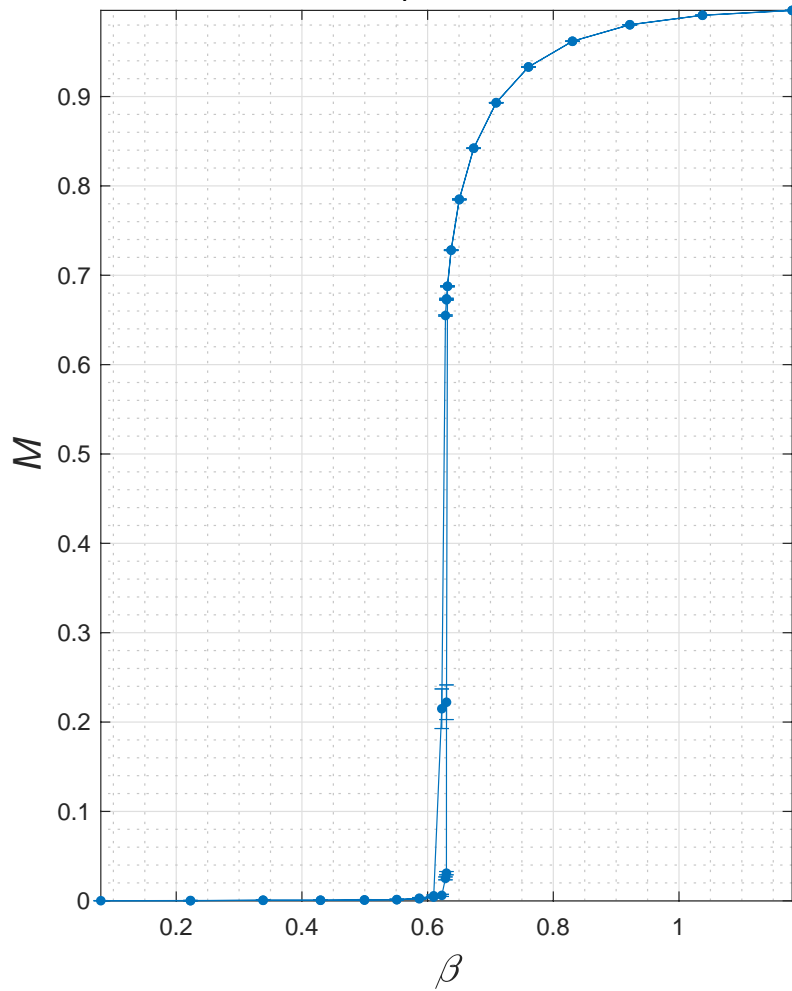
UNIVERSITÀ DI PARMA

Enrico Onofri

Lezioni di

METODI MONTE CARLO DELLA FISICA

Magnetization of $Potts_4$ on a periodic 64^3 lattice




© Università di Parma, A.A. 2023–2024
Seconda Edizione (aggiornata Gennaio 2024)
Disponibile su <https://eoinfnpr.it>

Enrico Onofri
affiliato a I.N.F.N., Sezione di Milano Bicocca
Gruppo Collegato di Parma
c/o Dipartimento di Scienze Matematiche, Fisiche e Informatiche
Università di Parma
43121 Parma, Italy

È con grande piacere che ringrazio il Prof. **Giovanni Cicuta** per interessanti discussioni su Statistica, Teoria della Probabilità e sulla teoria delle *Matrici Random*. Ho trovato molto spesso aiuto e ispirazione da discussioni con il Prof. **Claudio Destri** su problemi di carattere generale e su tematiche riguardanti lo sviluppo di algoritmi nel linguaggio `matlab`. Da un punto di vista tecnico, i miei ringraziamenti vanno a **Donald E. Knuth** (T_EX + algoritmi), **Leslie Lamport** (L_AT_EX), **Richard Stallman** (Gnu emacs), e agli sviluppatori di *MacTex*, di *PDF Expert* e *Homebrew* che hanno reso facile mettere queste lezioni in formato adatto per il web. Per i problemi che hanno richiesto calcolo numerico ci siamo avvalsi del linguaggio `matlab` (THE MATHWORKS™) e, per calcolo simbolico, di *Mathematica* (WOLFRAM RESEARCH) e *Form* di *J. Vermaseren*.

Nota. In copertina: *La magnetizzazione del modello di Potts a quattro stati in tre dimensioni studiata con il metodo di Metropolis. Si veda il Cap. 4.*

Typeset with $\mathcal{A}\mathcal{M}\mathcal{S}$ -T_EX, Euler fonts. Copying and printing this document is granted to anyone for scholar purposes, provided its content is not modified in any way.

Composto su  utilizzando TeXLive nel formato `amsbook`, utilizzando packages addizionali `amsmath`, `amssymb`, `amsthm`, `bbm`, `bbding`, `booktabs`, `dirtytalk`, `epigraph`, `euler`, `eurosym`, `fancybox`, `fancyhdr`, `fancyvrb`, `ifsym`, `resize`, `imakeidx`, `manfnt`, `marginnote`, `MnSymbol`, `mathtools`, `subfigure`, `hyperref`, `xcolor`.

Indice

	Pag.
Premessa	7
Capitolo 1. Introduzione	9
1.1. Caso e probabilità in Fisica	9
1.2. Bibliografia ragionata	11
1.3. Convenzioni adottate nel testo	12
Ringraziamenti	12
Capitolo 2. Le basi matematiche della Probabilità	13
2.1. Definizioni	13
2.2. Primi esempi	15
2.2.1. Gioco del Poker	15
2.2.2. Campionamento	17
2.3. Esercizi elementari	20
2.3.1. Un problema con i calzini	20
2.3.2. Un problema di matrici casuali	24
2.3.3. Problemi di geometria e di aritmetica aleatorie	24
2.3.4. Giochi al Casinó	26
2.3.5. Problemi di code	30
2.3.6. Alcuni problemi “classici”	36
2.3.7. Problemi con la crittografia	38
2.3.8. Biglie quantistiche	41
2.4. Sviluppi del formalismo	43
2.4.1. Indipendenza statistica	43
2.4.2. Probabilità condizionata e la formula di <i>Bayes</i>	44
2.4.3. La formula di inclusione/esclusione	49
2.5. Variabili aleatorie	54
2.5.1. Funzione di distribuzione	54
2.5.2. Variabili (statisticamente) indipendenti	56
2.5.3. Funzione generatrice	56
2.5.4. Somma di variabili aleatorie	60
2.5.5. La disuguaglianza di Chebyshev	62
2.6. La legge dei grandi numeri	63
2.6.1. Il Teorema di De Moivre-Laplace	63
2.6.2. Distribuzione di <i>Poisson</i>	65
2.6.3. Decadimento radioattivo	67
2.6.4. Teorema del Limite centrale	68
2.6.5. Non universalità della distribuzione normale	70
2.6.6. Successioni di eventi - Lemmi di <i>Borel-Cantelli</i>	70

2.7. Analisi statistica di un set di dati con matlab	71
<i>histfit</i>	72
<i>fitdist</i>	72
<i>fitgmdist</i>	72
2.8. Probabilità quantistica: il teorema di Bell	73
Capitolo 3. Processi di Markov	81
3.1. Il modello di Ehrenfest	81
3.1.1. La distribuzione asintotica nel modello di Ehrenfest	81
3.2. Catene di Markov	81
3.2.1. Analisi spettrale	83
3.2.2. Ritorno al punto di partenza	85
3.3. Processi di Markov in tempo continuo	92
3.3.1. Proprietà di convergenza	93
Capitolo 4. Il metodo Monte Carlo	99
4.1. Come realizzare successioni casuali	100
4.2. Il metodo degli “alias”	101
4.2.1. Un’applicazione elementare	103
4.3. Montecarlo dinamico	106
4.3.1. Bilancio dettagliato e l’algoritmo di Metropolis	107
4.3.2. Il modello di Ising in una dimensione	109
4.3.3. Il modello di Potts	109
4.3.4. Esperimenti numerici - modello di Ising	110
4.3.5. Vegas	112
4.4. Equazione di Langevin	112
4.4.1. Metodi di soluzione dell’equazione di Langevin	113
4.4.2. Equazione di Langevin e Meccanica Quantistica	115
4.4.3. Un esperimento nel caso più semplice: LangHO	117
4.4.4. Il calcolo della correlazione $\langle x(t) x(s) \rangle$	118
4.4.5. Galleria Montecarlo	120
Appendice A. Complementi	133
A.1. Generazione di successioni di numeri pseudo-casuali	133
A.2. Distribuzioni di probabilità elementari	133
A.2.1. Distribuzione di Bernoulli	134
A.2.2. Distribuzione binomiale	134
A.2.3. Distribuzione di Poisson	134
A.2.4. Distribuzione di Gauss	134
A.2.5. Distribuzione di Gauss multidimensionale	134
A.3. Il metodo del bootstrap	135
A.4. La distribuzione di Poisson nel decadimento α	135
A.5. I teoremi di Gershgorin e Perron-Frobenius	136
A.6. Soluzione esatta del modello di Ehrenfest	137
A.7. Calcolo del tempo medio di ritorno	139
A.8. Processi diffusivi	141
A.9. Dimostrazione della Eq.(66), correlazione quantistica	144
Appendice B. Codici matlab	145
B.1. Probabilità applicata al poker	145
B.2. Il problema del botteghino	146

B.3.	Il problema di De Mère	147
B.4.	Il problema dei “derangements”	148
B.4.1.	Subfactorial	148
B.5.	Il problema di Buffon	149
B.6.	Algoritmi per le equazioni differenziali stocastiche	149
B.6.1.	Il codice Langevin/Parisi per le correlazioni	150
B.6.2.	Il codice <code>form</code> per lo sviluppo perturbativo	154
B.7.	Codici per matrici random	155
B.7.1.	Matrici random reali simmetriche	155
B.7.2.	Matrici random Hermitiane	155
B.7.3.	Matrici random unitarie	156
B.7.4.	Matrici random a valori interi	156
B.7.5.	Matrici random stocastiche	156
B.8.	Esercizi	157
B.9.	Soluzioni	158
Appendice C.	Galleria di ritratti	161
Appendice.	Indice analitico	169
Appendice.	Bibliografia	173

Premessa

Pure mathematics is, in its way, the poetry of logical ideas

Albert Einstein

Non esprimerti mai in forma più chiara dei tuoi pensieri

Niels Bohr

Queste lezioni intendono fornire agli studenti del corso di laurea di primo livello nella classe “Fisica” alcuni concetti base di Teoria della Probabilità con l’obiettivo di acquisire alcuni metodi di calcolo che si sono rivelati molto utili nella ricerca in Fisica, sia teorica che sperimentale, e che hanno vaste applicazioni anche al di fuori dello stretto ambito fisico, i cosiddetti “metodi Monte Carlo”. Il carattere delle lezioni sarà molto pratico: l’accento è posto sulle tecniche matematiche adatte per affrontare efficacemente un dato problema e sulle simulazioni mediante l’uso del computer, piuttosto che sull’analisi matematica più astratta. Questa impostazione è suggerita dal fatto che una presentazione minimamente completa dal punto di vista matematico della Teoria della Probabilità assorbirebbe facilmente tutto il tempo a disposizione (circa 50 ore di lezione) senza riuscire ad arrivare a esplorare problemi realistici. La possibilità di simulare numericamente situazioni anche complesse con il minimo sforzo permette invece di addentrarsi rapidamente nelle problematiche delle variabili aleatorie (*random variables*), delle catene di Markov, delle camminate aleatorie (*random walks*), dei processi diffusivi retti da equazione differenziali stocastiche¹. È chiaro che per loro natura le simulazioni non permettono mai di raggiungere risultati sicuri al 100%, ma forniscono in ogni caso una buona guida per congetturare il risultato cercato, e nella stragrande maggioranza dei casi realistici rappresenta l’unica risorsa disponibile (oltre all’esperimento reale) in mancanza di un approccio puramente analitico.

Per trarre profitto da queste lezioni sono richiesti alcuni prerequisiti nella preparazione degli studenti. Questi comprendono le nozioni di Analisi Matematica acquisite nel primo biennio dei corsi di Fisica, Matematica o Ingegneria, i concetti base di Analisi Lineare (matrici, teoria spettrale) e un minimo di capacità di comprendere e sviluppare autonomamente un codice numerico in un linguaggio come `matlab` che sarà utilizzato ampiamente in questo corso.

¹Il termine *stocastico* è derivato dal Greco $\sigma\tau\acute{o}\chi\omicron\sigma$ ed è utilizzato quasi come sinonimo di *casuale* o *probabilistico*.

CAPITOLO 1

Introduzione

1.1. Caso e probabilità in Fisica

In quali circostanze parliamo di “caso” e basiamo perciò le nostre previsioni su valutazioni di probabilità? Rimanendo nel contesto della Fisica Classica, il caso fa la sua comparsa in fenomeni che sono condizionati da un grande numero di fattori imponderabili, troppo numerosi per potere essere tenuti in conto in modo sistematico, ovvero troppo difficili da misurare contemporaneamente; con l’approccio probabilistico si è imparato a sfruttare il grande numero di questi fattori per prevedere, almeno entro certi limiti, i possibili risultati delle osservazioni, assegnando a ciascuno una valutazione di plausibilità. I casi più comuni e familiari sono dati da fatti legati al gioco: l’uscita di un numero al lotto, il lancio dei dadi, il gioco della roulette etc. In tutti questi casi, il meccanismo che porta alla vincita o alla sconfitta cadrebbe in linea di principio sotto le leggi della meccanica (di Newton) se fossimo in grado di tenere sotto controllo la posizione e la velocità iniziale di rotazione della roulette, la velocità e la posizione iniziale della pallina e il suo movimento rotatorio (l’“effetto” di spin conferito dal croupier), lo stato della superficie della roulette che modifica l’aderenza, l’attrito e la presenza di microcorrenti d’aria, le variazioni di temperatura e umidità, l’attrito del perno della roulette che determina il tempo di rotazione (cos’altro?). Bene, nessuno di questi fattori è alla portata delle misure del giocatore istante per istante (e, per fortuna, per la maggior parte, *neppure del croupier*) e perciò il massimo che si possa concludere è che per ragioni di pura e semplice simmetria ogni numero della roulette è *a priori ugualmente probabile*, ossia che ci aspettiamo che in una lunga serie di rotazioni della roulette nessuno dei numeri esca più frequentemente degli altri in modo sistematico¹. Allo stesso modo l’uscita del lancio di due dadi o i numeri al lotto sono esempi di eventi assolutamente imprevedibili in quanto determinati da numerosi fattori e da condizioni iniziali non controllabili. Non ci addentreremo nella discussione su quale sia la definizione corretta di probabilità. Spesso si è caduti nella trappola di definire la probabilità in modo logicamente circolare. Perciò adotteremo un punto di vista molto pragmatico². La probabilità $\mathbb{P}(\mathfrak{E})$ di un evento \mathfrak{E} è un numero compreso tra zero e uno e sulla base della conoscenza di questo numero siamo pronti a scommettere che in N esperimenti preparati nelle stesse condizioni l’evento si produrrà approssimativamente $N \times \mathbb{P}(\mathfrak{E})$ volte, ossia $\mathbb{P}(\mathfrak{E})$ è una stima a priori della *frequenza* relativa di un evento casuale in una lunga serie di esperimenti³. L’impostazione di un problem di probabilità è dunque nei casi più semplici fondata su alcuni passaggi fondamentali: i) *l’individuazione degli elementi che contribuiscono alla definizione del fenomeno (eventi*

¹Utilizziamo il termine “probabile” dandolo per intuitivo e appartenente al linguaggio comune. Nel seguito il concetto di “probabilità” e “probabile” verrà chiarito.

²Ci ispiriamo a questo riguardo a quanto espresso da *Marc Kac* nella introduzione al suo libro [Kac59]: “... and so we shall not even attempt to state what probability theory is. Instead we shall show what it does”.

³Questa impostazione è detta *Luglio frequentista* e non è accettata da molti studiosi che si ispirano all’approccio “Bayesiano” [D’A03]. Ritorneremo su questo aspetto della questione che riguarda peraltro principalmente i fondamenti filosofici della teoria.

elementari); ii) *l'assegnazione a questi eventi elementari una misura di probabilità \mathbb{P}* ; iii) *l'utilizzo dell'analisi combinatoria e dell'algebra degli insiemi (unione, intersezione) che permette di valutare la probabilità di un evento complesso una volta ridotto ai suoi elementi*. Allora, piuttosto che analizzare in modo rigoroso le basi dell'approccio probabilistico (Laplace, Poincaré, De Finetti, Kolmogorov, Feller) procediamo adottando lo schema matematico della probabilità come ipotesi di lavoro. Le conferme che arriveranno dalla sua applicazione a problemi concreti permetteranno allora di convincerci della validità dell'approccio ⁴.

Prima di iniziare, un paio di osservazioni che sono per me molto importanti. Innanzitutto dobbiamo realizzare il fatto che il manifestarsi del caso in fenomeni fisici non è legato esclusivamente a quanto detto finora (la presenza di numerosi fattori imponderabili). Sappiamo ormai per certo che la fisica a livello microscopico, a partire dalle molecole e su scale inferiori, è caratterizzata da comportamenti casuali non riconducibili all'idea che esistano *agenti sub-microscopici e/o non misurabili*; il caso cioè sembra entrare nelle leggi fisiche in modo intrinseco, anche se questo fatto non è mai stato accettato dal grande Albert. L'emissione spontanea di radiazione dagli atomi e molecole, l'effetto tunnel, il decadimento radioattivo nei nuclei, i fenomeni di diffrazione e interferenza di particelle materiali, sono tutti fenomeni descritti dalla Fisica quantistica in termini probabilistici, ma è ormai escluso che la probabilità entri in questo campo di fenomeni per una nostra mancanza di informazione dettagliata sulle variabili fisiche in gioco. Gli esperimenti condotti al riguardo nell'arco degli ultimi cinquant'anni hanno mostrato che certe misure di correlazione sono incompatibili con una impostazione probabilistica in senso tradizionale (*disuguaglianze di Bell* [Bel64]). Il fatto fondamentale, ormai basato su esperimenti molto affidabili, consiste nella impossibilità di configurare la probabilità a livello quantistico in termini di “spazio di eventi elementari” come nell'impostazione della teoria della probabilità dovuta a Laplace, Poincaré e Kolmogorov ⁵. Non ci occuperemo di probabilità in ambito quantistico in queste lezioni, tranne che per un cenno alla fine del secondo capitolo (vedi la Sez. 2.8,) ma la problematica è molto interessante e tuttora rappresenta un campo di ricerca attiva con diramazioni verso la tecnologia (crittografia quantistica, *quantum computing*, etc.).

Una seconda osservazione è che la ragione per cui un numero sempre crescente di fisici si sono occupati di tecniche probabilistiche sta nel fatto che *queste tecniche si rivelano molto potenti nello sviluppare algoritmi di calcolo per la soluzione di problemi che di per sé non hanno nulla di casuale*. Ciò è dovuto a particolari coincidenze per cui, ad esempio, la soluzione di certe equazioni differenziali (come l'equazione di Laplace o quella della diffusione del calore) è equivalente alla soluzione di problemi di “camminate aleatorie” (*random walks*, RW per brevità). In molte situazioni è più agevole simulare un RW piuttosto che risolvere una complicata equazione alle derivate parziali. Anche il *calcolo di un integrale in un numero molto elevato di variabili* non presenta aspetti aleatori e tuttavia può risultare impraticabile con metodi classici di quadratura, mentre tecniche di simulazione “Montecarlo” sono adattabili ad una grande varietà di problemi di interesse fisico (si veda il Cap. 4).

Nei primi tempi in cui si sono resi disponibili calcolatori elettronici, circa 70 anni fa, i Fisici hanno applicato questa nuova potenza di calcolo alla soluzione di problemi inattaccabili con l'analisi matematica classica. Una prima applicazione fu quella alla simulazione delle reazioni a catena nella “pila atomica” e quindi alla dinamica dei neutroni. L'abstract di un lavoro

⁴Per chi fosse interessato ad apprendere i diversi approcci filosofici al concetto di probabilità segnaliamo il testo di Gillies [Gil00].

⁵Per una introduzione elementare a questi concetti si può consultare la Tesi triennale di B. Micciola che si trova su arxiv.org/abs/2003.08000.

apparso nel 1949 [MU49] e che segna l'inizio dell'era delle simulazioni al computer recita: “*We shall present here the motivation and a general description of a method dealing with a class of problems in mathematical physics. The method is, essentially, a statistical approach to the study of differential equations, or more generally, of integro-differential equations that occur in various branches of the natural sciences.*” Gli autori, *Nicholas Metropolis e Stan Ulam*, dettero così inizio all'uso delle simulazioni numeriche che, con l'avvento dei calcolatori veloci in questi ultimi decenni, permette oggi di affrontare problemi complessi in tutti i campi della Fisica e della Tecnologia, dallo studio delle proteine, allo studio delle interazioni forti. Alla base di tutto, la teoria matematica della probabilità. Un ingrediente indispensabile per le applicazioni della teoria a problemi di Fisica è costituito dalla disponibilità di algoritmi efficienti per la generazione di successioni di numeri casuali, sia interi che reali. Non si parla per la precisione di numeri strettamente casuali ma piuttosto “pseudo-casuali” che vengono generati da procedure del tutto deterministiche e riproducibili ma che dal punto di vista statistico presentano caratteristiche indistinguibili da vere successioni casuali. Per queste lezioni non è indispensabile approfondire questa problematica (utilizzeremo le funzioni `rand`, `randn`, `randi` di matlab “a scatola chiusa”), ma chi desideri capire come si realizzano questi algoritmi trova un'ottima introduzione nelle lezioni di *A. Pelissetto* alla Scuola di Parma del 1994 [Pel94]; copie delle lezioni sono disponibili su richiesta. Da allora si sono affermati altri algoritmi tra cui il *Mersenne Twister* [MN98] comunemente utilizzato dalla maggior parte dei compilatori di uso comune (il nome viene dal fatto che l'algoritmo è basato sul 2^{40} numero primo di Mersenne $2^{19937} - 1$, un numero di circa 6000 cifre). Questo algoritmo è utilizzato nelle versioni recenti di matlab per generare numeri casuali con le routines `rand`, `randn`, `randi`.

1.2. Bibliografia ragionata

Normalmente la bibliografia si dà alla fine. Tuttavia è giusto citare subito le fonti per queste lezioni. Mi sono basato sull'agile testo di Rozanov [Roz69], ricco di esempi. Come esercizi iniziali consiglio il libretto di Mosteller [Mos65], con esempi tratti dalla vita di tutti i giorni (più o meno). Un ottimo trattato di riferimento è quello di *Prokhorov e Rozanov* [PR69]. A un livello più elementare il testo di *Gnedenko* [Gne79]. Inoltre sulla collezione [Wax54] si trovano lavori originali che hanno fatto la storia della applicazione di concetti di probabilità alla teoria fisica (*Chandrasekhar, Ornstein-Uhlenbeck, Kac*). Il testo di *Mark Kac* [Kac59] è molto avanzato dal punto di vista matematico e presenta la connessione con la probabilità quantistica. Le problematiche connesse al ruolo della probabilità in fisica quantistica sono presentate in molti testi di meccanica quantistica. Il libro di Bell [Bel87] contiene una raccolta di scritti del fisico irlandese, alla base di molti sviluppi recenti. Ma una presentazione per non specialisti si può trovare utilmente nel libro di Penrose [Pen89], edizione italiana [Pen97]. Il trattato di *Feller* [Fel68] costituisce il riferimento standard per la teoria delle probabilità e permette di approfondire tutti gli argomenti coperti in queste lezioni. Un trattato più compatto ma ricco di esempi descritti diffusamente è quello di *Fisz* [Fis63]. Per le connessioni tra probabilità, random walk e teoria quantistica dei campi si può consultare il primo volume di Itzykson e Drouffe [ID89]. Un altro testo ricco di idee, esempi, problemi e che presenta un approccio originale alla teoria della probabilità è il libro di Backlawski, Cerasoli e Rota [BCR84]. Un testo recente con un forte orientamento all'uso della simulazione numerica, abbastanza vicino in spirito a queste lezioni è dato dal trattato di Rotondi, Pedroni e Pievatolo [RPP12], che può essere

⁶Si trovano informazioni affidabili consultando Wikipedia alle voci “Mersenne prime” e “Mersenne twister”.

conveniente consultare ma troppo esteso per un breve corso semestrale. Per l'approfondimento delle tecniche Montecarlo una buona introduzione è la raccolta [Reb83] e un testo più leggibile come introduzione è il libro di Creutz [Cre83]. Un recente testo universitario da tenere a portata di mano per questo corso è infine quello di Boffetta e Vulpiani [BV12] adatto per un corso monografico anche a livello di laurea magistrale. Infine, per chi volesse approfondire questioni di fondo riguardo la natura più filosofica che matematica della teoria delle probabilità consiglio il testo di D. Gillies [Gil00] già citato. Il problema di impostare la probabilità in termini non-frequentistici è particolarmente evidente in casi quali la meteorologia: “qual’è la probabilità che domani piova?”: è evidente che non si possono sperimentare n volte le medesime condizioni e calcolare il quoziente. Il contesto in cui ci si mette in queste lezioni è tuttavia più propriamente legato a problemi fisici in cui le condizioni al contorno siano perfettamente sotto controllo. Per applicazioni a “Fisica e Finanza” si può consultare il testo di Lax *et al* [LCM06] e per quanto riguarda l’approccio Bayesiano esistono delle lezioni di Jaynes [Jay58] disponibili su INTERNET e numerosi lavori di G. D’Agostini orientati all’applicazione dell’approccio alla Bayes all’analisi dati in Fisica delle Alte Energia (si veda il folder `~/Dropbox/MPF2023/Documents/Bayes`). Un testo recentissimo e che presenta caratteristiche di originalità è quello di Ariel Amir [Ami21]; a questo testo ci ispireremo prossimamente per introdurre qualche nuovo esercizio.

1.3. Convenzioni adottate nel testo

Utilizziamo i simboli standard di teoria degli insiemi \cup (unione), \cap (intersezione), \setminus (sottrazione), \emptyset (insieme vuoto). Il *complementare* di un sottoinsieme è indicato con una barra: se indichiamo con Ω l’insieme completo di tutti gli elementi, allora $\bar{A} = \{\omega \in \Omega | \omega \notin A\} \equiv \Omega \setminus A$. Un esercizio è corredato spesso dalla sua soluzione; il testo riprende dopo il simbolo ■.

Ringraziamenti

Desidero ringraziare gli studenti che hanno frequentato il corso negli anni 2004-2007, per gli stimoli che mi hanno fornito rendendo il compito di docente piuttosto gradevole; ringrazio in particolare gli allievi del 2006-2007 (Bertolini, Garlatti, Massari, Monteux e Parolini) che hanno anche curato la stesura dell’Appendice relativa ai processi diffusivi e gli studenti dal 2017 al 2023 (Armanini, Baistrocchi, Balzani, Bilzi, Boggio Tomasaz, Boraschi, Bossi, Caramanti, Carotti, Caruso, Celada, Chizzini, Cipressi, Dardani, Dilda, Gasparini, Gombi, Grossi, Lizzi, Mai, Marieni, Murasaki, Nervo, Piazza, Pocaterra, Sanfelici, Santonastaso, Savi, Siani, Tarani, Viappiani, Volpi, Zanni) per l’impegno e l’abilità che hanno profuso nello sviluppo di programmi Montecarlo, nell’analisi dei risultati o nell’approfondimento della problematica della probabilità quantistica.

Infine un grazie particolare all’amico Prof. Augusto Schianchi, collega del Dipartimento di Economia, per la curiosità sempre vivace che dimostra per vari problemi di Matematica e di Fisica.

CAPITOLO 2

Le basi matematiche della Probabilità

2.1. Definizioni

La Teoria della Probabilità si è sviluppata come teoria matematica a partire da Pascal e Laplace; successivamente Poincaré e altri matematici nel XX secolo le hanno dato un impulso notevole e oggi la teoria è una branca della matematica con agganci all'analisi matematica, all'analisi funzionale, alla geometria, alla teoria della misura *etc* e con applicazioni molto estese in campo scientifico in senso lato. Gli elementi sono molto semplici. Si introduce lo *spazio degli eventi* costituito da un insieme Ω i cui elementi sono detti *eventi elementari*. Ad Ω si applicano le definizioni caratteristiche della teoria degli insiemi, che nella sua forma elementare daremo per scontata. Ad una opportuna collezione di sottoinsiemi $\sigma \subset \Omega$ si attribuisce un numero non-negativo $\mathbb{P}(\sigma)$ che verrà definito *la probabilità* di σ e che deve soddisfare alcune condizioni di consistenza che vedremo poi. In sintesi la definizione assiomatica di probabilità comporta l'introduzione di una tripletta di oggetti $\mathfrak{P} = \{\Omega, \Sigma, \mathbb{P}\}$, detta *spazio di probabilità*, così definiti:

- i) Ω è un insieme detto *lo spazio degli eventi elementari*;
- ii) Σ è una famiglia di sottoinsiemi di Ω , detti *eventi*;
- iii) \mathbb{P} è una *misura* su Σ , ossia una funzione $\mathbb{P}: \Sigma \rightarrow [0, 1]$, detta *misura di probabilità* ¹.

A questa struttura si impongono alcune condizioni di consistenza:

- a) Σ è chiusa sotto operazioni di *unione* ($\sigma_1 \cup \sigma_2$) e *intersezione* ($\sigma_1 \cap \sigma_2$), anche per successioni infinite e di “sottrazione” $\sigma_1 \setminus \sigma_2 \equiv \{x \in \sigma_1 \ \& \ x \notin \sigma_2\}$.
- b) $\mathbb{P}(\sigma_1 \cup \sigma_2) = \mathbb{P}(\sigma_1) + \mathbb{P}(\sigma_2)$ se σ_1 e σ_2 non hanno elementi elementari in comune, simbolicamente $\sigma_1 \cap \sigma_2 = \emptyset$ (l'insieme vuoto), e più in generale $\mathbb{P}(\cup_i \sigma_i) = \sum_i \mathbb{P}(\sigma_i)$ se gli eventi σ_i sono a due a due disgiunti ² ($\sigma_i \cap \sigma_j = \emptyset, \forall (i, j), i \neq j$)
- c) $\mathbb{P}(\emptyset) = 0, \mathbb{P}(\Omega) = 1$

ESERCIZIO 1. In base agli assiomi qual'è il valore di $\mathbb{P}(\sigma_1 \setminus \sigma_2)$?

Se l'insieme Ω è discreto, ogni evento elementare è a sua volta un elemento di Σ e gli si assegna un valore di probabilità. Nel caso continuo invece sono gli insiemi *misurabili* in Σ ad avere attribuita una misura di probabilità ³. Ad ogni elemento dell'insieme Ω si associa un particolare accadimento che può o non realizzarsi: se Ω è l'insieme delle facce di un dado da gioco, ad esempio, l'“evento elementare” consiste nell'uscita di un numero tra 1 e 6 allorché tiriamo il dado mentre un evento in generale può essere individuato nell'uscita di un numero

¹Manterremo la notazione \mathbb{P} per la misura di probabilità il cui argomento è un qualunque insieme misurabile, mentre utilizzeremo via via notazioni quali $P(x)$, P_i etc. per indicare la probabilità come funzione di variabili discrete o continue.

²Il caso generale in cui i sottoinsiemi σ_i possono avere intersezione non vuota è coperto dalla formula di *inclusione/esclusione* di Poincaré (vedere Pag. 49).

³Ad esempio, se $\Omega = \mathbb{R}$ non si assegna una probabilità a singoli numeri, ma a sottoinsiemi misurabili di \mathbb{R} , quali gli intervalli o i pluriintervalli.

qualunque dispari. Cioè l'insieme Ω serve per parametrizzare eventi casuali, così come la coordinata cartesiana x individua un punto sulla retta. L'insieme Ω può essere costituito da un insieme finito (le facce di un dado), infinito numerabile (i punti nel piano con coordinate intere), da una varietà geometrica (ad es. un segmento di \mathcal{R} , una regione piana in \mathcal{R}^2 , una qualunque superficie \mathcal{S} , un dominio in \mathcal{R}^n , etc.)⁴. I sottoinsiemi che costituiscono Σ devono essere *misurabili*, cioè per ogni $\sigma \in \Sigma$ è definito il valore $\mathbb{P}(\sigma)$. Potremmo dire che Ω rappresenta simbolicamente i fatti osservabili cui attribuiamo una stima di probabilità riguardo il loro effettivo realizzarsi.

Un possibile punto di vista è quello detto “frequentista”, secondo cui la probabilità di un evento rispecchia quella che è la frequenza con cui l'evento accade rispetto al numero totale di osservazioni quando questo numero è molto grande. Se gettiamo un dado molte volte troveremo che un dato numero uscirà in media un sesto delle volte. La costruzione di un modello probabilistico si basa in parte sull'esperienza e in parte su argomenti a priori; la simmetria del dado ci fa prevedere quanto detto, probabilità $1/6$, e se l'esperienza ci mostra che la media si discosta significativamente da questa previsione saremo indotti a investigare sul fatto che il dado *potrebbe essere truccato*. Resta naturalmente da specificare cosa si debba intendere per “numero molto grande” di osservazioni.⁵ La cosa assumerà gradualmente un assetto soddisfacente: provvisoriamente adottiamo un approccio assiomatico con l'intesa che quanto ciò si rispecchi nella realtà dovrà essere verificato nei fatti.

Esistono altri punti di vista riguardo i fondamenti della teoria delle probabilità assolutamente rispettabili ma che sarebbe lungo illustrare in questa sede. Nell'approccio di *Ramsey-De Finetti* ad es. si definisce la probabilità di un evento da un punto di vista soggettivo - *quanto sei disposto a scommettere che un certo evento si realizzi?* - e quantunque questo atteggiamento possa a prima vista apparire un po' vago, si può precisare in termini di condizioni di consistenza al punto che, secondo un teorema dei due autori citati, si ha perfetta equivalenza con l'approccio assiomatico che adottiamo in queste lezioni⁶.

Nota. L'applicazione dello schema astratto di “spazio di probabilità” a problemi concreti (in Fisica o in qualunque altro contesto) si basa sull'assunto che gli eventi elementari di Ω siano associati ad eventi che si possono osservare e registrare e su cui sia possibile effettuare rilevazioni statistiche. Se in una serie di osservazioni l'evento ω è osservato $n(\omega)$ volte su un totale di N , il nostro modello probabilistico dovrà essere in grado di prevedere che

$$(1) \quad \frac{n(\omega)}{N} \approx \mathbb{P}(\omega)$$

La questione dell'applicabilità di questi concetti a situazioni reali passa attraverso la precisazione del significato del simbolo \approx . Ci si aspetta che la *frequenza* $\frac{n(\omega)}{n(\Omega)}$ e la probabilità $\mathbb{P}(\omega)$ tendano a coincidere quando il numero di eventi sia sufficientemente grande. Quest'idea si realizza nei vari teoremi “dei grandi numeri”. Attenzione però che ci sono pochi concetti quali la *legge dei grandi numeri* ad essere maggiormente travisati dai non esperti, come vicende legate al Lotto dimostrano ampiamente.

⁴Risulta evidente che possono esserci diverse scelte dell'insieme Ω che si applicano allo stesso caso in esame. I diversi insiemi sono in corrispondenza biunivoca e deve essere puramente una questione di convenienza adottare l'una o l'altra descrizione.

⁵Si veda il cap.2.6.6 e gli esercizi a pag.94 per approfondire la questione.

⁶Lo studente interessato può consultare il testo di Gillies [Gil00] per una presentazione molto dettagliata anche dal punto di vista storico e quello di D'Agostini [D'A03] per una trattazione approfondita dell'approccio Bayesiano nella sua applicazione all'analisi di dati sperimentali.

2.2. Primi esempi

2.2.1. Gioco del Poker.

ESEMPIO 1. Dopo avere mescolato accuratamente un mazzo di 32 carte (composto da 7, 8, 9, ..., Q, K, A in 4 semi diversi ♥, ♦, ♣, ♠), si distribuiscono le carte a quattro giocatori di poker. Qual'è la probabilità che il primo giocatore abbia un poker servito?

SOLUZIONE. Innanzitutto individuiamo in questo caso in cosa consista lo spazio di probabilità. Lo spazio degli eventi elementari è costituito da tutte le possibili “smazzate”, cioè tutti i possibili ordinamenti del mazzo di carte, in numero di $32! \approx 2.6 \times 10^{35}$. L'algebra degli eventi Σ è data da tutte le possibili collezioni di smazzate caratterizzate da tutti i possibili requisiti (ad es. quelle in cui i primi quattro giocatori hanno un poker servito, quelle in cui il gioco migliore in mano ai quattro giocatori è una coppia, etc.). La misura di probabilità nel caso di un gioco onesto assegna la stessa probabilità a tutte le smazzate possibili e utilizzando le regole enunciate nel capitolo precedente si può dedurre la probabilità di qualunque $\sigma \in \Sigma$. Nel caso previsto dall'esercizio (il primo giocatore abbia un poker servito) dobbiamo iniziare a dire che ci sono $\binom{32}{5} = 201376$ possibilità e non è consigliabile elencarle qui tutte per contare quante contengono quattro carte uguali; in effetti, se non si ha voglia di utilizzare l'“arte combinatoria”, il problema si può affrontare utilizzando un computer: si costruiscono esplicitamente tutte le possibili mani e si conta quante di queste contengono un poker. Provate a immettere queste poche istruzioni al prompt di **matlab** (ponendo molta attenzione a tutti i particolari, 'Mani' deve avere 201376 righe e 5 colonne)

```

Poker servito
Carte = [1:8,1:8,1:8,1:8]; % 32 carte in quattro semi, 1='7', 2='8', ..., 8='A'
% elencare tutte le combinazioni possibili di 5 carte
Mani = nchoosek(Carte, 5);
P=sort(Mani'); % si mettono in ordine le carte di ciascuna mano
               % es.: 7889K, 8888Q, 9JJJJ
poker1 = P(1,:) == P(4,:); % l'unica carta diversa e' la piu' grande
poker2 = P(2,:) == P(5,:); % l'unica carta diversa e' la piu' piccola
% (numero di mani con poker servito)/(numero totale di mani possibili)
prob = sum(poker1+poker2)/nchoosek(32,5);
[N,D] = rat(prob) % prob = N/D deve essere un numero razionale

```

Osservazione. Notare che la funzione di matlab *nchoosek* ha due differenti impieghi: se il primo argomento è un numero intero il valore di *nchoosek*(*n*,*k*) è semplicemente il numero di combinazioni di *n* elementi presi *k* a *k*, che si indica con $\binom{n}{k}$ (coefficiente binomiale). Se invece il primo argomento è un vettore come nell'esempio precedente in cui il primo argomento è “Carte”, allora *nchoosek* ritorna tutte le combinazioni.



Basta un secondo o poco più per esaminare tutte le 201376 mani e il risultato (1/899) coincide esattamente con quello che calcoleremo adesso usando la tecnica combinatoria, ma per problemi appena più complicati l'enumerazione completa può richiedere tempi e spazio in memoria sconfinati⁷: o si riesce ad effettuare il calcolo per via analitica oppure si deve ricorrere a un “campionamento”, come discuteremo tra poco. In questo caso semplice si può trovare facilmente un argomento combinatorio: in un poker servito la carta diversa dalle altre può essere

⁷Si veda l'Es.16

scelta in 32 modi differenti, e per ogni scelta sono possibili 7 modi per fare poker⁸. Dunque la probabilità, come rapporto tra i casi favorevoli e i casi possibili, è data da

$$32 \times 7 / \binom{32}{5} = \frac{32 \times 7 \times 30 \times 4}{32 \times 31 \times 30 \times 29 \times 28} = \frac{1}{31} \frac{1}{29} = 1/(900 - 1) = 1/899,$$

cioè poco più di un caso su mille (e non è stato necessario utilizzare nemmeno una calcolatrice...). E infatti è ritenuto fortunato un giocatore cui si presenti questa circostanza durante la serata. In generale per n carte diverse si ha $\mathbb{P}(\text{pokerservito}) = 5/\binom{4n-1}{3}$ (lasciamo la dimostrazione per esercizio).

Una domanda interessante: qual'è la probabilità che ci siano due giocatori su quattro ad avere un poker servito? (A questa domanda conviene rispondere utilizzando un metodo di campionamento come descritto nel paragrafo seguente - provare il codice `pokermult4g.m`). ◀

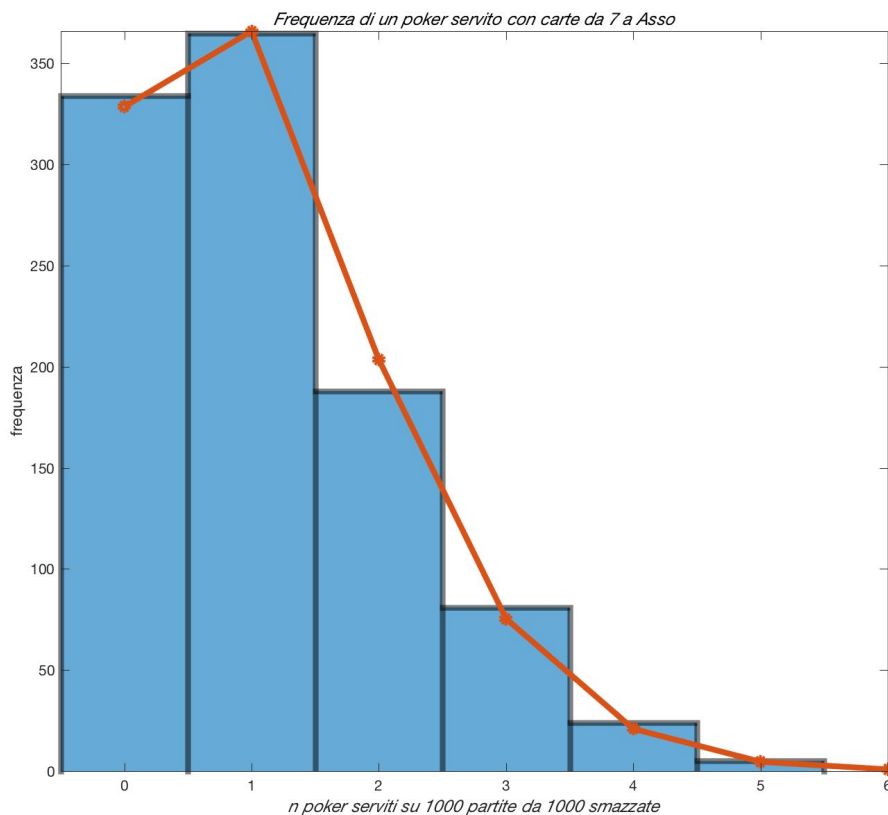


FIGURA 2.1. Simulazione del gioco del Poker: sono simulate 1000 serate di gioco con 1000 smazzate ciascuna.

ESERCIZIO 2. Una variante: qual'è la probabilità di trovarsi un “colore” servito, ossia cinque carte dello stesso seme da un mazzo di 32 carte?

⁸Alternativamente si può avere poker in 8 modi e la rimanente carta sarà una tra le 28 rimanenti

ESERCIZIO 3. Si mescolano accuratamente le carte di un mazzo da poker (32 carte). Qual'è la probabilità che in cima al mazzo ci siano tutte carte rosse (cioè le prime 16 carte siano carte di cuori e di quadri)?

SOLUZIONE. Ci sono $\binom{32}{16}$ modi di selezionare 16 carte dal mazzo, di queste scelte solo una corrisponde a tutte carte dello stesso colore rosso. Dunque la probabilità data da $16!/32! \approx 1.66 \times 10^{-9}$.

2.2.2. Campionamento. Vogliamo ora affrontare i problemi precedenti attraverso l'esame di un campione di possibili mani, anziché l'enumerazione esaustiva di tutte le possibilità: si esegue cioè una *simulazione*. Esistendo una trattazione combinatoria ⁹ è chiaro che questo costituisce giusto un primo approccio alle tecniche Montecarlo ¹⁰ in un contesto in cui possiamo verificarne l'efficacia. Avendo a disposizione un calcolatore e un linguaggio come **matlab** che permette facilmente l'estrazione di numeri a caso, possiamo simulare un numero arbitrario di smazzate e contare in quanti casi si presenta un poker servito. Il programma è dato in appendice a Pag. 145. Il risultato permette di valutare oltre il valore atteso (valore *medio*), anche la fluttuazione del fenomeno, cioè di quanto il numero osservato in ogni simulazione si scosti dal valore medio. La figura presenta attraverso un istogramma l'analisi di 1000 partite ognuna costituita da 1000 smazzate. L'istogramma è confrontato con una *distribuzione binomiale* ¹¹ con valore medio 1000/899. Il valore medio calcolato sugli eventi simulati risulta 1.11 non molto lontano da quello esatto. Tuttavia si noterà che abbiamo utilizzato un numero di eventi simulati (10^6) dell'ordine di grandezza e addirittura superiore a quello dell'intera popolazione. I casi in cui è veramente vantaggioso effettuare una simulazione sono quelli in cui lo spazio degli eventi è molto grande e il campione è scelto in modo mirato ma di dimensioni molto piccole rispetto al totale delle possibilità ¹². Potremmo esplorare allora cosa succede se simuliamo un insieme di “esperimenti” in numero molto inferiore a 200000. Ad es. prendiamo in esame l'estrazione casuale di 1000 smazzate ripetuta 20 volte (si considera un numero di “smazzate” dieci volte inferiore al totale). Si trova che la stima così ottenuta del numero di poker serviti può variare tra 0 e 5, con una incertezza ¹³ di 0.4. Si conclude che l'analisi del gioco del poker con il metodo della simulazione non è efficiente; se il campionamento è limitato a un numero piccolo, ad es. 10000, il risultato è affetto da errore statistico del 40%! Fortunatamente il metodo della simulazione Montecarlo in molte applicazioni interessanti per la Fisica non è affetto da questo problema. C'è anche da riflettere sul fatto che le forti fluttuazioni che si riscontrano nella simulazione sono un riflesso del fatto che anche nel gioco reale si hanno risultati che si discostano ampiamente dalle aspettative basate sul calcolo della probabilità, e questo contribuisce a rendere più imprevedibile, e forse più divertente, il gioco del poker...

ESEMPIO 2. Attraverso una simulazione valutare la probabilità che in una mano di poker ci siano due giocatori con un poker servito (oppure 3, o 4).

⁹Ogni problema di probabilità legato al poker è ampiamente trattato in letteratura. Ad es. si può chiedere a Wikipedia: http://en.wikipedia.org/wiki/Poker_probability.

¹⁰Un'introduzione al metodo “Montecarlo” si trova al cap. 4.

¹¹Diamo per scontato che il lettore conosce le distribuzioni statistiche elementari, Gauss, Cauchy, Bernoulli-binomiale, Poisson. Un richiamo si trova in Appendice.

¹²È questo il principio su cui si fondano i sondaggi - non avrebbe senso intervistare *tutti gli elettori*.

¹³L'incertezza della stima è data dalla *deviazione standard* (radice quadrata dello scarto quadratico medio) divisa per la radice quadrata del numero di esperimenti indipendenti. Daremo per acquisito questo concetto dalla teoria degli errori. Nel caso di variabili aleatorie si veda la sezione 2.5.

SOLUZIONE. Un codice matlab che affronta il problema si trova in `~dropbox/Problemi/Matlab/Cap.2/poker/pokermult4g.m`.

Notare che il programma non distingue il caso che un poker sia servito al primo giocatore o a uno qualunque degli altri, e di conseguenza la probabilità esatta cui confrontare il risultato della simulazione è $4/899$.

ESEMPIO 3. Si considerino i numeri reali compresi tra 0 e 1, ossia Ω consiste nell'intervallo reale $[0, 1]$. Si consideri poi la famiglia Σ costituita da tutti i possibili intervalli contenuti nel segmento $[0, 1]$ e tutte le loro possibili unioni (pluriintervalli). Ad un qualunque elemento di Σ è associata una lunghezza e questa si può assumere come valore di \mathbb{P} . Questo schema astratto di teoria delle probabilità può adattarsi a varie situazioni pratiche, ad es. lasciando cadere un ago su un foglio a righe, qual'è la probabilità che l'ago intersechi una riga? Questo è noto come il “problema di Buffon”¹⁴ che illustreremo attraverso un esperimento Montecarlo nel prossimo esempio.

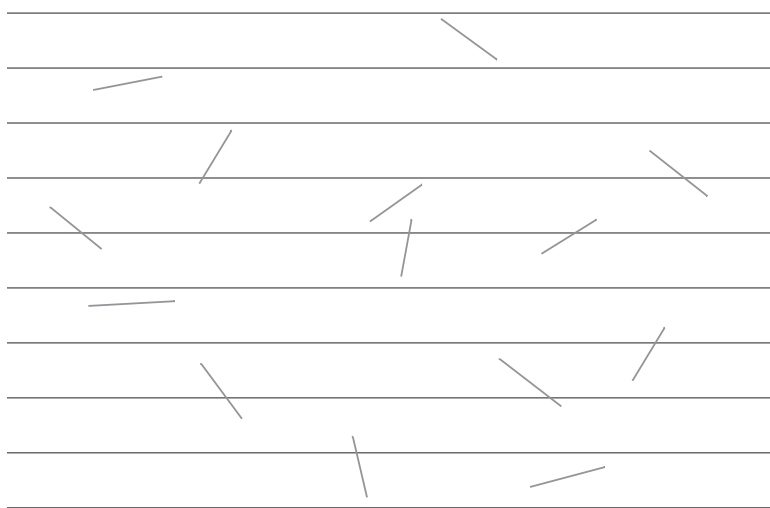


FIGURA 2.2. L'esperimento di Buffon

ESEMPIO 4 (Il problema di Buffon [Roz69]). Consideriamo un foglio a righe, idealmente di dimensioni grandi a piacere, la spaziatura delle righe sia L ; immaginiamo di lasciare cadere del tutto a caso degli aghi lunghi $\ell < L$ sul foglio: ci si chiede quale porzione degli aghi in media intersecherà una delle righe. Il problema è di natura geometrica. Orientiamo il foglio individuando una direzione positiva per l'asse y come si è abituati dalla geometria analitica, mentre le righe sono parallele all'asse x . Supponiamo per fissare le idee che l'estremità dell'ago con ordinata inferiore y cada a distanza δ ($0 \leq \delta < L$) dalla riga più vicina con ordinata $< y$ e che l'ago formi un angolo α con la direzione delle righe. In queste ipotesi, $0 < \alpha < \pi$. Si ha intersezione se $\ell \sin \alpha + \delta > L$. Ora sia δ che α assumono valori a caso, in particolare δ è distribuito

¹⁴A scanso di equivoci, il Buffon in questione si chiamava George-Louis Leclerc, Conte di Buffon, ed è vissuto nel XVIII secolo (vedi wikipedia "Buffon needle").

uniformemente ¹⁵ tra 0 e L, mentre α è distribuita uniformemente - e indipendentemente ¹⁶ da δ - sull'intervallo $(0, \pi)$. Ne concludiamo allora che la probabilità che l'ago intersechi una riga sarà data da

$$P = \mathbb{P}(\ell \sin \alpha + \delta > L) = \int_0^L \frac{d\delta}{L} \int_0^\pi \frac{d\alpha}{\pi} \theta(\ell \sin \alpha + \delta - L)$$

dove $\theta(x)$ vale 1 se l'argomento è positivo, zero altrimenti. Perciò si avrà

$$P = \int_0^\pi \frac{d\alpha}{\pi} \int_{L-\ell \sin \alpha}^L \frac{d\delta}{L} = \frac{\ell}{L\pi} \int_0^\pi \sin \alpha d\alpha$$

Ora, se sappiamo calcolare l'integrale (il risultato è ovviamente 2) possiamo concludere che la probabilità in questione vale

$$P = \frac{2\ell}{L\pi}$$

Oppure, se non conosciamo abbastanza l'analisi, potremmo provare a stimare l'integrale eseguendo una serie di esperimenti numerici ottenendo alla fine una valutazione approssimata dell'integrale (vedi il programma matlab `BuffonExp.m` riprodotto in fondo alle lezioni oppure l'altro codice `BuffonPlot.m` su dropbox). Si noti che tradizionalmente si diceva che l'esperimento di Buffon permetteva di *calcolare π attraverso una simulazione Montecarlo*. Una visione più moderna è quella che un programma di simulazione si può utilizzare per calcolare il valore di un integrale complesso - non è questo il caso, ma in pratica è proprio questa la logica di molte applicazioni Montecarlo dove gli integrali da calcolare sono su un numero di variabili talmente elevato da rendere impensabile un calcolo analitico (si vedano le lezioni di A. Sokal [Sok97]). In realtà l'ipotesi $\ell < L$ non appare molto sensata se pensiamo ai comuni fogli a righe (spaziatura 8-10 mm) e a un tipico ago da sarta (3-4 cm). Nel caso gli aghi abbiano una lunghezza superiore a L il calcolo della probabilità è più complicato ¹⁷. La formula che abbiamo considerato finora non può essere vera in generale per via che prevederebbe una probabilità maggiore di uno per $\ell > \pi L/2$. Il calcolo ci dà ¹⁸

$$P(\ell) = \begin{cases} \frac{2\ell}{\pi L} & (\ell < L) \\ 1 - \frac{2}{\pi} \left\{ \arcsin\left(\frac{L}{\ell}\right) - \left(\frac{\ell}{L} + \sqrt{\left(\frac{\ell}{L}\right)^2 - 1}\right)^{-1} \right\} & (\ell > L) \end{cases}$$

e correttamente P tende a uno per $\ell \rightarrow \infty$ (vedi Fig. 2.3). Notare che la funzione $P(\ell)$ è continua in $\ell = L$ ma *non è una funzione analitica*. In quel punto presenta una singolarità (punto di diramazione) ma la continuazione analitica della funzione lineare che vale per $\ell < L$ sarebbe la stessa funzione lineare in tutto il piano complesso. In realtà quello che succede è che la funzione per $\ell > L$ continuata analiticamente a $\ell < L$ presenta una parte immaginaria non nulla e la sua parte reale coincide con il valore corretto $2\ell/(\pi L)$.

¹⁵Questo fatto rappresenta in realtà un'ipotesi di lavoro. Dovremmo assicurarci che il metodo che utilizziamo per far cadere gli aghi non presenti qualche *bias*. Ad es. se gli aghi sono magnetizzati tendono ad allinearsi con il campo magnetico terrestre etc.

¹⁶Si incontra per la prima volta il concetto di indipendenza tra variabili casuali - il tema è sviluppato nel capitolo 2.4.1.

¹⁷Si tratta di utilizzare la formula che si trova sulla Tabella del Gradshteyn-Ryzhik 2.813.1 [GR65] e cioè $\int dx \sin^{-1} x = x \sin^{-1} x + \sqrt{1-x^2}$

¹⁸Se il calcolo analitico risultasse arduo per il lettore, c'è sempre la possibilità di chiedere a Mathematica: `2/(Pi L) Integrate[HeavisideTheta[x + l Sin[th] - L], {x, 0, L}, {th, 0, Pi/2}, Assumptions -> {l > L, L > 0}]`.

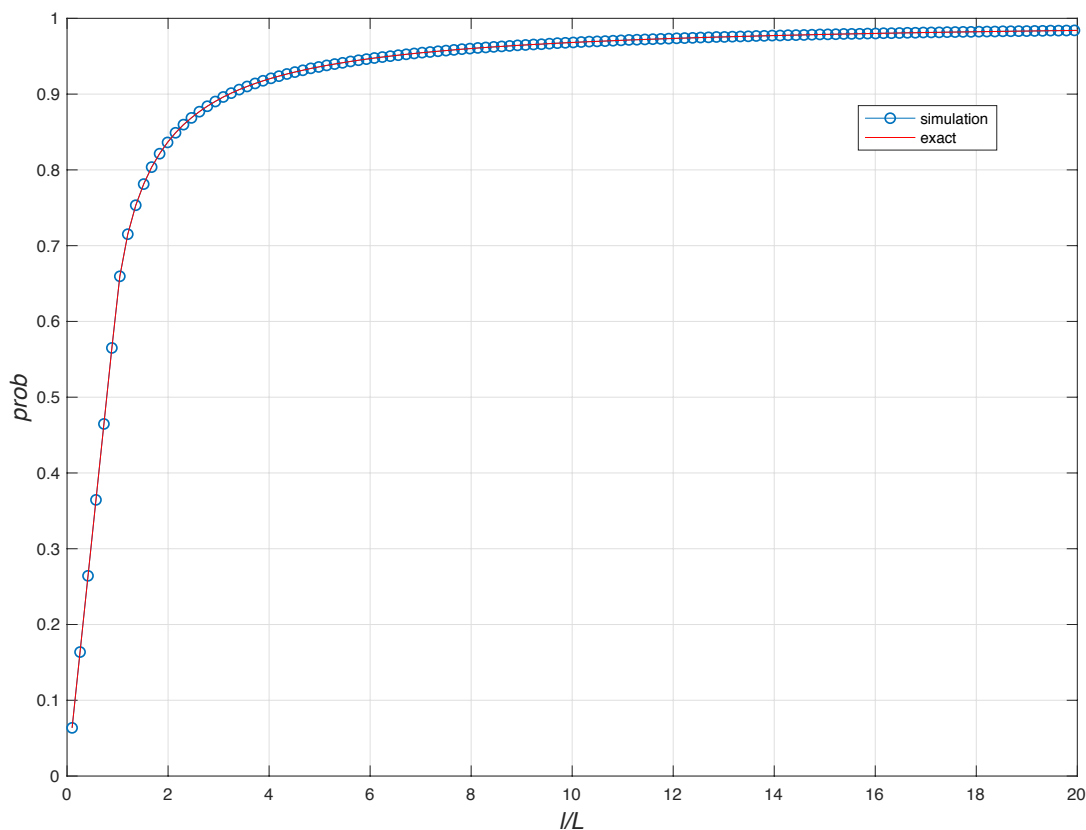


FIGURA 2.3. Esperimento di Buffon: la probabilità di intersecare una delle righe ha un andamento lineare per $\ell < L$ e va al limite 1 per $\ell \gg L$.

Nella figura 2.4 si riporta il risultato di cinquanta esperimenti di Buffon eseguiti per $\ell < L$; per ogni esperimento è riportato il valore medio e la deviazione standard. La riga blu indica la media complessiva insieme con l'errore della media ($\text{std}/\sqrt{50}$). Il risultato è ovviamente compatibile con il valore esatto di $\int_0^\pi \sin x = 2$. Va detto che per integrali in una variabile, o più in generale integrali fino a tre variabili, esistono algoritmi numerici “deterministici” del tutto affidabili; per un numero di variabili compreso tra quattro e dieci esistono algoritmi “quasi-Montecarlo” molto efficienti che sono realizzati in librerie quali la *gsl.lib* (gnu-scientific-library) o in *Mathematica*. Per dimensioni superiori del dominio di integrazione non c'è che affidarsi a un metodo Montecarlo.

Nota. Esaminando la Fig.2.5 si può constatare come variano le fluttuazioni al variare del numero di esemplari: il numero di *sweeps* viene quadruplicato passando dal primo al secondo esperimento e ugualmente si aumentano gli sweeps nei successivi. Si vede chiaramente come le fluttuazioni appaiano invariate ma la scala viene via via dimezzata e l'intervallo mostrato si restringe da 0.08 a 0.01. Questa è una caratteristica generale del Montecarlo, l'errore statistico scala come $1/\sqrt{N}$ dove N è il numero di configurazioni (misure) statisticamente indipendenti. [Sok97].

2.3. Esercizi elementari

2.3.1. Un problema con i calzini.

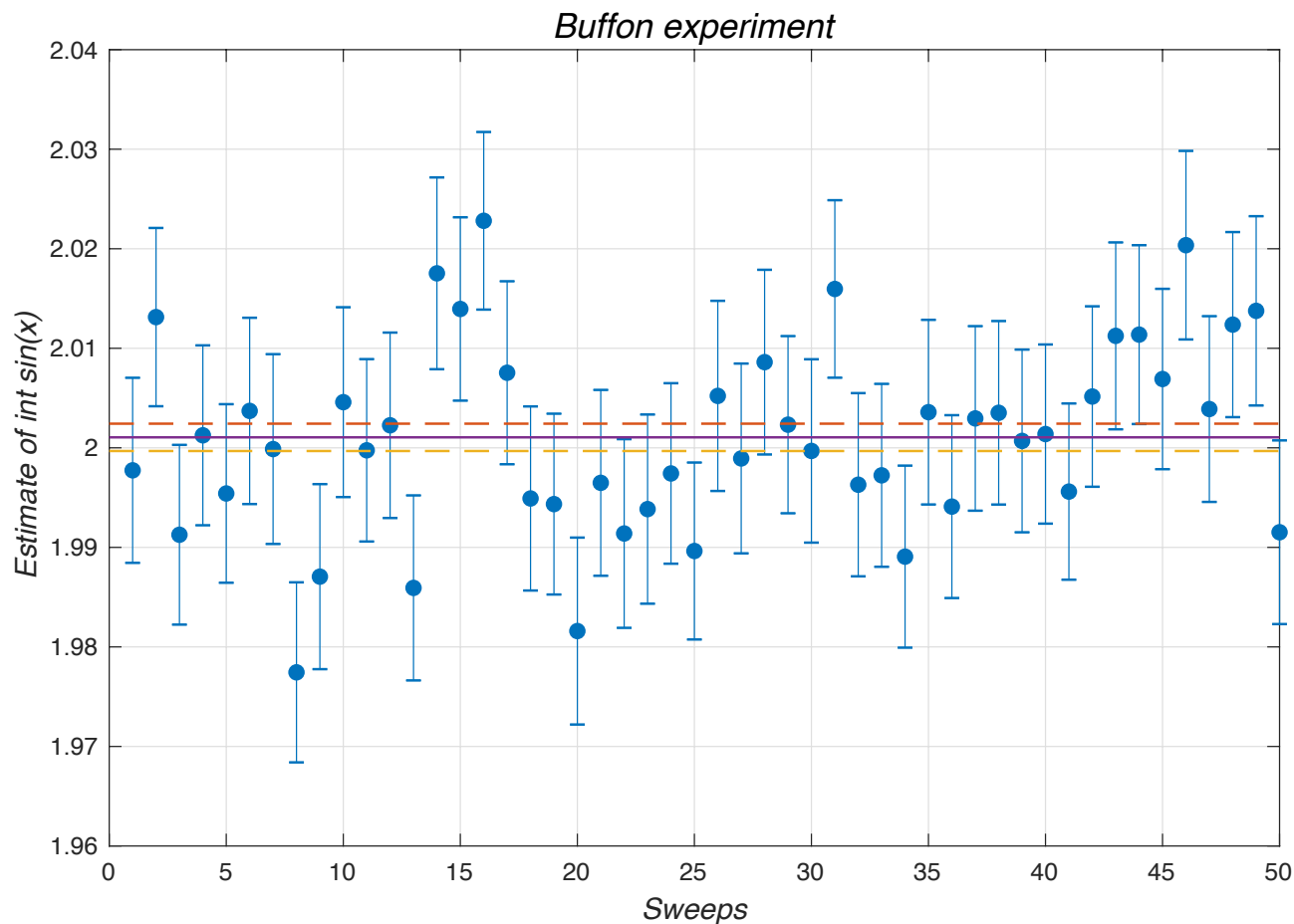


FIGURA 2.4. Esperimento di Buffon: 50 esperimenti indipendenti con 100 aghi lanciati 1000 volte. Media 2.0011 ± 0.0014 compatibile con il valore esatto.

ESERCIZIO 4 (*Il cassetto delle calze*, da Mosteller [Mos65]). Un cassetto contiene delle calze rosse e delle calze nere in numero imprecisato. Se si prendono al buio due calze a caso ci si trova con due calze rosse nella metà dei casi, avendo cura di rimpiazzare immediatamente le due calze nel cassetto in modo che la situazione sia sempre la stessa. Si chiede: 1) qual'è il numero minimo di calze presenti nel cassetto? 2) Stessa domanda nel caso in cui le calze nere siano in numero pari.

SOLUZIONE. Per impostare la soluzione indichiamo con n_r il numero di calze rosse e con n_b il numero di quelle nere. La probabilità che la prima calza scelta a caso sia rossa è evidentemente $n_r/(n_r + n_b)$. La probabilità di pescare adesso una seconda calza rossa sarà data dalla stessa formula con n_r diminuito di uno. In totale la probabilità richiesta è data da

$$P(n_r, n_b) = \frac{n_r}{n_r + n_b} \frac{n_r - 1}{n_r + n_b - 1}$$

e si tratta di determinare il minimo valore di $N = n_r + n_b$ tale che P valga $1/2$. Considerando valori piccoli del numero di calze si trova rapidamente che $n_r = 3, n_b = 1$ implica $P = \frac{3}{4} \frac{2}{3} = 1/2$, ma la ricerca di valori più alti in modo da ammettere anche valori pari per n_b è tutt'altro che semplice. Si tratta di un'equazione della classe "diofantina" il che spesso costituisce un

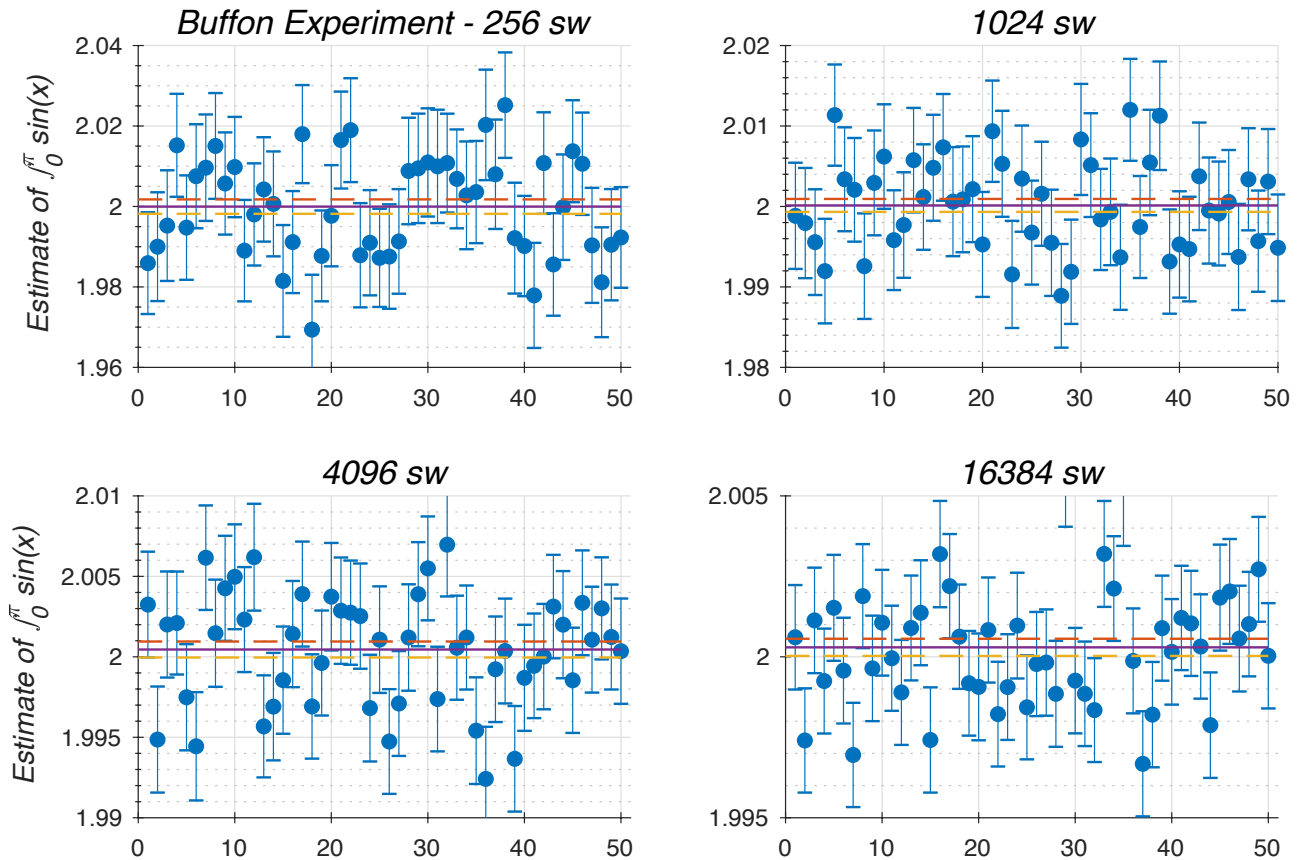


FIGURA 2.5. Esperimento di Buffon ripetuto quadruplicando successivamente il numero di campioni: prova visiva della dipendenza $1/\sqrt{N}$ per le fluttuazioni

problema spinoso. Per tagliare il nodo gordiano conviene affidarsi al computer e formulare il problema così:

```
function [i1,i2]=mosteller1(nmax)
n1 = ones(nmax,1)*(1:nmax);
n2 = n1';
N=n1+n2;
P = (n1./N).*(n1-1)./(N-1);
[i1,i2]=find(P==1/2);
```

Partendo con un numero di calze limitato a dieci per ogni colore si trova solo la soluzione già considerata. Bisogna arrivare a $n_r = 15$ per trovare la soluzione $n_r = 15, n_b = 6$. Per avere una soluzione con numeri più alti bisogna arrivare a migliaia di calze, la prima che si trova essendo $n_r = 2871, n_b = 1189$ e non esistono altre soluzioni almeno per $n_r < 20000$! Esiste una soluzione del problema ottenuta attraverso un'analisi esatta come si può verificare sul libro già citato, ma questo potrebbe costare intere notti di fatica... C'è da dire anche che la richiesta $P = 1/2$ sarebbe più ragionevolmente sostituita da $|P - 1/2| < \text{tol}$, cioè trattandosi di esperimenti di cui si tiene una statistica è più realistico richiedere che la probabilità risulti $1/2$ con una certa tolleranza, ammettendo una fluttuazione uguale a "tol". Se facciamo così troviamo che per $n_r = 10, N = 14$ la probabilità risulta 0.495 e per $N = 100$ ci sono ben ben 14 soluzioni con , $\text{tol}=0.001$. Dunque

la richiesta di avere la probabilità esattamente uguale a $1/2$ non sembra essere così ragionevole!



2.3.2. Un problema di matrici casuali.

ESERCIZIO 5. Utilizzare `matlab` per costruire un insieme di matrici in cui gli elementi di matrice sono distribuiti uniformemente a caso tra 0 e un valore massimo N . Verificare attraverso un numero adeguato di esperimenti numerici che gli elementi si distribuiscono intorno al valor medio $N/2$ con deviazione standard $\sqrt{\frac{N(N+2)}{12}}$. Costruire un algoritmo per generare matrici simmetriche a elementi interi uniformemente distribuiti in $0 : N$. Di queste matrici si potrà poi studiare lo spettro di autovalori, una problematica nata negli anni '50 ad opera di Eugene Wigner in un contesto di teoria quantistica dei nuclei atomici.

SOLUZIONE. La funzione `randi` ritorna numeri naturali distribuiti a caso in modo uniforme; per comodità conviene definire al prompt di `matlab` una funzione “anonima”¹⁹

```
>> F = @(n,r,c) randi(n+1,r,c) - 1;
```

che ritorna una matrice a “ r ” righe e “ c ” colonne i cui elementi sono distribuiti uniformemente tra 0 e n . Per ottenere matrici simmetriche conviene invece procedere come segue: si ritaglia la diagonale \mathbf{d} e la parte della matrice nel triangolo sopra la diagonale \mathbf{u} e poi si ricostruisce $\mathbf{A} = \mathbf{d} + \mathbf{u} + \mathbf{u}'$. Il codice è dato da `ISRM.m` che fa uso della routine `triu.m` di `matlab`.

La media e la varianza si calcolano molto semplicemente:

$$\begin{aligned}\mathbb{E}[\mathbf{A}] &= \frac{1}{n+1} \sum_{j=0}^n j = \frac{1}{2}n \\ \mathbb{E}[\mathbf{A}^2] &= \frac{1}{n+1} \sum_{j=0}^n j^2 = \frac{1}{6}n(2n+1) \\ \mathbb{D}[\mathbf{A}] \equiv \text{Var}[\mathbf{A}] &= n(n+2)/12\end{aligned}$$

Invece di generare un grande numero di matrici $N \times N$ con N piccolo conviene costruire una singola matrice molto grande: ad es. troviamo

$$\mathbf{A} = \text{F}(9, 100000, 1); \text{mean}(\mathbf{A}) = 4.4970; 4 * \text{std}(\mathbf{A})^2 = 33.0844;$$

Di quanto possiamo attenderci che la media in un singolo esperimento differisca dal valore teorico $n/2$? La deviazione standard dà una stima di quanto possiamo attenderci che un singolo numero estratto differisca dalla media teorica; ma la media su un grande numero di estrazioni tende a differire dal valore teorico per la deviazione standard **divisa per** \sqrt{N} . Comunque gli esperimenti numerici non si leggono sul libro ma si conducono in proprio! Si veda anche l'Es.B.8.4 dove si studia lo spettro delle matrici random a elementi interi. ◀

2.3.3. Problemi di geometria e di aritmetica aleatorie.

ESERCIZIO 6. Due punti vengono scelti a caso nell'intervallo $(0, L)$. Qual'è la probabilità che la loro distanza non superi ℓ ?

ESERCIZIO 7. Si sceglie un numero a caso compreso tra 1 e 1000. Qual'è la probabilità che il suo quadrato abbia la forma $10 \times a + 1$, cioè l'ultima cifra sia 1? Quale la probabilità che la quarta potenza abbia la stessa proprietà?

¹⁹più immediato che creare un file contenente la definizione della funzione!

SOLUZIONE. Il numero n sia dato da $10x+b$, $0 \leq b \leq 9$. Il suo quadrato è allora b^2+10y per qualche intero y , ossia $n^2 \equiv b^2 \pmod{10}$. e perciò per rispondere alla domanda è sufficiente considerare numeri naturali minori di 10. Il quadrato avrà l'ultima cifra 1 nel caso $b = 1$ e $b = 9$, dunque la probabilità è $2/10$. Per la seconda domanda, dovrà essere $b^2 \equiv 1 \pmod{10}$ oppure $b^2 \equiv 9 \pmod{10}$ e ciò vale per $b = 1, 3, 7, 9$, dunque la probabilità vale $4/10$. Questo problema può essere dunque risolto elencando tutte le possibilità e contando i casi favorevoli. Un'alternativa, utilizzando il calcolo numerico, è data dal seguente mini-programma matlab:

```
>> n = 0:9;
>> P1 = sum(mod(n.^2,10)==1)/10;
>> P2 = sum(mod(n.^4,10)==1)/10;
```

o in generale

```
>> P = @(n,t,m) sum(mod(n.^t, 10) == m)/10;
```

che permette anche di rispondere alla domanda “se prendo la potenza n^{10} qual'è la probabilità di trovare l'ultima cifra uguale a n (risposta: $4/10$, non banale da calcolare a mente ...)”. ◀

ESERCIZIO 8. Si sceglie a caso un numero intero compreso tra 1 e 1000. Qual'è la probabilità che il cubo del numero abbia le ultime due cifre uguali a 1? (cioè $n^3 = **11$).

SOLUZIONE. Il problema si imposta facilmente in termini aritmetici. Se $n = c_0 + 10c_1 + 100c_2$ con c_i compreso tra 0 e 9, si avrà $n^3 = c_0^3 + 30c_0^2c_1 \pmod{100}$ (in quanto tutti gli altri termini comportano un fattore 100 e quindi non influiscono sulle ultime due cifre; il 1000 non è compreso, ma è comunque da scartare). Ora il termine $30c_0^2c_1$ influenza solo la penultima cifra del risultato e quindi si dovrà avere $c_0^3 \equiv 1 \pmod{10}$ che implica $c_0 = 1$. Di conseguenza $3c_1 \equiv 1 \pmod{10}$ da cui segue $c_1 = 7$. Di tutte le scelte possibili per c_0, c_1 e c_2 la richiesta è soddisfatta da tutte le triplette $(1, 7, c_2)$ con c_2 arbitrario. Dunque la probabilità è uguale a $1/100$, e cioè i dieci valori possibili di c_2 rispetto al numero totale di 1000. Un semplice comando matlab trae dall'impiccio nel caso non si sia particolarmente forti in aritmetica (e neanche il presente autore lo è...):

```
>> n = 1:1000;
>> N3 = n.^3;
>> sum(mod(N3,100) == 11)/1000
```

ESERCIZIO 9 (Tre numeri a caso). Estrahendo a caso tre numeri interi qual'è la probabilità che questi non abbiano divisori in comune?

SOLUZIONE. Il problema sembra a prima vista mal formulato. Infatti non ha senso estrarre un intero da \mathbb{Z}^+ dato che una distribuzione uniforme su un set infinito non è normalizzabile. Possiamo intendere che le terne di interi vanno estratte in un intervallo finito $I = [1, N]$ con N grande. Allora si potrà ragionare così: i numeri pari figurano per la metà dunque la probabilità che nessuno dei tre sia pari²⁰ è $1 - 1/2^3$. I multipli di tre costituiscono un terzo dell'insieme I , e dunque la probabilità che nessuno dei tre sia multiplo di 3 sarà $1 - 1/3^3$. Ragionando così per

²⁰Il valore è corretto solo per il limite di N grande, infatti il numero di interi pari è $N/2$ se N è pari e $(N-1)/2$ se N è dispari!

tutti i numeri primi (minori di N) si trova

$$\text{Prob}(\text{'no-fattori-comuni'}) = \prod_{p \text{ primo}, p < N} (1 - p^{-3})$$

Una formula di Eulero²¹ lega questo prodotto nel limite di N grande alla funzione $\zeta(s)$ di Riemann (meglio funzione di Eulero-Riemann):

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s} = \prod_{p \text{ primo}} \frac{1}{1 - p^{-s}}$$

Per il caso di tre interi si ha dunque che la probabilità richiesta, prendendo N molto grande, è data da $1/\zeta(3) \approx 0.831907\dots$. Per il caso di quattro interi scelti a caso la probabilità che non abbiano fattori comuni è $P = 1/\zeta(4) = 90/\pi^4 \approx 0.923938\dots$

Nel folder `~/Dropbox/MPF2023/Programmi/Matlab/Cap.2/Coprimes` trovate il codice `coprimebatch.m` in `matlab` dove il risultato è messo alla prova attraverso una simulazione e il risultato è analizzato utilizzando `boxplot.m`. L'estrazione a caso di k interi tra 1 e N si ottiene con `n = randi(N,k,1)`; utilizzando `gcd()` si trova rapidamente il massimo comun divisore dei k interi e se questo è uno significa che i numeri sono primi tra loro. Si tratta di accumulare un gran numero di terne a caso, prendere la media e valutare la fluttuazione statistica. Prima di correre a vedere il codice provate voi a svilupparlo - bastano pochi minuti...

La simulazione aggiunge qualcosa all'analisi pura: lavorando con N piccolo, diciamo $N = 10$, si riscontrano deviazioni dalla formula analitica ben superiori alle fluttuazioni statistiche. Cioè si riscontra la presenza di un errore sistematico che qui è dovuto al fatto che la formula analitica vale nel limite $N \rightarrow \infty$; il produttorio nella formula di Eulero si deve restringere a numeri primi inferiori a N e questo costituisce una deviazione sistematica dalla formula $1/\zeta(k)$ che però risulta piccola per $N \gtrsim 100$.

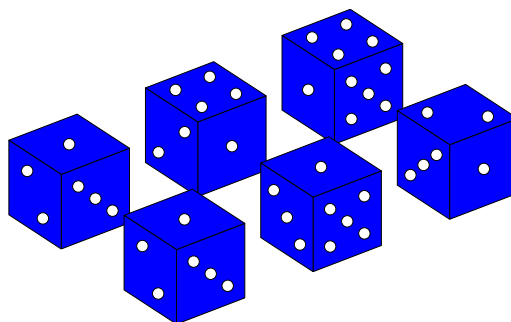
Il riferimento in letteratura è [Nym72], dove si mostra che il risultato corretto è $P(k) = \zeta(k)^{-1} + \mathcal{O}(1/N)$ se i numeri vengono estratti nell'intervallo $[1, N]$. Si riscontra infatti che un fit polinomiale dei risultati della simulazione assorbe molto bene il termine correttivo e può riprodurre il valore asintotico con estrema precisione (i.e. deviazione di 10^{-5} per $N < 300$ e $k = 5$). Nel codice `runcoprime.m` si effettua la simulazione per N primo $< N_{\max}$ e si effettua il fit in $1/N$.

2.3.4. Giochi al Casinó.

ESERCIZIO 10 (De Mère 1). Si tirano tre dadi. Si determini la probabilità che la somma dia 11 oppure 12. Nel '600, un certo De Mère, che passava molto tempo al tavolo da gioco, si era stupito del fatto che $P(11) > P(12)$, almeno stando alle sue osservazioni, ma Pascal diede la dimostrazione che il fatto era del tutto prevedibile sulla base di un argomento matematico²².

²¹Si può consultare il trattato di Whittaker e Watson [WW69] sec.13.3. Questo testo di Analisi Matematica rappresenta un riferimento importante anche se risale al 1902 (prima ediz.).

²²Questo fatto risulta da una corrispondenza tra Pascal e Fermat, ca. 1654, cui viene fatta risalire la nascita della teoria matematica della probabilità. Per notizie storiche di questo tipo si può consultare il bel libro di D. Gillies [Gil00] dove si fa cenno alla soluzione data molti anni prima da Galileo allo stesso problema! Molti dettagli su Pascal e De Mère si possono trovare inoltre su [Dav62].



SOLUZIONE. Le disposizioni dei tre dadi che portano a una somma uguale a 11 sono le 6 seguenti ²³: $6 + 4 + 1$, $6 + 3 + 2$, $5 + 5 + 1$, $5 + 4 + 2$, $5 + 3 + 3$, $4 + 4 + 3$, ma tenendo conto delle permutazioni possibili ($6 + 4 + 1$, $6 + 1 + 4$, $4 + 1 + 6$, $4 + 6 + 1$, $1 + 6 + 4$, $1 + 4 + 6$ etc.) si tratta in realtà **27** su 216 possibilità, mentre per il 12 si ha $6 + 5 + 1$, $6 + 4 + 2$, $6 + 3 + 3$, $5 + 5 + 2$, $5 + 4 + 3$, $4 + 4 + 4$, sempre 6 partizioni, ma solo **25** combinazioni. De Mère si era ingannato non tenendo conto delle permutazioni possibili. Si veda la trattazione generale all'Eserc.23 a pag. 36. ◀

ESERCIZIO 11 (De Mère 2). Un'altra osservazione attribuita a Monsieur *De Mère* riguarda il fatto che apparentemente era più facile vincere puntando sulla uscita di un sei in quattro lanci di un dado a sei facce piuttosto che puntando sull'uscita di due sei in ventiquattro lanci di due dadi (Si veda [BV12]). Potete giocare con i dadi, per verificarlo, ma ci vuole molto tempo, oppure simulare la cosa al computer, ma la domanda interessante è: quante giocate sono state necessarie a M. De Mère per scoprire questo fatto?

SOLUZIONE. Naturalmente si calcola facilmente la probabilità nei due casi: in quattro lanci la probabilità che non esca alcun sei è ovviamente $(5/6)^4$ e dunque la probabilità di successo è $1 - (5/6)^4 \approx .52$. Nell'altro caso la probabilità che non escano due sei in un singolo lancio di due dadi è $35/36$, in n lanci $(35/36)^n$ e dunque per $n = 24$ si trova che la probabilità di successo è $1 - (35/36)^{24} \approx .49$.

Noi qui consideriamo una simulazione che ci permette di concludere, osservando la fig.2.6, che fino a un migliaio di giocate non c'è modo di osservare differenza sensibile tra le due forme di scommessa; ben al di sopra di mille giocate si mette invece chiaramente in evidenza il risultato anche se con diecimila si ha maggiore sicurezza... Ma De Mère era un giocatore accanito e non ci sarebbe da stupirsi se il suo numero di giocate sia stato almeno a cinque cifre. Il codice matlab è `DeMereplot.m` che si può trovare in fondo alle lezioni.

Nota. Da quest'ultimo esempio si può apprezzare come la simulazione numerica permetta di andare più a fondo nel fenomeno aleatorio, con il minimo sforzo, permettendo di stimare valori medi e fluttuazioni, queste ultime non sempre ricavabili facilmente con un calcolo analitico.



ESERCIZIO 12 (De Mère 3). Riprendendo il problema del gioco dei dadi “alla De Mère”, immaginiamo che le puntate siano organizzate in modo che ad ogni serie di lanci dei dadi il giocatore versi una posta di P euro e ogni volta che esce un 6, o una coppia di 6 a seconda del

²³Per calcolare in modo sistematico tutte le possibilità si può ricorrere all'uso di una *funzione generatrice*. Si veda l'esercizio in fondo al capitolo a pag.37.

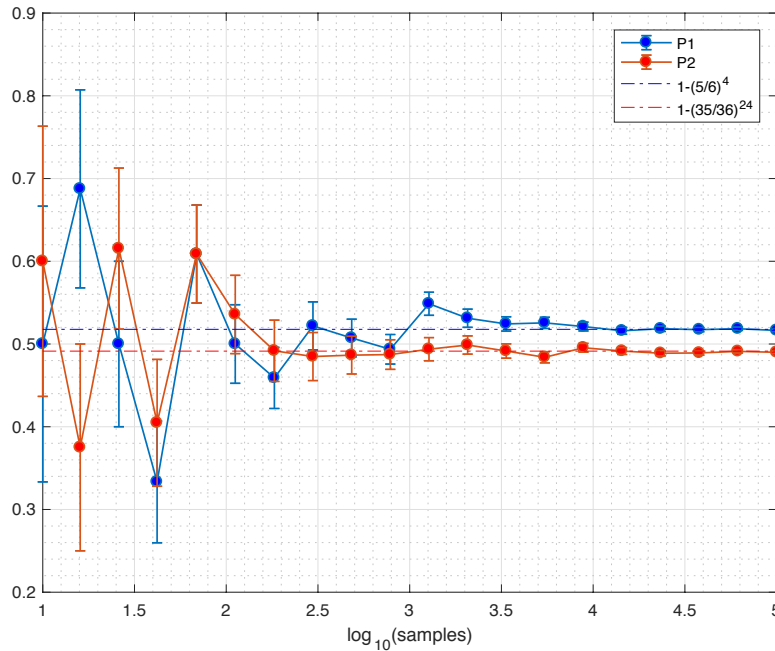


FIGURA 2.6. Simulazione delle giocate a dadi di De Mère: in blu le giocate sul 6 in quattro lanci.

tipo di gioco, sia premiato con Q euro. Qual'è il valore di Q nei due casi affinché il gioco sia equo?

SOLUZIONE. Consideriamo il primo caso: quattro tiri di un singolo dado. La probabilità che non esca il sei vale $p = (5/6)^4$. Il processo si configura come la ripetizione di quattro estrazioni dalla distribuzione di Bernoulli (vedi al cap. 2.5.1) con $q = 5/6$ e $p = 1/6$. Dunque la probabilità che esca il sei n volte su quattro è data dalla distribuzione binomiale (vedi al cap. 41) $p(n) = \binom{4}{n} q^n p^{4-n}$. Il valore medio del numero di successi (np) vale $4 \times \frac{1}{6} = \frac{2}{3}$ con una deviazione standard $npq = 4 \times \frac{1}{6} \times \frac{5}{6} = \frac{5}{9} \approx 0.556$. Nel caso della ricerca di una coppia di 6 su ventiquattro tiri di due dadi si avrà $q = 35/36$, $p = 1/36$ e dunque il valor medio del numero di uscite di una coppia di sei è dato da $24 \times \frac{1}{36} = \frac{2}{3}$, identico al primo risultato, con deviazione standard $24 \times \frac{1}{36} \times \frac{35}{36} = \frac{35}{54} \approx 0.648$. In entrambi i casi dunque al giocatore conviene giocare solo se la vincita Q è pari ad almeno $\frac{3}{2}P$. Nel programma su dropbox “DeMereBet.m” si simula in gioco nei due casi e si valuta il guadagno, o la perdita, nel caso $Q = \frac{3}{2}P$. ◀

ESERCIZIO 13. Un cubo di legno ha le facce colorate. Il cubo viene sezionato in dieci parti parallelamente alla faccia $x - y$, poi ancora i dieci parti secondo $x - z$ e infine secondo $y - z$. Qual'è la probabilità che scegliendo a caso uno dei mille cubetti questo abbia due facce colorate?

ESERCIZIO 14 (Giocatore ostinato, da [Mos65]). Un giocatore resta ore e ore alla roulette al Casinó e punta esclusivamente su un singolo numero. Una sessione è costituita da 36 puntate da 1 Euro. Qual'è il risultato atteso sul portafoglio del giocatore dopo ogni sessione? Si tenga presente che per l'uscita del numero giusto il banco paga 35 volte la posta e che ci sono 37 numeri sulla ruota (0:36). Una variante: nei Casino americani ci sono 38 numeri, oltre allo 0 esiste il doppio 0, anche questo a favore del banco ma la quota in caso di vincita è ancora 35 volte la posta. Qual'è il guadagno atteso in questo caso?

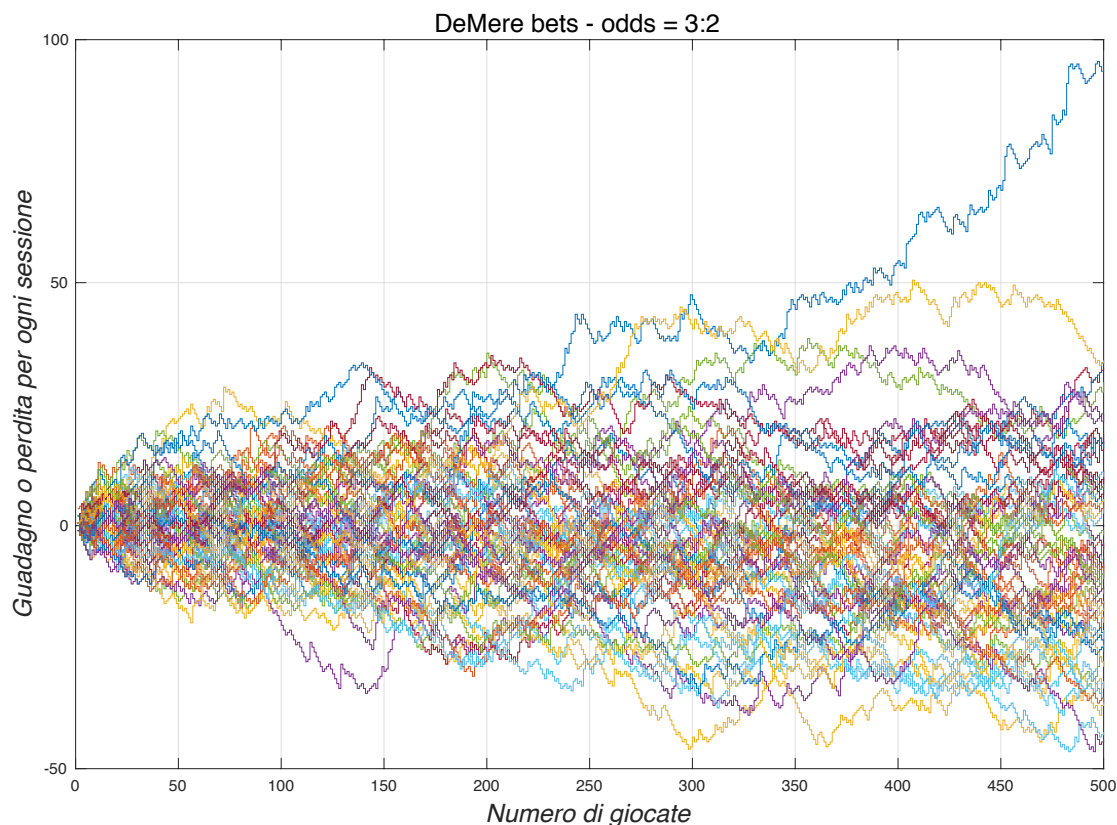


FIGURA 2.7. Una serie di giocate ai dadi; il gioco appare equo, come previsto, infatti le storie delle giocate sono concentrate intorno a zero.

SOLUZIONE. Il gioco rientra nel caso di ripetizione di estrazioni di Bernoulli, dunque la sessione di puntate è regolata dalla distribuzione binomiale: il numero di vincite è dunque in media np dove $n = 36$ e $p = 1/37$ ossia $\langle V \rangle = np = 36/37 \approx 0.97$. Il guadagno $G(k)$ per k successi è dato da $35k - (36 - k) = 36(k - 1)$ e in media $\langle G \rangle = 36(\langle k \rangle - 1) = 36(36/37 - 1) = -36/37 \approx -0.97$. La fluttuazione è calcolata con la formula $\Delta k = np(1 - p) = 36 \times \frac{1}{37} \times \frac{36}{37} \approx 0.92$. Per la roulette americana a 38 numeri il calcolo, a parità degli altri dati, dà $np = 36/38$ e $\langle G \rangle = 36(\frac{36}{38} - 1) \approx -1.89$. Le roulette europee sono più “oneste”. Lo sarebbero completamente se in caso di vittoria l’incasso fosse 36 invece di 35. Ma il Casino si riserva una quota che permette di vincere sempre.

Tra i programmi su dropbox (Cap.2) si troverà `Mosteller7.m` che studia il problema attraverso una simulazione e che permette rapidamente di variare i parametri a piacere. ◀

ESERCIZIO 15. Come nell’esercizio precedente, con la variante che il giocatore punta sistematicamente su numeri pair et rouge (pari e rosso). In caso di vincita il giocatore incassa tre volte la posta (la formula generale è $P = 36/n - 1$ dove n rappresenta il numero di casi favorevoli, in questo caso 9, e il “-1” è la quota garantita al banco).

SOLUZIONE. $p = 9/37 \approx 0.24$. Segue $\langle V \rangle = 36 \times 9/37 \approx 8.75$ e il bilancio sarà $G = 3 \times \langle V \rangle - (36 - \langle V \rangle) \approx -1$, cioè il giocatore perde in media la sua puntata per ogni sessione. ◀

2.3.5. Problemi di code.

ESERCIZIO 16 (*Il problema del botteghino*). Un centinaio di spettatori si presentano in fila ordinata alla biglietteria del cinema Astra, in possesso per la metà di biglietti da 5€ e l'altra metà di biglietti da 10€. La cassa del botteghino è inizialmente vuota e l'ingresso costa 5€. Qual'è la probabilità che nessuno nella coda debba aspettare per avere il resto?

SOLUZIONE. *Il problema rientra nella categoria generale delle “camminate aleatorie” (o random walks). In attesa di sviluppare qualche metodo di analisi, si può tentare una soluzione empirica basata sulla simulazione della coda. Il programma `rwstat.m` si trova su*

dropbox/MPF2023/Programmi/Matlab/Cap.2/Botteghino.

Il codice si può esaminare nell'appendice a pag. 146. Il codice realizza una successione di situazioni a caso per un numero qualunque $2n$ di spettatori. Esaminando l'output del programma per vari valori di n , risulta naturale congetturare che la probabilità che la coda non si arresti, $\mathbb{P}(\text{no-stop})$, sia approssimativamente $1/n$ (siveda il diagramma di Fig. 2.8). La simulazione si realizza così: ogni spettatore con biglietto da 5 aumenta la scorta della cassa, mentre uno con biglietto da dieci la fa diminuire. Dunque una qualunque coda si può mettere in corrispondenza biunivoca con una successione di $+1$ e -1 in quantità uguali a n . Si ottengono tutte le possibili code considerando tutti i possibili ordinamenti di $2n$ elementi di cui n uguali a $+1$ e n uguali a -1 . Il numero di queste combinazioni è uguale a $\binom{2n}{n}$ e si tratta quindi di prendere il vettore degli indici $(1, 2, 3, \dots, 2n)$, estrarre un qualunque sottoinsieme di n elementi ed assegnare il valore 1 alle componenti con gli indici selezionati e -1 alle rimanenti. Non è consigliabile tentare di enumerare tutte le code possibili e contare per quante di queste la cassa rimane sempre rifornita, in quanto per $n=100$ si tratta di un numero a 59 cifre! Però c'è il modo di estrarre a caso un numero ragionevole, ben inferiore, di combinazioni e sulla base di questo campionamento effettuare una stima.

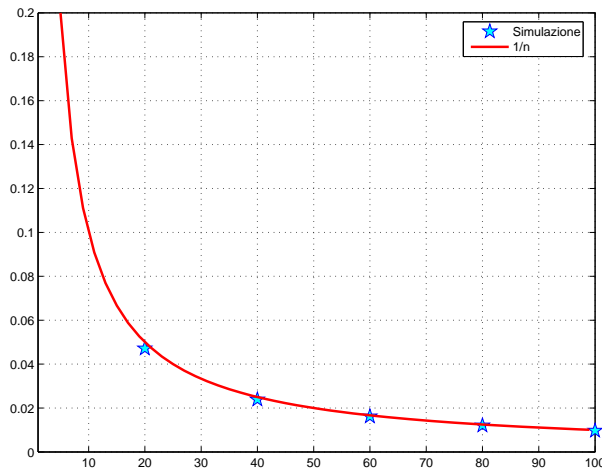


FIGURA 2.8. Simulazione della coda alla biglietteria: evidenza per la legge $\mathbb{P}(\text{no-stop}) \propto n^{-1}$

Naturalmente, dato che si tratta di una sorta di sondaggio a campione, bisognerà tenere presente che il risultato è soggetto a errore statistico che bisognerà quantificare. La parte più difficile della simulazione è infatti costituita di solito proprio dalla valutazione corretta delle fluttuazioni

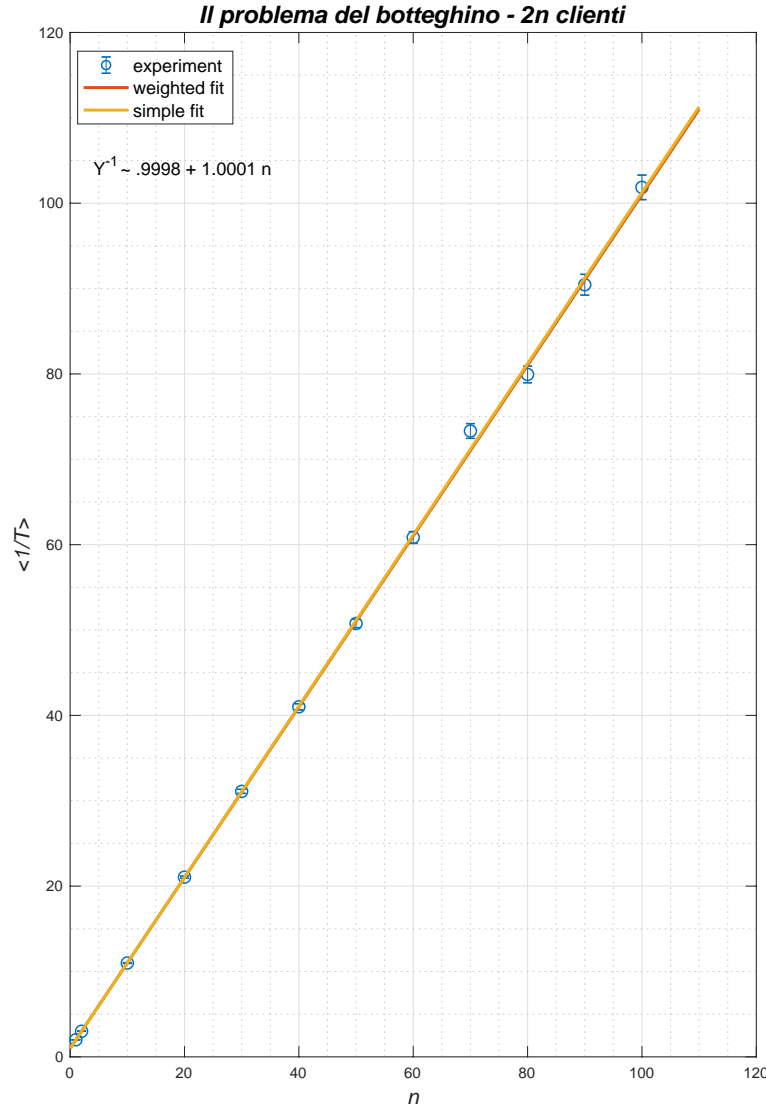


FIGURA 2.9. Best fit ottenuto con 100 esperimenti di 1000 campioni per ogni n .

statistiche. Nella fig. 2.9, una simulazione per $n = 4, 9, 14, 19, 24, \dots$ con 1000 campioni è analizzata ripetendo il campionamento per 100 volte e valutando la deviazione standard. Per rendere più evidente il risultato è conveniente visualizzare il reciproco del valor medio, che appare evidentemente lineare in n . Le barre di errore riportano l'errore della media²⁴. Risulta che, in questo esperimento, il valore più probabile della media è $1/(1.005n + 0.982)$, utilizzando il codice di best fit lineare `polyfitweighted.m`²⁵ che pesa maggiormente i dati con minore fluttuazione statistica. Il codice è riportato in dropbox `rwstat.m`. Senza la correzione che tiene conto dell'errore statistico i parametri sono chiaramente peggiori. Altri esperimenti mostrano che i coefficienti sono soggetti a fluttuazione ma sono consistenti con la formula $\mathbb{P}(\text{no-stop}) = 1/(n+1)$. Si noti che in questo caso ($n \leq 80$) l'approccio Montecarlo è efficace; infatti con 10^5

²⁴cioè si calcola la deviazione standard (`std`) dei dati e si divide per $\sqrt{N_{\text{exp}}}$, essendo N_{exp} il numero di esperimenti indipendenti. Si veda l'Es.B.8.1 in Appendice.

²⁵Il codice sviluppato da S. Rogers è disponibile sul deposito Mathworks.

estrazioni si catturano le caratteristiche del fenomeno che comporta $\binom{160}{80} \approx \mathcal{O}(10^{47})$ possibili eventi! La considerazione di code con un numero limitato di spettatori induce a confermare la formula. Infatti consideriamo i casi $n = 1, 2, 3$. È ovvio che $\mathbb{P}_{\text{no-stop}}(n = 1) = 1/2$; per $n = 2$ abbiamo un totale di 6 code possibili, ma solo le due combinazioni $++--$ e $+--+$ sono favorevoli, dunque $\mathbb{P}_{\text{no-stop}}(n = 2) = 1/3$. Infine per $n = 3$ si hanno in totale 20 possibilità e le code senza attesa sono date da $+++---$, $++-+--$, $+-+--+$, $+-+-+-$ e $+-+-+-$, quindi $\mathbb{P}_{\text{no-stop}}(3) = 1/4$. A questo punto abbiamo una congettura suffragata da un esperimento Montecarlo. Si può fare meglio: la soluzione si può determinare con il metodo della riflessione speculare (si veda [Gne79], Cap.I, Es. 5 e la successiva nota), oppure con il metodo della funzione generatrice. Il metodo rappresenta un caso particolare di qualcosa che svilupperemo più avanti, a Pag. 85. Sia $N(n)$ il numero di casi favorevoli (nessuno tra i $2n$ spettatori è costretto ad attendere). Sia poi $N_0(n)$ il numero di casi in cui la cassa rimane sempre in possesso di almeno un biglietto da 5 €, ad eccezione del momento iniziale e di quello finale, in cui per ipotesi non vi sono biglietti da 5 disponibili. Allora siano $2m$ gli spettatori che passano dalla cassa prima che questa sia priva di biglietti da 5; ovviamente $1 \leq m \leq n$ e si avrà

$$N(n) = \sum_{m=1}^n N_0(m)N(n-m)$$

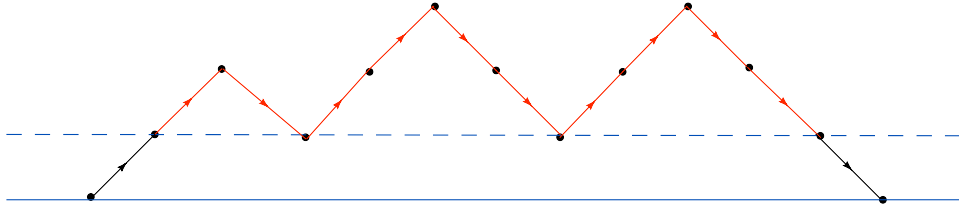


FIGURA 2.10.

Dal grafico in Fig. 2.10 si capisce anche che ad ogni evento favorevole si può associare un cammino lungo due passi in più che ritorna a zero per la prima volta dopo $2(n+1)$ passi (il problema del “primo ritorno” di un cammino aleatorio sarà studiato a Pag. 85).

Abbiamo perciò che $N_0(n+1) = N(n)$ e quindi vale una relazione di ricorrenza che permette di ricostruire facilmente $N(n)$:

$$N(n) = \sum_{m=1}^n N(m-1)N(n-m) = \sum_{m=0}^{n-1} N(m)N(n-m-1)$$

$$N(n+1) = \sum_{m=0}^n N(m)N(n-m)$$

Sia $F(x) = \sum_{n \geq 0} N(n)x^n$; moltiplicando i due termini della relazione di ricorrenza per x^{n+1} si trova

$$\sum_{m \geq 0} N(m+1)x^{m+1} = F(x) - 1 = xF(x)^2.$$

Si ottiene così $F(x) = \frac{1-\sqrt{1-4x}}{2x}$. I coefficienti dello sviluppo in serie si possono ricondurre a $\frac{1}{n+1}\binom{2n}{n}$ e sono noti come numeri di Catalan [Sta99]. Il risultato dimostra che la probabilità richiesta dal problema è esattamente $1/(n+1)$. ◀

Nota. Sul testo di Gnedenko [Gne79] si trova un altro metodo di soluzione estremamente semplice, ma difficile da scoprire. Si considerino tutti i possibili cammini $(0,0) \rightarrow (2n,0)$ su un reticolo a coordinate intere (si veda la fig.2.11). Uno spostamento $(1,1)$ corrisponde a uno spettatore con banconota da 5 € e uno spostamento $(1,-1)$ a uno spettatore con banconota da 10. I cammini favorevoli per i quali la coda non si arresta mai si situano al di sopra dell'asse delle ascisse. Invece la coda si arresta quando il cammino raggiunge la ordinata $y = -1$. Ad ognuno di questi cammini si può associare un cammino modificato ottenuto riflettendo attorno alla retta $y = -1$ tutta la parte rimanente del cammino, cioè dal primo contatto con la retta $y = -1$ fino alla fine (in rosso nel diagramma). Il cammino modificato congiunge $(0,0) \rightarrow (2n,-2)$ e i cammini così costruiti sono in corrispondenza biunivoca ai cammini $(0,0) \rightarrow (2n,0)$ che toccano l'asse $y = -1$. Possiamo allora contare i cammini sfavorevoli contando questi cammini modificati: questi sono caratterizzati da $n+1$ passi $(1,-1)$ e $n-1$ passi $(1,1)$. Il loro numero è pertanto pari al numero di combinazioni di $n+1$ oggetti in un insieme di $2n$, e cioè $\binom{2n}{n+1}$. Se ne conclude che la probabilità di arresto della coda è data dal rapporto tra il numero di cammini $(0,0) \rightarrow (2n,-2)$ e quello dei cammini $(0,0) \rightarrow (2n,0)$: $P_{\text{stop}} = \binom{2n}{n+1} / \binom{2n}{n} = n/(n+1)$. Ne segue il risultato già ottenuto in precedenza.

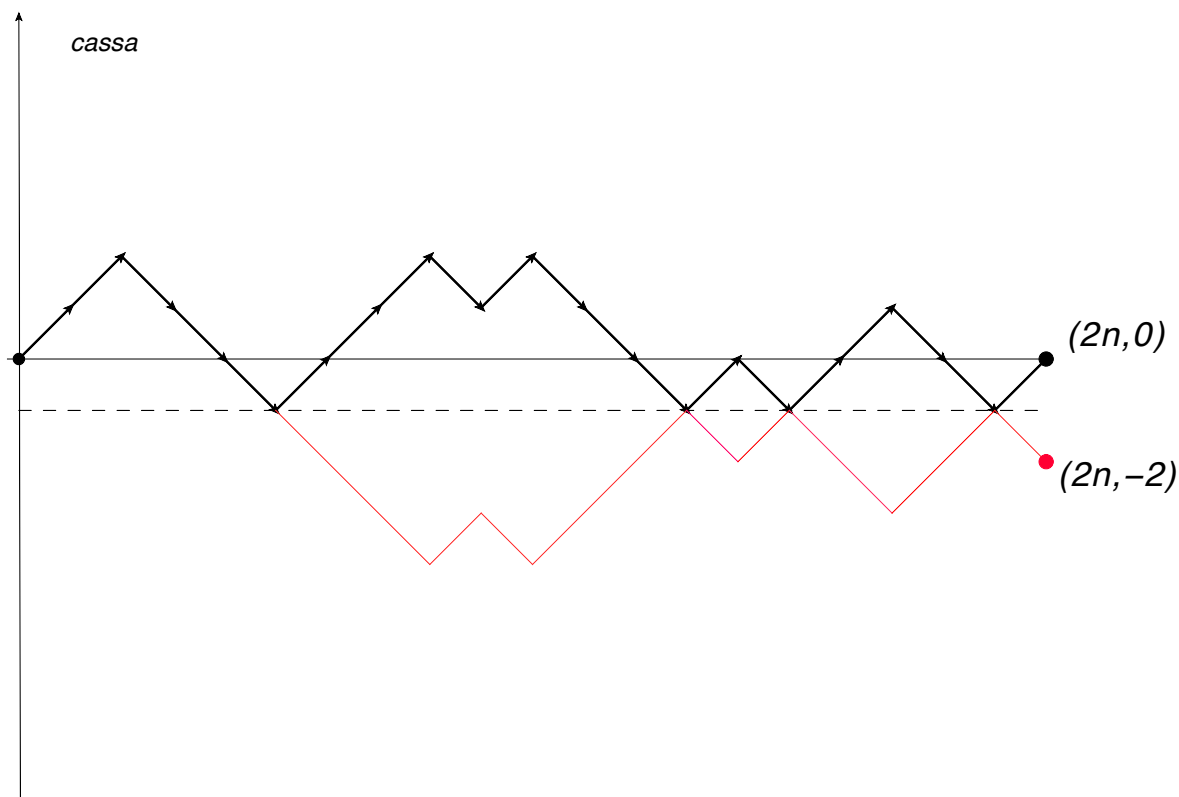


FIGURA 2.11. L'argomento di Gnedenko

ESERCIZIO 17. Una variante dell'esercizio precedente: supponiamo che nella coda ci siano $n+m$ spettatori muniti di banconote da 5 € e $n-m$ con 10 € (ovviamente m deve essere positivo, perchè con un eccesso di biglietti da 10 la coda si ferma con certezza). Come varia la probabilità che la coda non incontri interruzioni?

SOLUZIONE.

$$(2) \quad P = \frac{2m+1}{n+m+1}$$

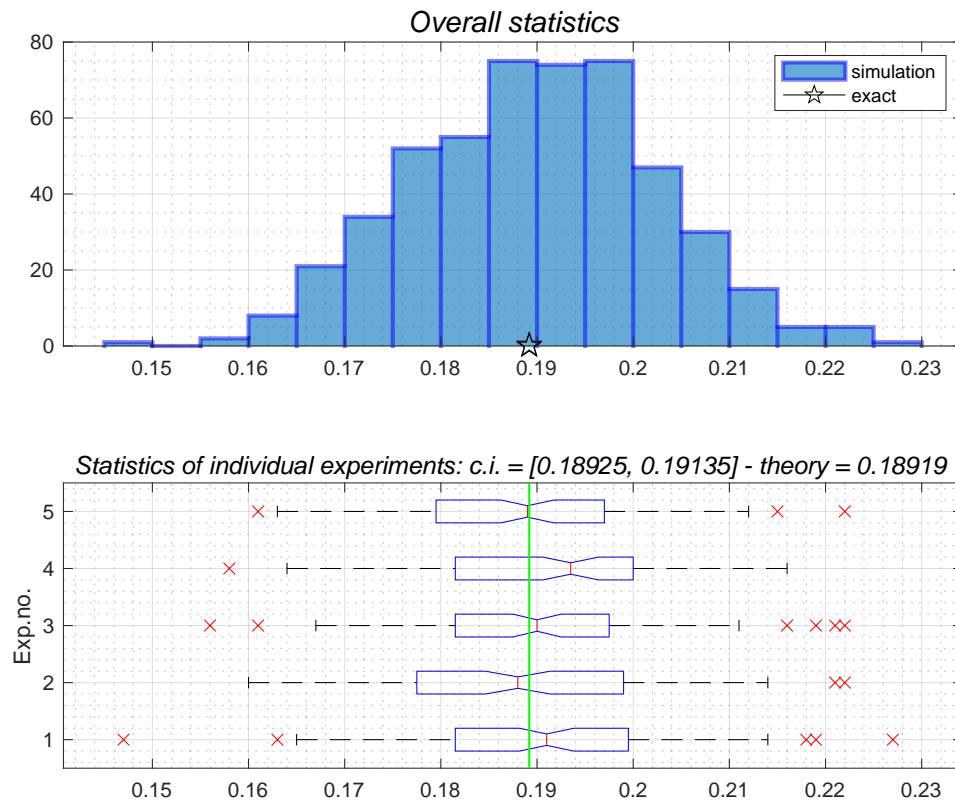


FIGURA 2.12. Risultato dell'esperimento visualizzato con `boxplot`. Nell'esperimento $n = 100$, $m = 10$, $c = 0$.

ESERCIZIO 18. Una seconda variante: in cassa ci siano inizialmente un numero c di banconote da 5 €; la domanda è la stessa dell'esercizio precedente.

SOLUZIONE. Il programma matlab `Cap.2/Botteghino/Bott.m` (vedi Appendice B, Sez. B.2) contiene la soluzione. Prima però di correre a vederlo, provate da soli a risolvere il problema applicando uno dei metodi che abbiamo imparato - partendo da numeri piccoli e cercando di congetturare la soluzione, sviluppando una variante del programma `rw.m` o tentando

una modifica alle strategie analitiche. Nel diagramma in Fig. 2.12 le informazioni di cinque esperimenti indipendenti sono convogliate con la tecnica del **boxplot**. Ogni elemento grafico contiene in modo sintetico e intuitivo una caratteristica dell'esperimento. La scatola è posizionata in modo che il segmento centrale verticale coincide con la mediana dei dati; il lato più a destra indica il valore oltre cui si trova il 25% dei dati superiori alla mediana e quello più a sinistra indica dove si collocano il 25% dei dati inferiori. La scatola in sé copre il 50% dei dati centrali e il segmento più lungo tratteggiato indica l'intervallo di 3σ intorno alla mediana. I lati della scatola sono ristretti verso il centro su una regione che è proporzionale alla **std** dei dati scalati con la radice quadrata degli esemplari statisticamente indipendenti (errore della media). Questo viene evidenziato in modo sovrabbondante riportando anche la statistica effettuata con **errorbar**. Come si vede una semplice chiamata a **boxplot** si incarica di convogliare tutti i dati statistici rilevanti, compresi i dati che vengono considerati "outliers", cioè eventi eccezionali, indicati con delle crocette, al di fuori dell'intervallo considerato "normale". In questi cinque esperimenti i dati erano i seguenti: $n = 100, m = 10, c = 0$ e si deve applicare l'Eq.(2). Ogni esperimento consiste in 100 serie di 10^3 estrazioni a caso delle code al botteghino. La suddivisione in pacchetti di simulazioni, anziché effettuare un'unica serie di 10^5 estrazioni, permette di stimare meglio le fluttuazioni statistiche. Si veda la documentazione Matlab per **boxplot**. ◀

ESERCIZIO 19 (Ballottaggio). I due candidati Alice e Bob per l'elezione a Sindaco hanno ricevuto al ballottaggio $N_A = p$ e $N_B = q$ voti con $p > q$. Qual'è la probabilità che nel corso dello spoglio delle schede Alice figuri sempre avanti rispetto a Bob?

SOLUZIONE. Il problema è molto simile a quello della coda al botteghino. Qui però si richiede che i voti favorevoli ad Alice siano costantemente superiori in numero a quelli di Bob (non \geq). La risposta si trova sul Feller [Fel68] Sec. III.1, dove si mostra che $P_> = \frac{p-q}{p+q}$. Il codice **matlab** **ballot.m** permette di provare il caso " $>$ " e anche " \geq ", leggermente diversi. Notare che la probabilità dipende soltanto dal vantaggio percentuale. Per ottenere il risultato possiamo partire dalla formula per il problema 17. Utilizzando un diagramma come in Fig.2.13 si può argomentare che ad ogni percorso permesso nello scrutinio (sempre positiva la differenza nei voti) corrisponde un percorso non-negativo nel problema del botteghino. Aiutandosi con la Fig.9 si vede che si tratta di identificare $p + q = 2n + 1$ e $p - q = 2m + 1$. Inoltre il numero totale di traiettorie è dato da $\binom{2n}{n+m}$ nel caso del botteghino e da $\binom{p+q}{p}$ in quello dello scrutinio. Allora si ha che il numero totale di possibili traiettorie per il ballottaggio con $N_A > N_B$ dato da

$$N_{\text{ballot}}(p, q) = P_>(p, q) \binom{p+q}{p}$$

deve coincidere con il numero di traiettorie senza attesa per i $2n$ spettatori con un surplus di banconote da 5 € pari a $2m$ secondo le identificazioni $(p, q) \leftrightarrow (n, m)$. Dunque

$$N_{\text{botteghino}}(n, m) = P_{\geq}(n, m) \binom{2n}{n+m}$$

Si ottiene in pochi passaggi che da $P_{\geq} = \frac{2m+1}{n+m+1}$ segue $P_> = \frac{p-q}{p+q}$. ◀

ESERCIZIO 20. [Il classico problema "della rovina del giocatore"] Giocando a testa-o-croce un giocatore gioca ogni volta un Euro. Assumendo che il suo capitale iniziale sia di x €, il giocatore si ripropone di continuare ad oltranza fino a raggiungere un capitale di m Euro ($m > x$) oppure perde tutto. Qual'è la probabilità di uscire con il portafoglio vuoto?

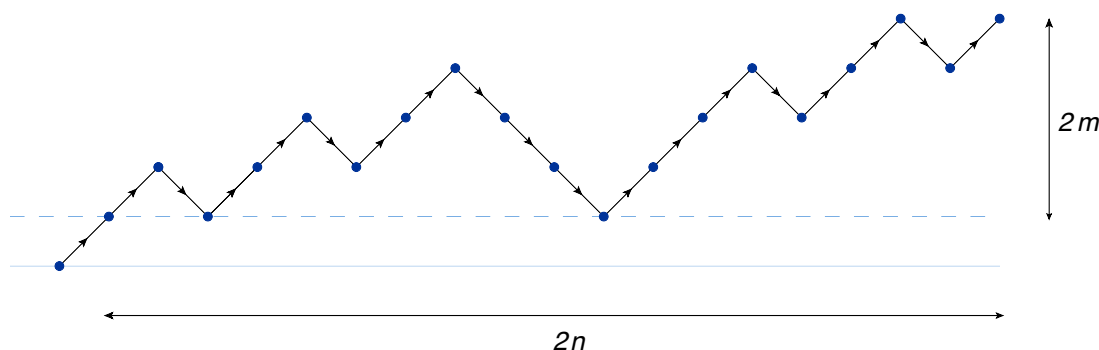


FIGURA 2.13. Traiettorie ammesse per il problema dello scrutinio.

SOLUZIONE. Se indichiamo con $p(x)$ la probabilità della “rovina” per un capitale iniziale x . Al primo tentativo, dopo una vincita la nuova probabilità sarà $p(x+1)$ e in caso di perdita $p(x-1)$. I due casi sono equiprobabili, perciò

$$p(x) = \frac{1}{2}(p(x+1) + p(x-1))$$

Dobbiamo allora cercare una soluzione a questa equazione con le condizioni $p(0) = 1$ e $p(m) = 0$. L'equazione è lineare e ammette due soluzioni indipendenti $p = 1$ e $p = x$. La soluzione va quindi cercata come $p(x) = \alpha + \beta x$. Imponendo le due condizioni, la soluzione è $p(x) = 1 - x/m$. Questo implica, ad esempio, che se si vuole raddoppiare il portafoglio tanto vale puntare tutto su una singola giocata, l'unica differenza essendo il tempo trascorso a giocare. Il programma `gambler.m` simula una giocata a caso e permette di valutare qualcosa in più rispetto alla analisi svolta fin qui, ad esempio risponde alla domanda “dato x e m quanto tempo in media passa prima della rovina?”. Risposta: $t = x(m-x)$. ◀

2.3.6. Alcuni problemi “classici”.

ESERCIZIO 21. In un gioco a premi, il presentatore mette sul tavolo tre buste di cui una sola contiene il premio. Una volta scelta la busta, al concorrente viene indicata quale tra le due buste rimanenti è sicuramente vuota e gli si offre la possibilità di modificare la scelta. Il concorrente argomenta che a questo punto la probabilità di vincere è al 50% e tanto vale mantenere la scelta iniziale. Tuttavia il concorrente sbaglia a non approfittare dell'offerta. Perché?

ESERCIZIO 22. Discutere il seguente problema: si consideri una circonferenza di raggio r . Qual'è la probabilità che una corda presa a caso sia più corta del lato $r\sqrt{3}$ del triangolo equilatero inscritto nella circonferenza? (Avvertenza: il problema è mal posto: ci sono diversi modi non equivalenti di definire una misura di probabilità in questo caso - ciò è del tutto ragionevole: per una risposta non ambigua dovremmo specificare come *esattamente* si sceglie la corda, ad es. si fissa un punto a caso all'interno del cerchio e si prende un diametro che lo contiene e quindi la perpendicolare al diametro passante per il punto, oppure etc. La questione è nota come *Paradosso di Bertrand* e se ne trova traccia nel folder `dropbox/MPF2023/Documents`.)

ESERCIZIO 23. Si giochi ai dadi usandone n numerati da 1 a f . Si mostri che il numero dei possibili risultati con somma N si può ottenere come il coefficiente di x^N nel polinomio ²⁶
 $P_{n,f}(x) = \left(\sum_{k=1}^f x^k\right)^n$.

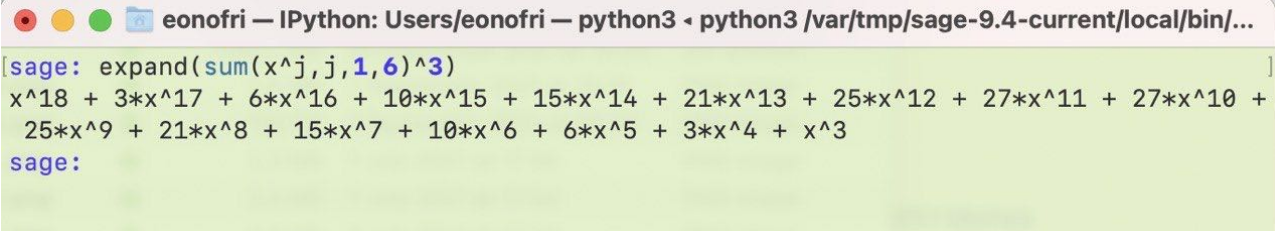
SOLUZIONE. La spiegazione sta nel fatto che nello sviluppo di $P_n(x)$ contribuiscono alla potenza x^N e con peso uguale a uno tutti i termini $x^{k_1} x^{k_2} \dots x^{k_n}$ con $\sum k_j = N$. Per automatizzare il calcolo si può utilizzare la formula di Mathematica

```
| G[f_, n_] := CoefficientList[(Sum[x^j, {j, 1, f}])^n, x]
```

ad es. nel caso $n = 3, f = 6$ si ottengono i valori

output	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
prob.×216	1	3	6	10	15	21	25	27	27	25	21	15	10	6	3	1

per il numero di possibili uscite con somma delle facce compresa tra 3 e 18 e così ritroviamo il conteggio di 27 e di 25 dell'esercizio 10 a pag.26. Ma adesso sappiamo come calcolare le probabilità per la somma delle uscite di un numero qualunque di dadi a forma di ottaedro ($f = 8$) o di dodecaedro ($f = 12$) e per quest'ultimo caso l'enumerazione senza l'ausilio della formula generale terrebbe impegnati per molte notti... ²⁷ Per non scomodare un programma come Mathematica ci si può limitare al più snello SageMath²⁸



```
eonofri — IPython: Users/eonofri — python3 - python3 /var/tmp/sage-9.4-current/local/bin/...
[sage: expand(sum(x^j, j, 1, 6)^3)
x^18 + 3*x^17 + 6*x^16 + 10*x^15 + 15*x^14 + 21*x^13 + 25*x^12 + 27*x^11 + 27*x^10 +
  25*x^9 + 21*x^8 + 15*x^7 + 10*x^6 + 6*x^5 + 3*x^4 + x^3
sage:
```

Infine si può chiedere a matlab di lavorare in modo simbolico:

```
| >> syms('x');
>> n=1:6;
>> expand(sum(x.^n).^3)
>> coeffs(ans)
```

Un consiglio per sfruttare le immense potenzialità della rete: se troviamo la successione 1, 3, 6, 10, 15, ... e ci chiediamo come ricostruire la formula analitica che la produce, possiamo fare ricorso a una sorta di oracolo che si trova all'indirizzo oeis.org ("On-line Encyclopedia of Integer Sequences")²⁹. Si scopre che questo problema è attivamente studiato e per esempio la formula che ci risolve le probabilità per n dadi a f facce si può riscrivere in Mathematica come

```
| G[f_, n_] := Table[Sum[(-)^j Binomial[n, j] Binomial[n+k-1- f j, n-1],
  {j, 0, Floor[k/f]}] , {k, 0, n(f-1)}]
```

dove "Binomial[n,j]" rappresenta il coefficiente binomiale $\binom{n}{j}$. Questi coefficienti sono una generalizzazione dei coefficienti binomiali e hanno una lunga storia, dal momento che la loro

²⁶Questo è un caso particolare della tecnica che incontreremo più avanti, quella che utilizza la funzione generatrice della distribuzione di probabilità, vedi Cap.2.5.3

²⁷Per divertimento potete verificare che se gettate quattro dadi a forma di ottaedro l'uscita più probabile è una somma pari a 18 (8.4%).

²⁸Ricordarsi di definire le variabili simboliche prima di utilizzarle: `j=var('j')`.

²⁹Porre attenzione al fatto che ci sono infinite successioni che coincidono nei primi termini e OEIS suggerisce quelle che hanno una struttura algebrica identificata. Quindi anche avvalendosi di OEIS è prudente non scommettere sul prossimo elemento della successione...

introduzione risale al 1875 (si veda [Com74]). In generale si definisce

$$(1 + x + x^2 + \dots + x^{f-1})^n = \sum_{j=0}^{n(f-1)} R_j^{n,f} x^j.$$

ad es. $R_8^{3,6} = 27$ (notare che l'indice j differisce di n dal risultato del tiro dei dadi). I coefficienti R soddisfano una regola di ricorrenza che costituisce la generalizzazione di quella che vale nel caso binomiale, e cioè

$$R_k^{n,f} = R_k^{n-1,f} + R_{k-1}^{n-1,f} + \dots + R_{k-f+1}^{n-1,f}$$

Utilizzando questa semplice regola si può ricavare lo specchio riportato sopra con semplici manipolazioni senza utilizzare i grandi mezzi! Si tratta di sommare sei termini ($f = 6$) nella riga soprastante a partire dalla stessa posizione in colonna. Semplice come costruire il diagramma di Newton dei coefficienti binomiali. Abbiamo evidenziato il caso che porta a un 27 nella terza riga, nona colonna, che corrisponde al solito risultato per la somma dei dadi uguale a 11 ($= j+n$).

$j \rightarrow$	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
$n = 1$	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0
$n = 2$	1	2	3	4	5	6	5	4	3	2	1	0	0	0	0	0
$n = 3$	1	3	6	10	15	21	25	27	27	25	21	15	10	6	3	1



2.3.7. Problemi con la crittografia.

ESERCIZIO 24. Si chiede: scegliendo a caso una lettera di un testo scritto in italiano qual'è la probabilità che si tratti di una 'e'? Stesso problema nel caso di altre lingue europee.

SOLUZIONE. Deve essere chiaro che a questo problema non si risponde con una formula analitica per la probabilità. La frequenza delle lettere è determinata da come il linguaggio si è sviluppato nei secoli ed è continuamente in evoluzione. Inoltre la risposta può dipendere dal carattere del testo (letterario, scientifico, etc). L'unico modo di procedere consiste nell'esaminare un gran numero di testi e trarne la statistica della presenza delle varie lettere. Se non si vuole andare alla ricerca del risultato nelle pubblicazioni di linguistica si può effettuare un esperimento in piccolo prendendo un testo rappresentativo della lingua e calcolando la frequenza. Ciò si può fare facilmente utilizzando un calcolatore. Nel folder `~/Dropbox/MPF2023/Programmi/Matlab/linguistica` si trovano testi significativi in quattro lingue europee e ci si può esercitare nell'analisi statistica. Nella fig.2.3.7 si riscontra una sostanziale somiglianza tra Italiano, Francese, Inglese e Spagnolo. La lettera 'e' è dominante e si noterà come ci sono alcuni caratteri distintivi che permettono di individuare la lingua: la 's' così frequente nello Spagnolo, l'alta frequenza delle vocali 'a-e-i-o' in Italiano, 'w-y' praticamente assenti se non in Inglese, $p('n') > p('o')$ in Francese. ◀

ESERCIZIO 25. Immaginiamo di avere tre contenitori uguali e tre biglie di colore diverso tra loro (diciamo blu-rosso-verde). Con un procedimento del tutto casuale le biglie vengono poste nei tre contenitori - ad esempio utilizzando la caduta lungo un piano inclinato con opportuni

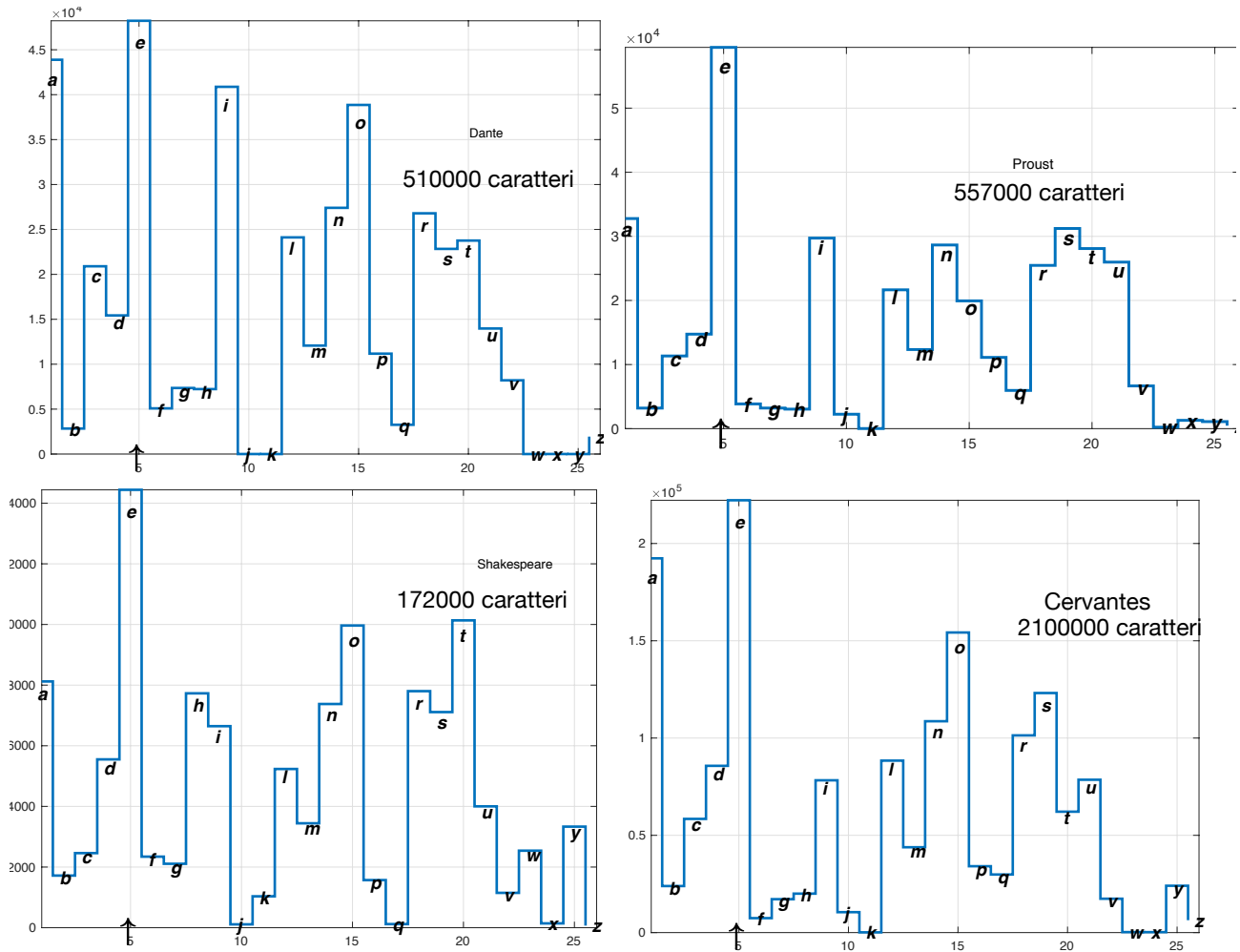


FIGURA 2.14. Esempio di analisi statistica delle lettere di cui è formato un documento.


















































































ostacoli che deviano la traiettoria di ciascuna biglia ³⁰ Se ammettiamo che ogni biglia cada con la stessa probabilità in un qualunque contenitore, si chiede quale sia la probabilità che

- (i) tre biglie finiscano in un medesimo contenitore
- (ii) le tre biglie finiscano in tre contenitori differenti.

SOLUZIONE. Il calcolo è abbastanza banale, ma il problema si presta ad approfondimenti molto interessanti. Prima domanda: posto che la prima biglia può arrivare in un contenitore qualunque, la seconda e la terza hanno entrambe $1/3$ di probabilità di finire nello stesso contenitore della prima, dunque la probabilità di trovare le tre biglie insieme è $1/9$. Con lo stesso

³⁰Un esempio storicamente importante di esperimenti del genere è quello della “tavola di Galton” - si veda *Galton board* su Wolfram MathWorld o *bean machine* su Wikipedia. Nel caso specifico tuttavia la disposizione degli ostacoli obbliga le biglie a numerosi urti e il risultato è una distribuzione non uniforme che per un grande numero di biglie e di ostacoli tende alla Gaussiana.

argomento la probabilità di avere una biglia in ciascun contenitore è $2/9$. Un'analisi più approfondita è la seguente. Supponiamo che le tre biglie siano di colore blu, verde e rosso; avremo dunque le 27 possibilità

i)	  		
i)		  	
i)			  
	 		
	 		
	 		
	 		
	 		
	 		
		 	
		 	
		 	
		 	
		 	
		 	
			 
			 
			 
			 
			 
			 
ii)			
ii)			
ii)			
ii)			
ii)			
ii)			

da cui la risposta $3/27$ e $6/27$ (la prima colonna indica a quale domanda corrispondano le disposizioni di biglie). ◀

Per rendere il calcolo nell'esercizio precedente meno faticoso è possibile affidarsi all'algebra: consideriamo

$$(x + y + z)^3 = x^3 + y^3 + z^3 + 3x^2y + 3xy^2 + 3x^2z + 3xz^2 + 3y^2z + 3yz^2 + 6xyz$$

Il coefficiente di $x^{n_1}y^{n_2}z^{n_3}$ è un *coefficiente multinomiale* e precisamente $3!/(n_1!n_2!n_3!)$ e ci dà il numero di possibili distribuzioni di biglie con n_i biglie nel i -esimo contenitore. È chiaro che questo ci permette di generalizzare immediatamente al caso di un numero qualunque di contenitori e di biglie. Se indichiamo con N il numero di contenitori e con n il numero di biglie, cerchiamo la probabilità $P(n_1, n_2, \dots, n_N)$ di trovare n_1 biglie nel primo contenitore, n_2 nel secondo e così via. Il numero totale di possibilità è ovviamente N^n e la probabilità cercata si ottiene come il coefficiente di $x_1^{n_1} x_2^{n_2} \dots x_N^{n_N}$ nello sviluppo di $((x_1 + x_2 + \dots + x_N)/N)^n$ ossia

$$P(n_1, n_2, \dots, n_N) = \frac{n!}{N^n \prod_{k=1}^N n_k!}$$

Per quattro biglie in tre contenitori si trova ad es. $P(3, 1, 0) = \frac{4!}{3^4 3!} = 4/81$ e si verifica il conteggio considerando la lista

$$\{(brv, g, -), (rvvg, b, -), (vgb, r, -), (gbr, v, -)\}$$

o anche $P(2, 2, 0) = \frac{4!}{3^4 2! 2!} = 6/81$ e queste sono le sei disposizioni:

$$\{(br, vg, -), (bv, gr, -), (bg, rv, -), (rv, gb, -), (rg, bv, -), (vg, br, -)\}.$$

2.3.8. Biglie quantistiche. Chiediamoci ora cosa succederebbe se le biglie fossero del tutto indistinguibili (stesso colore, stesso raggio, stesso peso, etc.). Dal punto di vista della probabilità sembrerebbe non cambiare nulla, in fondo possiamo sempre pensare di etichettare le biglie con qualche caratteristica nascosta che non ne modifichi le proprietà sensibili e quindi ritornare al caso di biglie distinguibili. Se invece stiamo strettamente all'aspetto identico per tutte la tabella precedente diventerebbe

● ● ●		
	● ● ●	
		● ● ●
● ●	●	
● ●		●
●	● ●	
	● ●	●
●		● ●
	●	● ●
●	●	●

10 possibilità in tutto. Secondo questo nuovo schema sembrerebbe che la probabilità di trovare tre biglie nel primo contenitore sia $1/10$ anziché $1/27$, ben diverso! Il fatto è che per biglie con cui possiamo sperimentare, biglie di vetro colorato con cui una volta si giocava da bambini, il conteggio corretto è senz'altro $1/27$. *La sorpresa è che se invece di biglie parliamo di molecole o di atomi identici il conteggio corretto è proprio $1/10$!* Questo fatto è da inquadrare nella “statistica quantistica” di Bose-Einstein che vale per particelle identiche *bosoniche* ossia a spin

intero. In questa statistica conta solo il *numero di occupazione* e ogni scelta di numeri di occupazione si presenta con la stessa probabilità, come se l'algebra di riferimento prevedesse

$$(x + y + z)^3 = x^3 + y^3 + z^3 + x^2y + xy^2 + x^2z + xz^2 + y^2z + yz^2 + xyz \quad (!?)$$

Il numero di termini si trova facilmente essere $\binom{N+n-1}{n}$ e la probabilità risulta uguale al reciproco di questo valore. Sul volume I del trattato di Feller [Fel68] si troveranno maggiori dettagli.

Bose-Einstein non è l'unica possibile statistica microscopica. Se siamo in presenza di “biglie microscopiche” a spin semi-intero allora il principio di Pauli esclude tutti i casi con più di una biglia nello stesso contenitore, come fosse in opera un'algebra bizzarra secondo cui

$$(x + y + z)^3 = xyz \quad (!?)$$

La statistica viene detta di Fermi-Dirac e le “biglie” che sono assogettate a questa statistica sono detti *fermioni*, mentre quelle a spin intero sono dette *bosoni*. Questi argomenti vengono sviluppati nei corsi di Meccanica Quantistica, Meccanica Statistica e Teoria dei Campi Quantistica. Torneremo brevemente alla probabilità quantistica nella sezione 2.8.

ESERCIZIO 26. Prendiamo dal libro di Amir [Ami21] un problema di genetica, può trattarsi di incroci tra semi di pianta o unione tra individui evoluti. Ogni individuo è caratterizzato da un particolare gene che può avere due forme, diciamo “a” o “A”; un individuo con entrambi geni “a” soffre di una rara malattia, ma la maggioranza degli individui presentano due copie del gene “A” o una coppia “aA”. Gli individui con gene “aA” non soffrono della malattia ma possono trasmettere il gene “a” alla progenie. Ora il problema è il seguente: il gene recessivo “a” tenderà o no a scomparire nelle successive generazioni? Chiamando “madre” e “padre” i due individui che passano uno dei loro geni alla progenie si possono facilmente rappresentare tutte le possibilità con l'aiuto di questa tabella (tra parentesi la probabilità che si presenti il risultato indicato):

ma/pa	aa	AA	aA
aa	aa(1)	aA(1)	aA(1/2), aa(1/2)
AA	aA(1)	AA(1)	AA(1/2), aA(1/2)
aA	aa(1/2), aA(1/2)	AA(1/2), aA(1/2)	aa(1/4), AA(1/4), aA(1/2)

Indichiamo con p la percentuale iniziale di coppie AA, con $2q$ quella della coppia aA e con r quella della coppia aa ($p + 2q + r = 1$). Nel libro citato si riporta un risultato pubblicato nel 1908 da G. Hardy e W. Weinberg; la popolazione raggiunge in una sola generazione uno stato di equilibrio caratterizzato dalla identità $pr = q^2$ e la percentuale di individui con gene aa si mantiene stabile nel corso delle successive generazioni. L'argomento puramente algebrico è riprodotto nella soluzione, piuttosto presentiamo qui un approccio di simulazione. Il codice `Hardy.m`, che si può trovare su [dropbox/Programmi/Matlab/Genetics](#), è organizzato in modo molto semplice: si assegna un codice numerico alla tre varietà, precisamente $aa \rightarrow 1$, $AA \rightarrow 2$, $aA \rightarrow 3$ e si preparano due array M e F di dimensione $(n_m, 1)$ e $(n_f, 1)$ e riempite con 1, 2, 3 in proporzione rispettivamente a r , p , $2q$ e questa costituisce la prima generazione. La seconda generazione si ottiene estraendo a caso un certo numero di volte (proporzionale a n_f) un elemento da M e uno da F : sulla base della tabella l'unione tra questi due individui darà luogo a un nuovo individuo sulla base delle probabilità ricavate dalla tabella. Si analizza la percentuale dei tre geni in questa generazione e si passa alla generazioni successive. Se si tratta di microorganismi il numero di generazioni può essere scelto anche molto grande, nel

caso di umani, se non si vogliono esplorare millenni, su cui comunque non si può pensare di avere dati disponibili, il numero di generazioni è limitato, diciamo non oltre 5 o 6. Il programma mostra che per popolazioni sufficientemente ampie l'equilibrio di Hardy-Weinberg è raggiunto immediatamente; naturalmente si riscontrano fluttuazioni intorno alla configurazione di equilibrio, come accade in termodinamica, fluttuazioni percentualmente più importanti per piccole popolazioni. Provare a sperimentare con il programma utilizzando Matlab (*Hardy1.m*). Una variante prevede che gli individui caratterizzati dalla copia **aa** risultino sterili (*Hardy.m*). Il programma permette di esplorare la dinamica genetica anche in questo caso.

SOLUZIONE. *Dalla tabella del libro di Amir si calcola facilmente la media delle coppie di geni nella prossima generazione, tenendo conto che la prima riga riguarda una percentuale di popolazione r , la seconda p e la terza $2q$, e lo stesso vale per le tre colonne. Dunque si vede che la coppia **aa** proviene dagli elementi di matrice $(1,1), (1,3), (3,1), (3,3)$ con probabilità r^2, rq, rq, q^2 in totale $r_2 = (r+q)^2$. Per **AA** si trova allora $p_2 = (p+q)^2$ e per **aA** la probabilità risulta $q_2 = (r+q)(p+q)$. Ne segue che nella seconda generazione, qualunque siano le proporzioni nella prima, si ha l'identità $r_2 p_2 = q_2^2$. Questo risultato riguarda le medie su una grande popolazione e ci si deve aspettare la presenza di fluttuazioni più o meno significative a seconda della dimensione della popolazione. Lo studio attraverso la simulazione numerica permette di approfondire il discorso sulle fluttuazioni. ◀*

2.4. Sviluppi del formalismo

2.4.1. Indipendenza statistica. Quanto è influenzato l'evento $A \subset \Omega$ dal verificarsi di un altro evento B ? Ad es. se $B \subset A$ ovviamente A si verifica sempre in seguito a B , mentre se $A \cap B = \emptyset$ i due eventi non possono verificarsi contemporaneamente e si diranno incompatibili. Si definisce allora la probabilità condizionale di A dato B attraverso la formula

$$(3) \quad \mathbb{P}(A|B) \equiv \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

Se $\mathbb{P}(A|B) = \mathbb{P}(A)$, ciò significa che la conoscenza dell'avverarsi dell'evento B non ha alcuna influenza sulla nostra aspettativa relativamente ad A . Si dice allora che A e B sono statisticamente indipendenti. Esplicitamente la condizione suona così

$$(4) \quad \mathbb{P}(A \cap B) = \mathbb{P}(A) \mathbb{P}(B).$$

La relazione di indipendenza statistica è dunque simmetrica. In generale, dati n eventi A_1, A_2, \dots, A_n , questi si dicono *mutuamente indipendenti* se per ogni combinazione di indici e ogni k , ($2 \leq k \leq n$),

$$(5) \quad \mathbb{P}(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k}) = \mathbb{P}(A_{i_1}) \mathbb{P}(A_{i_2}) \dots \mathbb{P}(A_{i_k})$$

Per capire meglio il significato della definizione conviene considerare un esempio di eventi che sono indipendenti a due a due ma *non mutuamente* è il seguente: nel lancio di due dadi, $A = \{\text{il primo dado dà un numero dispari}\}$, $B = \{\text{il secondo dado dà un numero dispari}\}$, $C = \{\text{la somma dei dadi è dispari}\}$. Si ha come è ovvio

- $\mathbb{P}(A) = 1/2$
- $\mathbb{P}(B) = 1/2$,
- $\mathbb{P}(C) = 1/2$ (primo dado pari e il secondo dispari (prob=1/4) o viceversa (+1/4).

Ora si verifica facilmente che

- $\mathbb{P}(A \cap B) = 1/4$,
- $\mathbb{P}(A \cap C) = 1/4$,
- $\mathbb{P}(B \cap C) = 1/4$;

e tuttavia $\mathbb{P}(A \cap B \cap C) = 0 \neq 1/8!$ Dunque i tre eventi *non* sono mutuamente indipendenti.

ESERCIZIO 27. Dimostrare la formula

$$\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A) \mathbb{P}(B|A) \mathbb{P}(C|A \cap B) .$$

Bayes

2.4.2. Probabilità condizionata e la formula di Bayes. Si noti che la probabilità $\mathbb{P}(A \cap B)$ è simmetrica in A e B dunque

$$\mathbb{P}(A|B) \mathbb{P}(B) = \mathbb{P}(B|A) \mathbb{P}(A)$$

che porta immediatamente alla *formula di Bayes*

$$(6) \quad \mathbb{P}(A|B) = \frac{\mathbb{P}(B|A) \mathbb{P}(A)}{\mathbb{P}(B)}$$

Supponiamo che la collezione di eventi $\mathfrak{B} = \{B_j | j = 1, \dots, n\}$ sia caratterizzata dal fatto che ogni coppia di eventi siano incompatibili ($B_j \cap B_k = \emptyset$) e inoltre $\bigcup_{k=1}^n B_k = \Omega$. \mathfrak{B} si dice un insieme completo di eventi e vale la relazione generale, detta *formula della probabilità totale*:

$$(7) \quad \mathbb{P}(A) = \sum_{k=1}^n \mathbb{P}(A|B_k) \mathbb{P}(B_k) .$$

Si noterà la somiglianza formale con la formula che esprime lo sviluppo di uno stato quantistico su una base

$$|\psi\rangle = \sum_n |n\rangle \langle n|\psi\rangle$$

ma la differenza sta nel fatto che in teoria delle probabilità si compongono i valori di \mathbb{P} , non le ampiezze.

ESERCIZIO 28. Sia dato un grafo rappresentato in figura. Si considera un cammino casuale sul grafo che inizia in O. Ad ogni diramazione il cammino sceglie un ramo con la stessa probabilità (tranne che ritornare indietro). Qual'è la probabilità di raggiungere il nodo A?



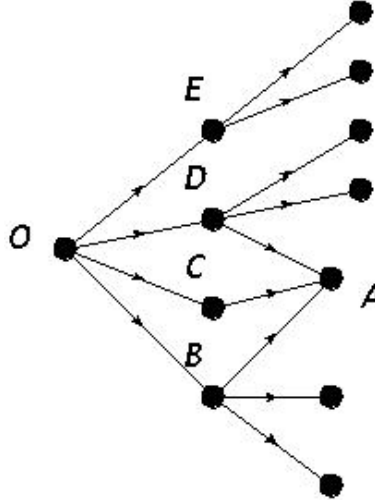
Se $\mathcal{I}_1, \mathcal{I}_2, \dots, \mathcal{I}_n$ è un insieme completo di eventi incompatibili tra loro allora vale la formula

$$(8) \quad \mathbb{P}(\mathcal{I}_k|A) = \frac{\mathbb{P}(A|\mathcal{I}_k) \mathbb{P}(\mathcal{I}_k)}{\mathbb{P}(A)} = \frac{\mathbb{P}(A|\mathcal{I}_k) \mathbb{P}(\mathcal{I}_k)}{\sum_{k=1}^n \mathbb{P}(A|\mathcal{I}_k) \mathbb{P}(\mathcal{I}_k)}$$

La formula precedente, pur essendo una conseguenza immediata della definizione di probabilità condizionata, può risultare utile in molti contesti ed è l'equazione centrale nel ragionamento "Bayesiano". La formula si può leggere in modo più intuitivo così

Bayes' Formula

$$(9) \quad \mathbb{P}(\mathcal{I}_k|D) \propto \mathbb{P}(D|\mathcal{I}_k) \mathbb{P}(\mathcal{I}_k)$$



Pensiamo a \mathcal{I}_k come “ipotesi” di cui conosciamo la probabilità a priori $\mathbb{P}(\mathcal{I}_k)$ (“priors”). Se D viene interpretato come un complesso di “dati” di nuova acquisizione, la formula ci dà la valutazione delle nuove probabilità a posteriori (“posteriors”) $\mathbb{P}(\mathcal{I}_k | D)$ sulla base delle probabilità $\mathbb{P}(D | \mathcal{I}_k)$ dette “likelihood”; il denominatore che abbiamo nascosto nel passare all’ultima formula è semplicemente il fattore di normalizzazione.

ESERCIZIO 29 (The Monty Hall problem). Un quesito famoso e che ha sollevato anche diatribe interpretative è il seguente: in un gioco a premi il concorrente è invitato a scegliere tra tre scatole chiuse di cui una contiene il premio e le altre due sono vuote. Una volta che il concorrente ha fatto la sua scelta, prima di aprire la scatola, il conduttore, che sa dove si trova il premio, apre quella delle altre due scatole che sa non contenere il premio e la sceglie a caso se il premio è nella scatola scelta dal concorrente. Al concorrente viene concesso di cambiare idea e di optare per la terza scatola. L’offerta deve essere respinta o accettata? Qual’è la strategia vincente su una lunga lista di concorrenti?

SOLUZIONE. Una risposta molto frequente è che si può tenere la prima scelta o cambiare senza differenza alcuna, in quanto ormai l’alternativa è tra solo due possibilità ritenute equivalenti. Il ragionamento corretto invece indica che la strategia giusta è quella di accettare il cambio. La cosa è del tutto evidente se si pensa che l’azione del conduttore equivale a dire che il concorrente ha in realtà a disposizione due scatole su tre e il conduttore gli ha risparmiato la fatica di aprirne una ...

Un argomento basato sulla formula di Bayes corre come segue: le scatole siano contrassegnate a, b e c e siano A, B, C gli eventi che corrispondono alla presenza del premio in una delle tre scatole, cioè “ A : il premio è nella scatola a ” etc. Allora a priori $\mathbb{P}(A) = \mathbb{P}(B) = \mathbb{P}(C) = 1/3$ (“priors”). Diciamo che il concorrente sceglie a e il conduttore mostra che b è vuota. Possiamo calcolare facilmente le probabilità condizionate dal fatto che venga aperta la scatola b partendo da

$$\mathbb{P}(b|A) = 1/2, \mathbb{P}(b|B) = 0, \mathbb{P}(b|C) = 1.$$

Si trova allora applicando la formula di Bayes

- $\mathbb{P}(A|b) \propto \mathbb{P}(b|A) \mathbb{P}(A) = 1/6$

- $\mathbb{P}(B|b) \propto \mathbb{P}(b|B) \mathbb{P}(B) = 0$
- $\mathbb{P}(C|b) \propto \mathbb{P}(b|C) \mathbb{P}(C) = 1/3$

Le probabilità a posteriori sono dunque rispettivamente $1/3, 0, 2/3$. ◀

ESERCIZIO 30. Un'urna contiene 10 palline di cui 6 sono bianche e 4 nere. Si estrae a caso una pallina e, senza rimetterla nell'urna, se ne estrae una seconda. Qual'è la probabilità che quest'ultima sia bianca? (da [RPP12]).

SOLUZIONE. Si applica la formula delle probabilità totali utilizzando $\mathbb{P}(B) = \text{"prima pallina estratta è bianca"} = 6/10$, $\mathbb{P}(N)$ invece $\text{"prima pallina estratta è nera"} = 4/10$ e alla seconda estrazione $\mathbb{P}(B|B) = 5/9$, $\mathbb{P}(B|N) = 6/9$:

$$\begin{aligned} P &= \mathbb{P}(B) \mathbb{P}(B|B) + \mathbb{P}(N) \mathbb{P}(B|N) \\ &= \frac{6}{10} \frac{5}{9} + \frac{4}{10} \frac{6}{9} = 3/5. \end{aligned}$$

◀

ESERCIZIO 31. In dieci contenitori si hanno nell'ordine 2 palle bianche e 2 nere nei primi nove, mentre nell'ultimo sono contenute 5 bianche e una nera. Si estrae a caso una palla senza sapere da quale contenitore. Se si è scelta una palla bianca, qual'è la probabilità che ce ne siano altre quattro nello stesso contenitore?

SOLUZIONE. Indichiamo con $\mathbb{P}(B)$ la probabilità di estrarre una palla bianca senza altre condizioni; indichiamo poi con C_{5-1} il contenitore con 5 palle bianche e una nera. La probabilità richiesta è $\mathbb{P}(C_{5-1}|B)$ cioè la probabilità condizionata che avendo estratto una palla bianca questa provenga dal contenitore C_{5-1} che ne contiene altre 4 dello stesso colore. Si trova immediatamente che $\mathbb{P}(B) = \frac{1}{2} \frac{2}{10} + \frac{5}{6} \frac{1}{10} = \frac{8}{15}$, e dalla formula di Bayes ricaviamo

$$\mathbb{P}(C_{5-1}|B) = \mathbb{P}(B|C_{5-1}) \frac{\mathbb{P}(C_{5-1})}{\mathbb{P}(B)} = \frac{5}{6} \times \frac{1/10}{8/15} = \frac{5}{32}$$

dove appunto si intende che C_{5-1} indica che viene scelta la palla dal contenitore con 5 palle bianche e una nera. ◀

ESERCIZIO 32. Ci sono n urne contenenti ciascuna un numero b di palle bianche e r di palle rosse. Si estrae senza guardare una palla dalla prima urna e la si pone nella seconda. Poi si estrae una palla dalla seconda e la si mette nella terza, sempre alla cieca. Si procede così fino all'ultima urna e da questa si estrae un'ultima volta una palla da riporre nella prima urna, ma questa volta si registra il colore della palla estratta. Qual'è la probabilità che si tratti di una palla rossa?

SOLUZIONE. Indichiamo con p_j il vettore a due componenti che rappresenta la probabilità di estrarre una palla bianca o una rossa dalla j -sima urna. Si ha ovviamente $p_1 = \begin{pmatrix} b/N \\ r/N \end{pmatrix}$ dove $N = b + r$. Per la seconda urna si ragiona così: si estrae una palla bianca con probabilità $(b+1)/(N+1)$ se dalla prima urna abbiamo estratto una palla bianca oppure $b/(N+1)$ in caso

contrario, dunque

$$p_2 = \begin{pmatrix} \frac{b+1}{N+1} \frac{b}{N} + \frac{b}{N+1} \frac{r}{N} \\ \frac{r+1}{N+1} \frac{r}{N} + \frac{r}{N+1} \frac{b}{N} \end{pmatrix} = \begin{pmatrix} \frac{b}{N} \\ \frac{r}{N} \end{pmatrix}$$

il che significa che la probabilità di estrarre una palla bianca dalla seconda urna, anche se la cosa può risultare poco intuitiva, è la stessa b/N che si ha per la prima urna. Allora non cambierà nulla nelle successive estrazioni e la risposta è $p_n = \begin{pmatrix} b/N \\ r/N \end{pmatrix}$ anche all'ultima urna. Però per capire più a fondo cosa sta succedendo conviene guardare al problema da un altro punto di vista: possiamo dire che si passa da p_1 a p_2 utilizzando la matrice ³¹

$$M = \begin{pmatrix} \frac{b+1}{N+1} & \frac{b}{N+1} \\ \frac{r}{N+1} & \frac{r+1}{N+1} \end{pmatrix}$$

Ma si trova che p_1 è giusto l'autovettore di M con autovalore 1! Ne segue $p_2 = M p_1 = p_1$ e dunque ciò si estende ad ogni successivo indice j , $p_{j+1} = M p_j = p_1$. Si può affrontare il problema attraverso una simulazione. Il codice `matlab` `Nurne.m` effettua un numero N_{exp} di esperimenti e ciascuno comporta un certo numero `sw` di passate cicliche $1 \rightarrow N \rightarrow 1$ registrando il numero di palle rosse in ciascuna urna. Così è possibile valutare valor medio e deviazione standard del numero di palle bianche e rosse in ciascuna urna. Il risultato di un esperimento con 12 urne, $b=20$, $r=10$ e 20 esperimenti per 200 sweeps è riportato in Fig.2.15.



ESERCIZIO 33. È noto da studi statistici accurati che una certa malattia X ha un'incidenza sull'1‰ della popolazione. Esiste un test che risulta accurato solo nel 95% dei casi. Se fate il test e risultate positivi dovrete davvero preoccuparvi? Si veda ad es. [BV12] per una trattazione che fa uso della formula di Bayes.

SOLUZIONE. Indichiamo con vari simboli le probabilità condizionali coinvolte:

- $\mathbb{P}(p|M) = 0.95$ la probabilità di risultare positivi nel caso in cui si sia malati;
- $\mathbb{P}(n|M) = 0.05$ la probabilità di risultare negativi nel caso in cui si sia malati;
- $\mathbb{P}(n|S) = 0.95$ la probabilità di risultare negativi nel caso in cui si sia sani.
- $\mathbb{P}(p|S) = 0.05$ la probabilità di risultare positivi nel caso in cui si sia sani;

Troviamo allora

- $\mathbb{P}(M|p) \propto \mathbb{P}(p|M) \mathbb{P}(M)$ (probabilità di essere malati in caso di test positivo)
- $\mathbb{P}(S|p) \propto \mathbb{P}(p|S) \mathbb{P}(S)$ (probabilità di essere sani in caso di test positivo)

Ne segue

$$\frac{\mathbb{P}(S|p)}{\mathbb{P}(M|p)} = \frac{\mathbb{P}(p|S) \mathbb{P}(S)}{\mathbb{P}(p|M) \mathbb{P}(M)}$$

Sostituendo $\mathbb{P}(M) = 1/1000$, $\mathbb{P}(S) = 0.999$, $\mathbb{P}(p|S) = 0.05$, $\mathbb{P}(p|M) = 0.95$ troviamo che $\mathbb{P}(S|p) = \mathbb{P}(M|p) \times \frac{0.05 \times 0.999}{0.95 \times 0.001} \approx 52.6$ $\mathbb{P}(M|p) = 52.6 (1 - \mathbb{P}(S|p))$. Ne ricaviamo che $\mathbb{P}(S|p) \approx 98\%$,

³¹Si tratta di un caso particolare di matrice stocastica, di cui parleremo nel capitolo 3.

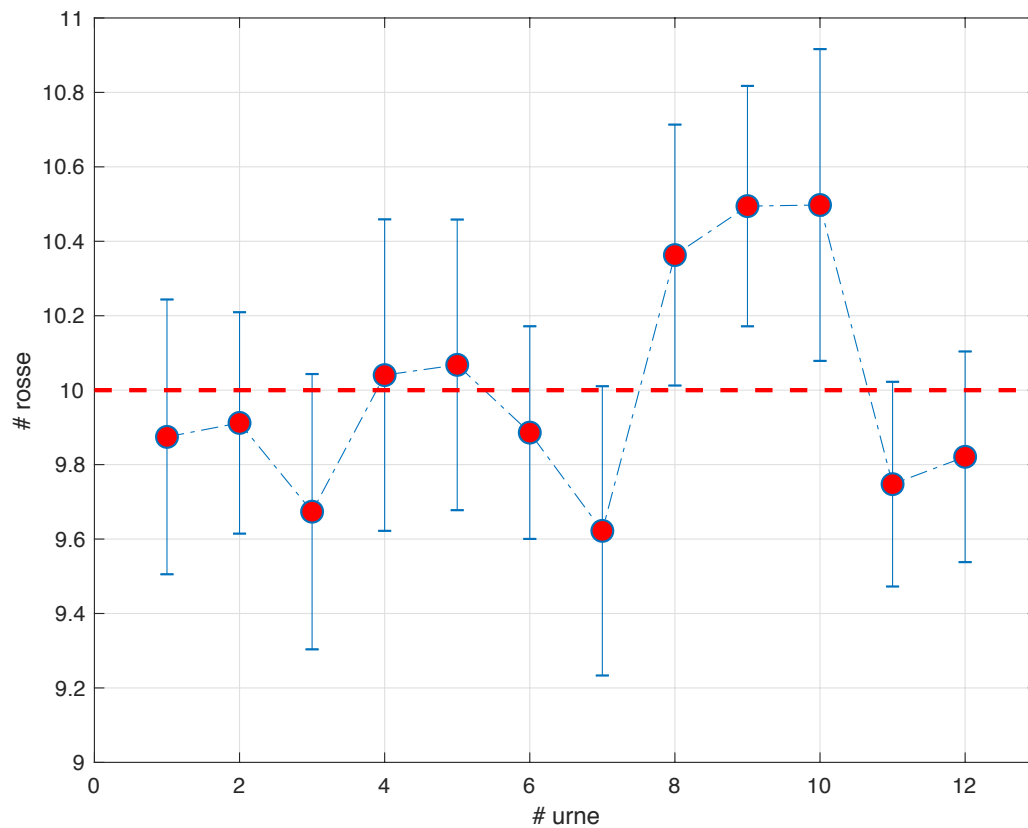


FIGURA 2.15. Il problema delle n urne. In ordinata il numero medio di palle rosse in ciascuna urna e la deviazione standard dei valori nel corso dell'esperimento.

dunque il test non è molto significativo e possiamo stare relativamente tranquilli! Si può ragionare anche molto più semplicemente: su una popolazione di N persone ci si aspettano $N/1000$

ammalati di cui 0.95 N/1000 saranno i “veri positivi”; per contro tra i 999/1000 N sani risulteranno 5/100 999/1000 N falsi positivi. Ne risulta che la probabilità di essere ammalati quando si è risultati positivi al test risulta inferiore al 2%. Per una risposta più affidabile sarebbe necessario un test cento volte più accurato. ◀

ESERCIZIO 34. La zecca ha prodotto un miliardo di pezzi da 1€, quelli con l’“uomo Vitru-



viano” di Leonardo su una faccia e la dicitura “1 Euro” sull’altra. Si è calcolato che un pezzo su un milione è difettoso in quanto riproduce l’uomo di Leonardo su entrambe le facce. Voi prendete, senza guardarla, una moneta da un Euro e la tirate per dieci volte (si diceva “testa o croce”, ma adesso questo modo di dire si applica solo a qualche vecchia moneta), ottenendo per dieci volte lo stesso risultato, l’uomo di Leonardo; qual’è la probabilità che si tratti di una delle monete difettose?

SOLUZIONE. Indichiamo con F il caso che la moneta sia fallata, con V quello che si tratti di una moneta valida e con L_{10} l’evento di dieci uscite uguali. Si avrà vilupp

$$\mathbb{P}(F) = 10^{-6}$$

$$\mathbb{P}(V) = 1 - 10^{-6}$$

$$\mathbb{P}(L_{10}|F) = 1$$

$$\mathbb{P}(L_{10}|V) = (1/2)^{10}$$

La formula di Bayes e la formula delle probabilità totali ci danno allora

$$\mathbb{P}(F|L_{10}) = \frac{\mathbb{P}(L_{10}|F) \mathbb{P}(F)}{\mathbb{P}(L_{10})} = \frac{1 \times 10^{-6}}{\mathbb{P}(L_{10}|F) \mathbb{P}(F) + \mathbb{P}(L_{10}|V) \mathbb{P}(V)} \approx 1/1000.$$

Conclusione: un evento che ha probabilità di uno su un milione diventa probabile all’un per mille se si verificano dieci uscite uguali - si usa dire che la probabilità “a priori” è $1/10^6$ e quella “a posteriori”, essendosi verificato l’evento di dieci uscite uguali, è $1/1000$. Anche se verrebbe spontaneo scommettere che la moneta è difettosa, il risultato mostra che non sarebbe prudente farlo. La formula generale per il caso di N uscite uguali è

$$\mathbb{P}(F|L_N) = \frac{\mathbb{P}(F)}{\mathbb{P}(F) + 2^{-N}(1 - \mathbb{P}(F))}$$

Si provi a valutare quante uscite uguali ci permettono di scommettere sul fatto che la moneta è difettosa con probabilità di vittoria al 50%. ◀

2.4.3. La formula di inclusione/esclusione. Siano A_1, A_2, \dots, A_N sottoinsiemi dello spazio degli eventi Ω . Ci si chiede qual’è la probabilità che si realizzi un evento contenuto in uno qualunque dei sottoinsiemi? La risposta $\mathbb{P}(\cup_i A_i) = \sum_{i=1}^N \mathbb{P}(A_i)$ (?) è ovviamente affetta da errore se gli eventi non sono a due a due disgiunti. Viene spontaneo correggere la formula compensando per le intersezioni

$$\mathbb{P}(\cup_i A_i) = \sum_i \mathbb{P}(A_i) - \sum_{i>j} \mathbb{P}(A_i \cap A_j)(?)$$

ma anche questa formula non funziona se esistono triple intersezioni $A_i \cap A_j \cap A_k$. La risposta corretta è data dalla formula di inclusione/esclusione ([Roz69] (Teor.2.2), [Poi12] (pag.59); si

veda [Sta99] per un'introduzione più ampia a questo risultato; oppure vedere GCNewsletterII-5, II-8, nel folder *Documents*.)

TEOREMA 2.1. Per ogni insieme di eventi $\{A_j \mid j = 1, \dots, n\}$ vale l'identità

$$(10) \quad \begin{aligned} \mathbb{P}(\cup_i A_i) &= \sum_i \mathbb{P}(A_i) - \sum_{i>j} \mathbb{P}(A_i \cap A_j) + \\ &\quad \sum_{i>j>k} \mathbb{P}(A_i \cap A_j \cap A_k) + \dots + \\ &\quad - (-1)^n \sum_{i_1>i_2>\dots>i_n} \mathbb{P}(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_n}) \end{aligned}$$

Dimostrazione: consideriamo un evento elementare che non appartenga ad alcun sottoinsieme A_i : nella formula questo pesa giustamente per zero. Se un evento elementare appartiene a k tra gli A_i nella formula esso contribuisce con un peso

$$k - \binom{k}{2} + \binom{k}{3} - \dots = 1 - \sum_{j=0}^k (-1)^j \binom{k}{j} = 1.$$

che è quanto deve risultare senza sovra conteggi. Il risultato risale a Poincaré, [Poi12]. Il “problème de la rencontre” è trattato nel problema B.8.6.



60

CHAPITRE III.

D'où, une généralisation du théorème des probabilités totales : la probabilité pour que l'un, au moins, des événements se produise est

$$\Sigma p_i - \Sigma p_{ik} + \Sigma p_{ijk} - \dots$$

23. Problème de la rencontre. — Dans une urne, il y a μ boules numérotées de 1 à μ ; je les tire les unes après les autres, jusqu'à ce que l'urne soit vide. Il y a rencontre si, au i^{e} tirage, je tire la boule numérotée i .

FIGURA 2.16. Estratto da Poincaré, *Calcul des Probabilités*

Derangements

ESERCIZIO 35. Si consideri il gruppo S_n delle permutazioni su n oggetti. Si chiede qual'è la probabilità che una permutazione scelta a caso non lasci fisso alcun elemento? (Una variante del problema dice: 40 signori lasciano il proprio cappello all'ingresso del teatro. All'uscita viene a mancare l'energia elettrica e i cappelli vengono riconsegnati al buio. Qual'è la probabilità che nessuno ritorni in possesso del proprio cappello? Un'altra variante ancora considera il caso

S_4	<i>punti fissi</i>
1 2 3 4	4
1 2 4 3	2
1 3 2 4	2
1 3 4 2	1
1 4 2 3	1
1 4 3 2	2
2 1 3 4	2
2 1 4 3	0
2 3 1 4	1
2 3 4 1	0
2 4 1 3	0
2 4 3 1	1
3 1 2 4	1
3 1 4 2	0
3 2 1 4	2
3 2 4 1	1
3 4 1 2	0
3 4 2 1	0
4 1 2 3	0
4 1 3 2	1
4 2 1 3	1
4 2 3 1	2
4 3 1 2	0
4 3 2 1	0

TABELLA 1. Punti fissi nelle permutazioni su 4 elementi

di un segretario distratto che, preparate n lettere aventi ognuna un differente destinatario e preparate le buste con gli indirizzi, infila noncurante le lettere senza badare che gli indirizzi coincidano. Qualè in media il numero di lettere che raggiungono in destinatario?).

SOLUZIONE. Per fissare le idee, iniziamo da n piccolo, ad es. $n = 3$: S_3 è costituito dalle sei permutazioni

S_3	<i>punti fissi</i>
1 2 3	3
1 3 2	1
2 3 1	0
2 1 3	1
3 1 2	0
3 2 1	1

dunque $D(3) = 1/3$. Per $n = 4$ troviamo dalla Tab. 1 quanti punti fissi ci sono per tutte le permutazioni e da qui otteniamo $D(4) = 9/24 \approx 0.375$. E qui si finisce perchè già $n = 5$ riempirebbe cinque pagine. Per $n = 16$ poi avremmo circa 1.3×10^{12} permutazioni, dunque bisogna trovare una tecnica di calcolo sintetica, oppure risolversi ad effettuare un campionamento. Utilizzando

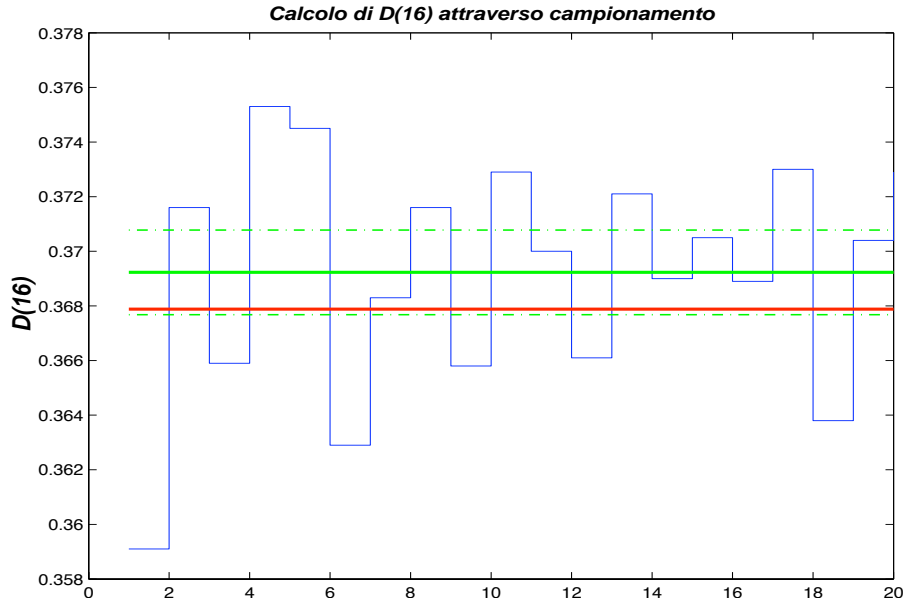


FIGURA 2.17. Simulazione del problema dei cappelli.

la funzione `randperm` (vedi l'Eserc.1, Append. B) è molto facile farsi un'idea di quanto può valere $D(16)$. Secondo la simulazione (20 pacchetti di 10000 permutazioni scelte a caso, tempo richiesto circa 10") si ha che $D(16) \approx .369 \pm 0.004$, compatibile con il risultato $\exp(-1) \approx .3679$.

Vediamo cosa ci può dire la formula di inclusione/esclusione. Partiamo dal fatto che per ogni i ($1 \leq i \leq n$) l'insieme delle permutazioni che lascia fisso i è un gruppo di permutazioni su $n-1$ elementi, che indichiamo con A_i . Inoltre $A_i \cap A_j$ è l'insieme di permutazioni che lascia due elementi fissi, e così via. Il numero di permutazioni in cui non ci sono elementi fissi sarà perciò ottenuto sottraendo dal totale $n!$ il numero dell'insieme $A_1 \cup A_2 \dots A_n$ che ci viene dato dalla formula inc/excl. Avremo perciò (indicando con $\#(A)$ il numero di elementi di A)

$$\begin{aligned} n! D(n) &= n! - \sum_i \#(A_i) + \sum_{i>j} \#(A_i \cap A_j) - \sum_{i>j>k} \#(A_i \cap A_j \cap A_k) + \dots \\ &= n! - \binom{n}{1} (n-1)! + \binom{n}{2} (n-2)! - \dots + (-1)^k \binom{n}{k} (n-k)! + \dots + (-1)^n \\ &= \sum_{k=0}^n (-1)^k \frac{n!}{k!(n-k)!} (n-k)! \end{aligned}$$

per cui

$$D(n) = \sum_{k=0}^n (-1)^k \frac{1}{k!}.$$

Per n grande si ha perciò $D(n) \rightarrow \exp(-1)$. Avendo a disposizione il software **Mathematica**, la probabilità è data da $D[n] = \text{Subfactorial}[n]/n!$, funzione che si trova anche in **Matlab**; si veda anche il codice **Matlab** su **Dropbox**.

Possiamo dire qualcosa di più riguardo al numero di permutazioni aventi un numero r ($0 \leq r \leq n$) di punti fissi. Se teniamo fissi i numeri $1, 2, \dots, r$ il numero è dato dalle rimanenti permutazioni senza punti fissi su $n-r$ elementi: $(n-r)! D(n-r)$. Tuttavia possiamo scegliere

	1	2	3	4	5	6
1	0	1	2	3	4	5
2	1	0	1	2	3	4
3	2	1	0	1	2	3
4	3	2	1	0	1	2
5	4	3	2	1	0	1
6	5	4	3	2	1	0

TABELLA 2. Vincita nel gioco dei due dadi dell'Es.36.

gli r elementi in $\binom{n}{r}$ modi e quindi in totale otteniamo

$$\binom{n}{r}(n-r)! D(n-r) = n! D(n-r)/r!$$

e cioè la probabilità di estrarre a caso una permutazione con esattamente r punti fissi è data da

$$D(n, r) = D(n-r)/r!.$$

Si verifica per consistenza che $\sum_r D(r, n) = 1$. Con questa formula si ritrovano le cifre della tabella 1. Il modulo matlab `permfp.m` permette di calcolare il numero di **permutazioni** con dato numero di **fixed points**. Si veda [Mos65] per ulteriori dettagli (Probl.N.46). ◀

Riprendiamo l'esempio del lancio dei due dadi considerato alla sezione 2.4.1: abbiamo verificato A , B e C non sono mutuamente indipendenti. Chiediamoci qual'è la probabilità $\mathbb{P}(A \cup B \cup C)$. La formula di incl/escl ci dà subito il valore $1/2 + 1/2 + 1/2 - 1/4 - 1/4 - 1/4 = 3/4$. Corretto, infatti i casi possibili sono $(d, *, *)$, $(*, d, *)$, $(*, *, d)$, indicando con $*$ un risultato qualunque. Ma $(d, *, *)$ contiene anche i casi (d, d, p) e (d, p, d) che sono già conteggiati in $(*, d, *)$ e $(*, *, d)$ e dunque sono da sottrarre $(-1/4)$; analogamente per gli altri due casi che portano in totale a sottrarre $3/4$. Invece (d, d, d) non può verificarsi.

Consideriamo lo stesso problema con $A' = (p, *, *)$, $B' = (*, p, *)$ e $C' = (*, *, p)$. Si avrà

$$\mathbb{P}(A' \cup B' \cup C') = 1/2 + 1/2 + 1/2 - 1/4 - 1/4 - 1/4 + 1/4 = 1$$

l'ultimo contributo dato da $\mathbb{P}(A' \cap B' \cap C') = 1/4$. Il risultato si capisce dal fatto che i casi che non rientrano in $A' \cup B'$ sono dati da $\overline{A' \cup B'} = \overline{A'} \cap \overline{B'}$ cioè i due dadi hanno entrambi valori dispari, ma questi casi rientrano in C' e quindi si ottiene la totalità dei casi.

ESERCIZIO 36. In un recente numero de “*Le Scienze*” si propone questo problema. A un tavolo di Casinò una volta versata la posta di 2 Euri si lanciano due dadi e si ottiene in ritorno la *differenza*, in valore assoluto, tra i due valori che sono usciti. Dimostrare che per rendere il gioco equo si deve introdurre la regola che nel caso dell'uscita di due “6” si può tirare di nuovo la coppia di dadi senza aggiungere un'altra posta.

SOLUZIONE. Considerando la matrice della Tab.2 si trova che si ha una vincita di 5 in due casi, di 4 in quattro casi, di 3 in sei casi, etc. Dunque il ritorno medio è dato da $\frac{1}{36} \sum_{j,k=1}^6 |j-k| = 2 \sum_{j=1}^6 j(6-j) = \frac{1}{36} \frac{N(N^2-1)}{3}$ (N è il numero di facce). Per $N = 6$ si ha perciò un ritorno medio $70/36$; mancano dunque $2/36$ affinché il gioco sia equo. Questo fatto è corretto modificando la regola in modo che per l'uscita di due 6 si ritornano i due Euri. La formula valida per N qualunque permette di affrontare il caso di due dadi a forma di dodecaedro...

2.5. Variabili aleatorie

Dato uno spazio di probabilità $(\Omega, \Sigma, \mathbb{P})$ si definisce *variabile aleatoria* (random variable) una funzione a valori reali

$$(11) \quad X: \Omega \rightarrow \mathcal{R}$$

Le variabili aleatorie, nel loro complesso, costituiscono un sistema di coordinate che permettono di individuare gli eventi di Ω in modo simile alle coordinate che si utilizzano in geometria. Ad es., l'evento rappresentato dall'estrazione di una data carta può essere messo in corrispondenza con una variabile aleatoria a valori interi nell'intervallo $[1 - 32]$ e il \heartsuit corrisponde al valore 7. Avendo valori reali, le variabili aleatorie ammettono tutta una serie di operazioni aritmetiche che hanno un significato probabilistico/statistico. Vediamo alcune di queste operazioni.

Se X è una v.a., si definisce il valore di aspettazione (o valore medio)

$$(12) \quad \mathbb{E}[X] \equiv \langle X \rangle = \sum_{\omega \in \Omega} X(\omega) \mathbb{P}(\omega).$$

Inoltre, come è familiare dalla statistica, si definisce la *varianza*

$$(13) \quad \text{Var}[X] \equiv \mathbb{D}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$$

mentre la quantità $\text{std}[X] = \sqrt{\mathbb{D}[X]}$ è detta “deviazione standard”. Questa definizione si applica nel caso di spazi di probabilità discreti, in cui la probabilità è definita per gli eventi elementari. Nel caso continuo è necessario richiedere che la variabile aleatoria X soddisfi una condizione di compatibilità rispetto alla famiglia di eventi misurabili e alla funzione di probabilità \mathbb{P} . Dovremo richiedere cioè che gli insiemi $\sigma(x) \equiv \{\omega \in \Omega | X(\omega) < x\}$ appartengano all'algebra Σ degli insiemi misurabili. Questo permette di definire la *funzione di distribuzione* come sarà introdotta nella sezione successiva.

2.5.1. Funzione di distribuzione. Per ogni variabile aleatoria si definisce la *funzione di distribuzione* (f.d.d) attraverso la formula

$$(14) \quad \Phi(x) = \mathbb{P}(X < x).$$

La f.d.d. è non-decrescente. Vediamo un primo esempio:

La distribuzione di **Bernoulli** è la più semplice possibile. Lo spazio degli eventi ha solo due elementi ω_1 e ω_2 con $\mathbb{P}(\omega_1) = q, \mathbb{P}(\omega_2) = p$; p e q sono non negativi e a somma uno. Un secondo esempio è dato dalla distribuzione binomiale che illustreremo più avanti quando tratteremo della *somma di variabili aleatorie*.

Una variabile aleatoria si può definire nel modo più semplice con $X(\omega_1) = 0, X(\omega_2) = 1$. La funzione di distribuzione di X è allora

$$\Phi(x) = \begin{cases} 0 & (x < 0) \\ q & (0 \leq x < 1) \\ 1 & (1 \leq x) \end{cases}$$

Nel caso discreto la f.d.d. è rappresentata da una *scalinata* che cresce da 0 a 1 con salti di discontinuità in corrispondenza ai valori possibili della v.a.

Se la v.a. X assume valori x_1, x_2, \dots, x_n con probabilità p_1, p_2, \dots, p_n il diagramma della sua f.d.d. incrementa di p_k quando il valore della x passa attraverso x_k . Il diagramma rappresenta perciò la somma cumulativa dei valori p_k , quello che in `matlab` si esprime con `cumsum()`.

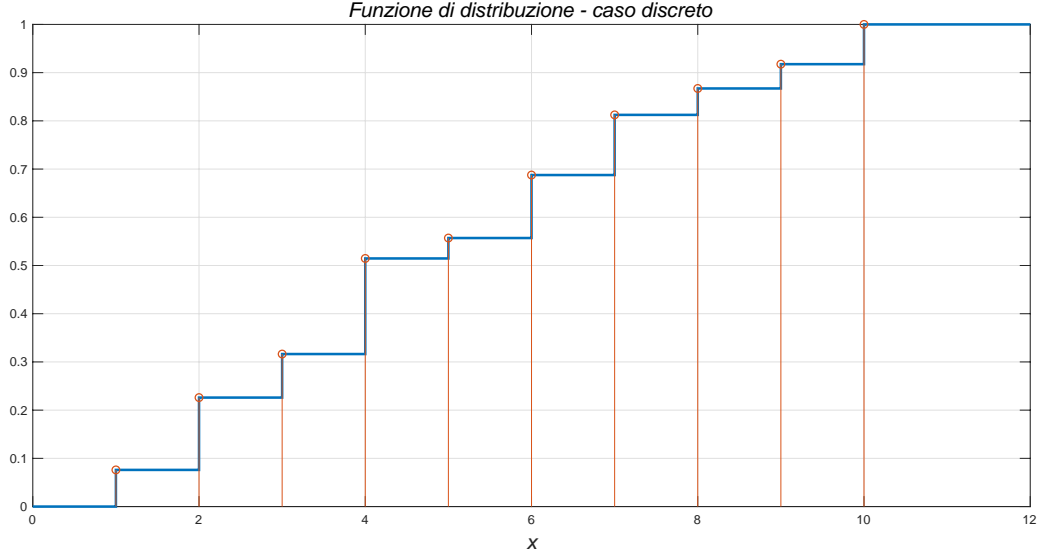


FIGURA 2.18. Esempio di funzione di distribuzione nel caso discreto.

Nel caso di distribuzioni di probabilità continue la f.d.d. è continua ed è esprimibile come $\Phi(x) = \int_{-\infty}^x \rho(\xi) d\xi$, e $\rho(\cdot)$ è detta *densità di probabilità*.

Il valore medio è in ogni caso dato dalla espressione

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x d\Phi(x)$$

Nel caso continuo si ha $d\Phi(x) = \rho(x)dx$ (è la densità ρ ad es. che è collegata in meccanica ondulatoria alla funzione d'onda ψ attraverso la relazione di Born $\rho = |\psi|^2$).

Nel caso discreto la Φ non è differenziabile ma l'integrale è definibile rigorosamente come *integrale di Stieltjes*. Il valore dell'integrale si definisce attraverso la formula di integrazione per parti:

$$\begin{aligned} \int_{-\infty}^{\infty} x d\Phi(x) &= \sum_k \int_{x_k - \varepsilon}^{x_k + \varepsilon} x d\Phi(x) \\ &= \sum_k ((x_k + \varepsilon) \Phi(x_k + \varepsilon) - (x_k - \varepsilon) \Phi(x_k - \varepsilon)) \\ &= \sum_k x_k p_k \end{aligned}$$

dove nell'ultimo passaggio si è tenuto conto che nel limite $\varepsilon \rightarrow 0$ la quantità che sopravvive è la discontinuità della Φ in ciascun punto x_k . Utilizzando la “funzione” $\delta(x)$ di Dirac si può scrivere

$$d\Phi(x) = \sum_k p_k \delta(x - x_k) dx.$$

La *funzione di distribuzione congiunta* per due variabili aleatorie si definisce attraverso la relazione

$$(15) \quad \Phi(x, y) = \mathbb{P}(\{X < x\} \cap \{Y < y\})$$

e in generale questa definizione si estende in modo naturale a un numero qualunque di variabili aleatorie. Si vedano altri esempi in Appendice.

2.5.2. Variabili (statisticamente) indipendenti. Il concetto di *indipendenza statistica* si estende in modo naturale alle variabili aleatorie. Siano X e Y variabili aleatorie definite nello stesso spazio di probabilità; si dirà che esse sono statisticamente indipendenti se ciò è vero per tutti i sottoinsiemi $\{X < x\}$ e $\{Y < y\}$. In concreto ciò equivale a dire che la f.d.d. congiunta è fattorizzata:

$$\Phi(x, y) = \Phi_X(x) \Phi_Y(y)$$

Un indicatore importante per due variabili aleatorie è la loro *correlazione* definita attraverso

$$(16) \quad C(X, Y) = E[XY] - E[X]E[Y]$$

Se X e Y sono indipendenti si ha ovviamente $C(X, Y) = 0$.

ESEMPIO 5. Due variabili gaussiane sono definite del tutto in generale dalla f.d.d.

$$\Phi(x, y) = \int_{-\infty}^x \int_{-\infty}^y \mathcal{N} \exp\{-\alpha\xi^2 - \beta\eta^2 - 2\gamma\xi\eta\} d\xi d\eta$$

dove la costante \mathcal{N} è fissata in modo che $\Phi \rightarrow 1$ per $x \rightarrow \infty, y \rightarrow \infty$. (Si ha $\mathcal{N} = \sqrt{\alpha\beta - \gamma^2}/\pi$, con la condizione che $\gamma^2 < \alpha\beta$). Si calcoli la covarianza $C(X, Y)$ e si dimostri che essa risulta proporzionale a γ .

2.5.3. Funzione generatrice. Sia X una v.a. a valori interi con probabilità $\mathbb{P}(X = i) \equiv p_i$. Consideriamo l'espressione

$$F(z) = \mathbb{E}[z^X] = \sum_i p_i z^i.$$

Si tratta di un polinomio nella variabile z che contiene in sé ovviamente tutte le informazioni su X . Ad es. si avrà

$$\mathbb{E}[X] = \sum_i i p_i = F'(1)$$

$$\mathbb{D}[X] = \mathbb{E}[(X - \langle X \rangle)^2] = F''(1) + F'(1) - F'(1)^2.$$

La funzione F viene chiamata funzione generatrice della v.a. X . Per variabili a valori continui con distribuzione di probabilità Φ definita da

$$\Phi(x) = \mathbb{P}(\{X \leq x\})$$

e densità $\rho(x) = \Phi'(x)$ si utilizza comunemente una definizione un pó differente

$$F(t) = \mathbb{E}[e^{tX}] = \int e^{tx} d\Phi(x) = \int e^{tx} \rho(x) dx.$$

dalla quale si ottiene immediatamente

$$\mathbb{E}[X^k] = F^{(k)}(0)$$

e cioè lo sviluppo in serie di Taylor della funzione generatrice $F(t)$ ha come coefficienti tutti i momenti della distribuzione (divisi per $k!$). Talvolta si preferisce considerare $F(it)$ che consente di calcolare la f.g. come una trasformata di Fourier, nel qual caso bisogna tenere conto in modo ovvio del fattore i nella formula per $\mathbb{E}[X^k]$.

ESERCIZIO 37. (*Esercizio adatto a chi apprezza argomenti matematici più raffinati della media di queste lezioni*) Consideriamo la probabilità $D(n)$ che una permutazione su n oggetti

non ne lasci alcuno fisso (sia un “derangement”). Sia $N(n)$ il numero di derangements e consideriamo la funzione generatrice

$$F[z] = \sum_{n=0}^{\infty} D(n)z^n = \sum_{n=0}^{\infty} N(n)z^n/n!.$$

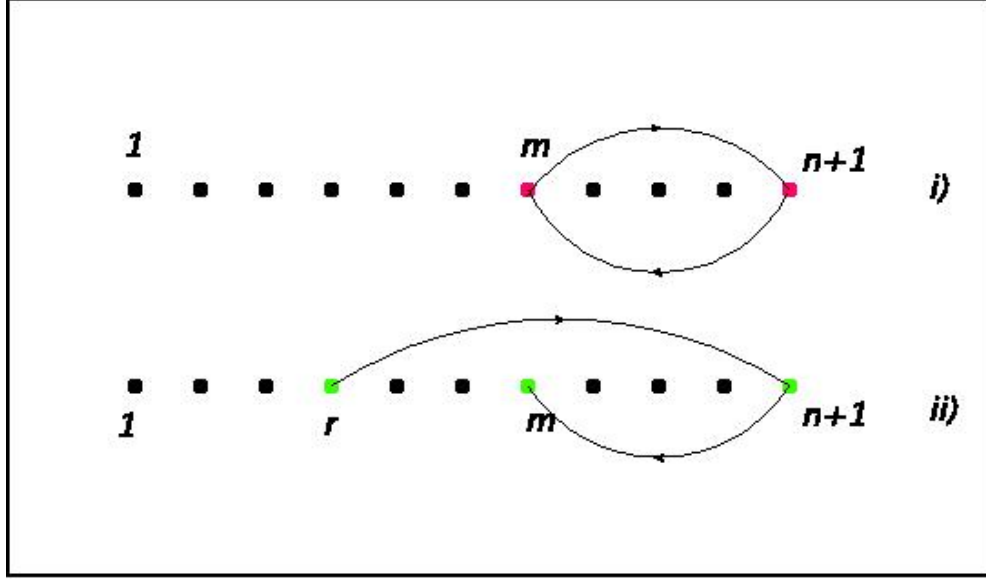


FIGURA 2.19. I due casi possibili $n+1 \rightarrow m \rightarrow n+1$ e $r \rightarrow n+1 \rightarrow m$, ($r \neq m$) che portano alla relazione di ricorrenza.

Un semplice argomento porta alla relazione di ricorrenza $N(n+1) = n(N(n) + N(n-1))$. Infatti data una permutazione su $n+1$ oggetti indicata con π che non lasci alcun punto fisso si danno due casi:

- (i) $\pi(n+1) = m$, $\pi(m) = n+1$ per qualche $m \leq n$;
- (ii) $\pi(n+1) = m$, e sia $r \neq m$ tale che $\pi(r) = n+1$.

Nel primo caso m e $n+1$ formano un ciclo e quello che rimane di π è un derangement su $n-1$ oggetti. Nel secondo caso collassando la sequenza $\pi: r \rightarrow n+1 \rightarrow m$ eliminando $n+1$ si rimane con un derangement su n oggetti. Dato che m può essere scelto in n modi si ha allora che

$$N(n+1) = n(N(n) + N(n-1)).$$

Calcoliamo $F'(z) = \sum n N(n) z^{n-1}/n! \equiv \sum N(n+1) z^n/n!$. Sostituendo la relazione di ricorrenza otteniamo

$$F'(z) = \sum n(N(n) + N(n-1)) z^n/n! = zF'(z) + zF(z)$$

ossia abbiamo un'equazione differenziale

$$(1-z)F'(z) = zF(z)$$

che ha come soluzione $F(z) = \frac{e^{-z}}{1-z}$, avendo tenuto conto della normalizzazione $F(0) = 1$. Sviluppando $F(z)$ in serie di Taylor otteniamo le probabilità $D(n)$ come coefficiente dello sviluppo in

serie:

$$F(z) = \sum_0^{\infty} D(n) z^n = \sum \frac{(-z)^k}{k!} \times \sum z^m = \sum \left(\sum_{k=1}^n (-)^k / k! \right) z^n.$$

ossia

$$D(n) = \sum_{k=1}^n (-)^k / k!$$

che per n molto grande converge a $\exp\{-1\} \approx 0.368$ un risultato che abbiamo già ottenuto applicando la *formula di inclusione/esclusione*. ◀

Gauss

ESERCIZIO 38. La distribuzione *normale* o *Gaussiana* è data dalla densità di probabilità

$$\rho(x) = \mathcal{N} \exp \left\{ -\frac{x^2}{2\sigma} \right\}$$

La costante di normalizzazione vale $\mathcal{N} = (2\pi\sigma)^{-\frac{1}{2}}$. Calcolare i momenti della distribuzione, cioè

$$m_k = \mathbb{E}[X^k] = \mathcal{N} \int_{-\infty}^{\infty} \exp \left\{ -\frac{x^2}{2\sigma} \right\} x^k dx$$

SOLUZIONE. A questo scopo conviene calcolare la funzione generatrice

$$\Phi(y) = \mathbb{E}[e^{yX}] = \mathcal{N} \int_{-\infty}^{\infty} \exp \left\{ yx - \frac{x^2}{2\sigma} \right\} dx$$

Un cambiamento di variabile $x' = x + a$ nell'integrale scegliendo a in modo che all'esponente si cancelli il termine lineare in x' permette di concludere

$$\Phi(y) = \exp \left\{ \frac{1}{2} \sigma y^2 \right\}$$

Dallo sviluppo in serie di Taylor di Φ otteniamo allora che il coefficiente di y^{2n} che per costruzione coincide con $\mathbb{E}[X^{2n}]/(2n)!$ è dato da $(\frac{1}{2}\sigma)^n/n!$ o in definitiva

$$\mathbb{E}[X^{2n}] = \frac{(2n)!}{2^n n!} \sigma^n.$$

Il valor medio di una potenza dispari di X è zero per simmetria e infatti $\Phi(y)$ ha coefficienti nulli per potenze dispari di y . ◀

Gaussiana n-dim

ESERCIZIO 39. Una distribuzione Gaussiana in n variabili è definita da una densità di probabilità

$$\rho(x_1, x_2, \dots, x_n) = \mathcal{N} \exp \left\{ -\frac{1}{2} \sum_{i,j} x_i (A^{-1})_{ij} x_j \right\}$$

dove la matrice $A = \|a_{ij}\|$ è *positiva definita*, cioè è reale-simmetrica e con autovalori tutti positivi. Calcolare la costante di normalizzazione \mathcal{N} e i valori medi $\mathbb{E}[x_i x_j]$.

SOLUZIONE. Dal momento che \mathbf{A} è simmetrica si può introdurre un sistema di riferimento ortogonale $\mathbf{x}'_i = \mathbf{R}_{ij} \mathbf{x}_j$ in cui \mathbf{A} risulta diagonale

$$\mathbf{A}' = \mathbf{R} \mathbf{A} \mathbf{R}^{-1} = \begin{pmatrix} \mathbf{a}_1 & 0 & \dots & 0 \\ 0 & \mathbf{a}_2 & \dots & 0 \\ 0 & \dots & \ddots & 0 \\ 0 & \dots & \dots & \mathbf{a}_n \end{pmatrix}$$

Ne segue che utilizzando le variabili \mathbf{x}'_i l'integrale di normalizzazione diventa

$$\int \dots \int d\mathbf{x}'_1 \dots d\mathbf{x}'_n \exp\{-\frac{1}{2} \sum \mathbf{x}'_j{}^2 / \mathbf{a}_j\}$$

e si può dunque valutare facilmente come il prodotto di n integrali in una sola variabile; risulta allora $\mathcal{N} = \Pi_j (2\pi \mathbf{a}_j)^{-1/2}$ e più sinteticamente

$$\mathcal{N} = (2\pi)^{-n/2} (\det \mathbf{A})^{-\frac{1}{2}}$$

Per calcolare i valori medi $\mathbb{E}[\mathbf{x}_i \mathbf{x}_j]$ converrà definire la funzione generatrice

$$\Phi(\mathbf{y}) = \mathbb{E}[e^{\sum \mathbf{y}_i \mathbf{x}_i}]$$

che ci permette di calcolare

$$\mathbb{E}[\mathbf{x}_i \mathbf{x}_j] = \frac{\partial^2}{\partial \mathbf{y}_i \partial \mathbf{y}_j} \Phi(\mathbf{y}) \Big|_{\mathbf{y}=0}$$

Il calcolo di Φ non è difficile, perché attraverso il cambiamento di base $\mathbf{x} \rightarrow \mathbf{x}'$ si riduce a n integrali in una sola variabile:

$$\begin{aligned} \Phi(\mathbf{y}) &= \mathcal{N} \int \dots \int \exp\{\sum \mathbf{y}_i \mathbf{x}_i - \frac{1}{2} \sum_{i,j} \mathbf{x}_i (\mathbf{A}^{-1})_{ij} \mathbf{x}_j\} \prod d\mathbf{x}'_j \\ &= \mathcal{N} \int \dots \int \exp\{\sum \mathbf{y}'_i \mathbf{x}'_i - \frac{1}{2} \sum_i \mathbf{x}'_i{}^2 / \mathbf{a}_i\} = \exp\{\sum_i \mathbf{a}_i \mathbf{y}'_i{}^2 / 2\} \\ &= \exp\{\frac{1}{2} \sum_{i,j} \mathbf{y}_i \mathbf{a}_{ij} \mathbf{y}_j\} \end{aligned}$$

Si può allora calcolare facilmente la correlazione

$$\mathbb{E}[\mathbf{x}_i \mathbf{x}_j] = \frac{\partial^2}{\partial \mathbf{y}_i \partial \mathbf{y}_j} \Phi(\mathbf{y}) \Big|_{\mathbf{y}=0} = \mathbf{a}_{ij} .$$

Infatti, ad es. per $i = 1, j = 2$

$$\mathbb{E}[\mathbf{x}_1 \mathbf{x}_2] = \frac{\partial^2}{\partial \mathbf{y}_1 \partial \mathbf{y}_2} \exp\{\mathbf{a}_{12} \mathbf{y}_1 \mathbf{y}_2 + \mathbf{Rem}\} \Big|_{\mathbf{y}=0} = \mathbf{a}_{12}$$

dove **Rem** indica tutti gli altri termini, che non contengono il prodotto $\mathbf{y}_1 \mathbf{y}_2$ e che quindi non contribuiscono al risultato. ◀

ESERCIZIO 40. Calcoliamo la funzione generatrice della distribuzione di probabilità per il problema del botteghino (dall'Es.16 a pag.30): ci sono $\binom{2n}{n}$ cammini possibili lunghi $2n$ in cui ogni passo può essere ± 1 e che raggiungono l'origine. Ciò è evidente considerando che possiamo modellizzare i cammini sulla funzione $(\mathbf{x} + \mathbf{x}^{-1})^{2n}$: il termine costante della funzione rappresenta tutti i contributi allo sviluppo del polinomio in cui \mathbf{x} e \mathbf{x}^{-1} entrano con potenza n , quindi con

coefficiente $\binom{2n}{n}$ e questo ci dà la funzione generatrice (per il numero totale di cammini che raggiungono 0 dopo $2n$ passi)

$$F(z) = \sum_1^{\infty} \binom{2n}{n} z^n$$

Utilizzando proprietà ³² della funzione $\Gamma[z]$ si può trasformare il coefficiente binomiale e ottenere

$$F(z) = \sum_{n>0} \binom{-\frac{1}{2}}{n} (-4z)^n = (1 - 4z)^{-\frac{1}{2}} - 1$$

Ora, come si era spiegato in precedenza, noi siamo interessati alla probabilità che il cammino *incontri l'origine per la prima volta* dopo $2n$ passi. Per una proprietà generale la funzione generatrice di questa probabilità è data da $F_1(z) = F(z)/(1 + F(z))$; questo fatto sarà dimostrato nell'Es.(3.2.2) a pag.85. Intanto sfruttiamo questo risultato per ottenere

$$F_1(z) = 1 - \sqrt{1 - 4z} = - \sum_{n>0} \binom{1/2}{n} (-4z)^n$$

Per ottenere la probabilità di primo arrivo nell'origine dopo $2n$ passi dobbiamo dividere per il numero totale di cammini, ossia

$$(17) \quad \text{Prob}(n) = \frac{F_1^{(n+1)}}{\binom{2n}{n}} = - \frac{\binom{1/2}{n+1} (-4)^{n+1}}{\binom{2n}{n}}$$

dove abbiamo indicato con $F_1^{(n)}$ l' n -esimo coefficiente di Taylor di $F_1(z)$. Tuttavia possiamo già concludere che il risultato coincide con quanto già trovato in quanto coincidono le funzioni generatrici ³³ ottenute qui e quella nell'Es.16 a pag.30. Il calcolo del risultato dell'Eq.(17) può essere noioso, ma l'utilizzo di Mathematica (si può utilizzare il modulo `botteghino.nb`) ci solleva dalla fatica e riscopriamo ³⁴ $\text{Prob}(n) = 2/(n+1)$.

2.5.4. Somma di variabili aleatorie. Le v.a. si possono manipolare come le funzioni reali dell'analisi matematica. Molto spesso si trova necessario *sommare* v.a.: ad es. il risultato del lancio di due dadi coincide con la somma di due v.a. D_1 e D_2 : siamo abituati a considerare le facce del dado identificate con il numero corrispondente, ma concettualmente il numero è una variabile aleatoria che identifica la faccia corrispondente, così come la coordinata cartesiana x non è il punto geometrico ma misura la sua distanza da un punto fisso. Ora la somma $D_1 + D_2$ assume valori nell'insieme $2, 3, \dots, 12$. La sua distribuzione di probabilità si determina facilmente (come?):

$$(18) \quad p^{(2)}(n) = \begin{pmatrix} 1 & 0 \\ 2 & 1/36 \\ 3 & 2/36 \\ 4 & 3/36 \\ 5 & 4/36 \\ \vdots & \\ 11 & 2/36 \\ 12 & 1/36 \end{pmatrix}$$

³² $\Gamma(2z) = (2\pi)^{-1/2} 2^{2z-1/2} \Gamma(z) \Gamma(z + \frac{1}{2})$, $\Gamma(z) \Gamma(1-z) = \pi / \sin \pi z$.

³³Non ci si lasci ingannare dal fattore x^{-1} , quello serve solo a spostare l'indice di uno.

³⁴Rimane da giustificare un fattore 2: il computo infatti non tiene conto che per ogni cammino che raggiunge per la prima volta l'origine in un dato numero di passi ne esiste uno simmetrico a valori negativi.

In generale la distribuzione di probabilità per la somma $X + Y$ è definita da

$$\mathbb{P}(\{X + Y = n\}) = \sum_j \mathbb{P}((X = j) \& (Y = n - j)) .$$

Per v.a. indipendenti la formula si può calcolare in modo esplicito dato che la distribuzione congiunta si fattorizza. Consideriamo il caso discreto, come nel caso del lancio dei dadi. Sia $p(n) = \mathbb{P}(D_i = n) = 1/6$; si ha ovviamente

$$\mathbb{P}(\{D_1 + D_2 = n\}) = \sum_{i=1}^6 p(i) \sum_{j=1}^6 p(j) \delta_{i+j=n} = \sum_{i=1}^6 p(i)p(n-i) = (p \star p)(n)$$

ossia la funzione di probabilità della somma di v.a. indipendenti con la stessa distribuzione di probabilità è data dalla *convoluzione* delle rispettive funzioni di probabilità. Il risultato vale anche per variabili a valori interi indipendenti con diversa distribuzione di probabilità, come è facile controllare. Ad es. il risultato della somma delle uscite di tre dadi ha una d.d.p. che si può calcolare con la convoluzione in matlab in modo immediato³⁵:

```
>> p = ones(1,6)/6;
>> p2 = conv(p,p) = [1 2 3 4 5 6 5 4 3 2 1]/36
>> p3 = conv(p2,p) = [1 3 6 10 15 21 25 27 27 25 21 15 10 6 3 1]/216
```

e così abbiamo calcolato la d.d.p. dell'Eq.(18) e ritroviamo i risultati già noti di De Mère dell'esercizio 10 a pag.26.

Possiamo ora apprezzare l'utilità della definizione di funzione generatrice. Si trova che la f.g. per la somma $D_1 + D_2$ è data da

$$F_{12}(z) = \sum_{n=0}^{\infty} \sum_{i=1}^6 p(i) \sum_{j=1}^6 p(j) \delta_{i+j=n} z^n = \sum_{i,j} p_i p_j z^{i+j} = F(z)^2;$$

ossia la f.g. per la somma di v.a. indipendenti con la medesima distribuzione di probabilità è il quadrato della funzione generatrice di ognuna. In generale per v.a. indipendenti la f.g. si ottiene moltiplicando le funzioni generatrici degli addendi. Ciò costituisce una bella facilitazione! Il risultato dell'equazione (18) è ottenuto giusto in questo modo. Si ricorderà il risultato dell'Es.23; senza saperlo abbiamo utilizzato il risultato relativo alla funzione generatrice! Il misterioso polinomio $P_{n,f}(x) = \left(\sum_{k=1}^f x^k\right)^n$ altro non è che la funzione generatrice per la somma dei risultati di n dadi a f facce.

ESERCIZIO 41. Distribuzione binomiale. Considerare le v.a. X_i di Bernoulli tra loro statisticamente indipendenti. Ogni X_i ha una f.g. data da

$$F(z) = q + pz .$$

Calcolare la f.g. per la somma di n variabili di Bernoulli indipendenti, e dedurne il valor medio e la varianza (distribuzione binomiale)

SOLUZIONE. Si ha immediatamente $F_n(z) = (q + pz)^n$ e dunque

$$\mathbb{E}[\sum X_i] = F'_n(1) = np$$

$$\mathbb{D}[\sum X_i] = F''_n(1) + F'_n(1) - (F'_n(1))^2 = n(n-1)p^2 + np - (np)^2 = nqp$$

³⁵Se si desidera lavorare con numeri razionali conviene omettere la normalizzazione $\frac{1}{6}$ nella definizione di p e aggiungere un denominatore $\frac{1}{36}$ e $\frac{1}{216}$ dopo avere calcolato le convoluzioni.

Inoltre sviluppando la funzione generatrice si ottiene immediatamente che la probabilità di k “successi” è data da

$$P_{n,k}(p, q) = \binom{n}{k} p^k q^{n-k}$$

nota come distribuzione binomiale. Analogamente per il caso in cui abbiamo la somma di n variabili aleatorie indipendenti X_i caratterizzate da valori interi nell'intervallo $[0, f]$ con probabilità p_0, p_1, \dots, p_f . La funzione generatrice per la somma $\sum X_i$ è semplicemente $F(z) = (\sum_j p_j z^j)^n$, come abbiamo già considerato nella Sez.23. Si parla, per estensione della terminologia, di distribuzione multinomiale. ◀

2.5.5. La disuguaglianza di Chebyshev. Un teorema del tutto generale caratterizza meglio il significato della varianza di una variabile aleatoria.

TEOREMA 2.2. Sia ξ una variabile aleatoria con media 0 e varianza $\mathbb{E}[\xi^2] = \sigma$. Allora la probabilità che ξ assuma valori che superano in valore assoluto n volte la deviazione standard $\sqrt{\sigma}$ rispetta la *disuguaglianza di Chebyshev*

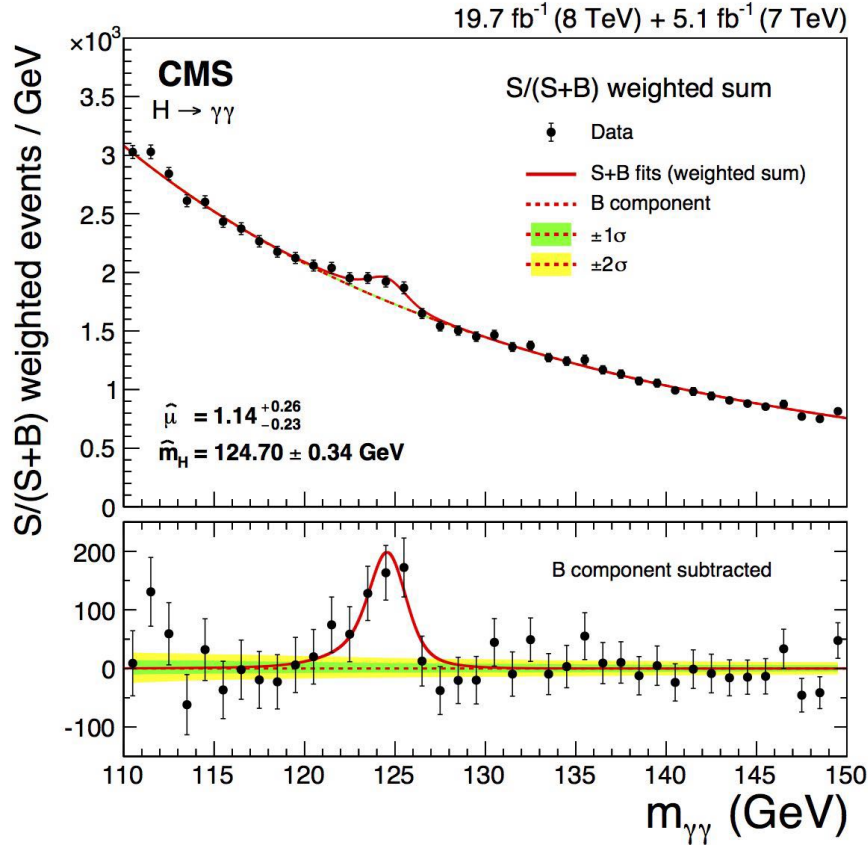
$$(19) \quad \mathbb{P}[|\xi| \geq n\sqrt{\sigma}] \leq 1/n^2$$

Per la dimostrazione, abbastanza semplice, si veda [Roz69], Cap.4.9. ◀

La disuguaglianza non è molto stringente, ad es. per $n = 1$ è del tutto vuota; ciò è abbastanza comprensibile in quanto la disuguaglianza si applica *qualunque sia la distribuzione di probabilità*. Ovviamente per una data distribuzione si potrà rinforzare la disuguaglianza. Ad es. per la distribuzione normale ($\sigma = 1/2$) si ha

$$(20) \quad \mathbb{P}[\xi \geq \frac{n}{\sqrt{2}}] = \mathbb{P}[\xi \leq -\frac{n}{\sqrt{2}}] \equiv P[n] = \frac{1}{2} (1 - \text{erf}(n/\sqrt{2}))$$

essendo $\text{erf}(z)$ la funzione degli errori $\text{erf}(z) = 2/\sqrt{\pi} \int_0^z e^{-x^2} dx$. Numericamente si trovano per $n = 1, 2, \dots, 5$ i valori $P[n] = \{0.16, 0.023, 1.35 \times 10^{-3}, 3.17 \times 10^{-5}, 2.87 \times 10^{-7}\}$. Sulla base di questi valori, si comprende come una deviazione di un segnale per più di 3σ da quanto atteso possa essere considerato un segnale significativo, cioè non attribuibile a semplice fluttuazione statistica. A titolo di esempio si veda la Fig. 2.20 (fonte CERN/CMS). Si nota che i dati sono soggetti a fluttuazioni di varia intensità ma intorno all'energia di 125GeV si riscontra un picco che si scosta sistematicamente per più di tre σ dal “fondo”. È la firma della particella di Higgs. [I Fisici sono peraltro molto cauti prima di annunciare una scoperta anche se il segnale provvisorio sembra superare il requisito dei “ 3σ ”, come il recente caso del segnale attorno a 750GeV ha mostrato chiaramente (Physics World, 5 agosto 2016). Per un approfondimento del problema della significatività di una misura nel campo della fisica delle alte energie si può consultare la Ref.[D'A12].

FIGURA 2.20. Segnale nel canale $\gamma\gamma$ nel decadimento della particella di Higgs.

2.6. La legge dei grandi numeri

Legge dei grandi

Do not worry about your difficulties in Mathematics. I can assure you mine are still greater

Albert Einstein

2.6.1. Il Teorema di De Moivre-Laplace. Sotto la denominazione “legge dei grandi numeri” sono compresi numerosi risultati che riguardano valori asintotici di distribuzioni di probabilità quando il numero di gradi di libertà è molto grande. L’idea generale è quella della meccanica statistica; un sistema costituito da un grande numero di particelle in debole interazione può essere studiato facilmente sfruttando le semplificazioni che si presentano nel limite di infinite particelle. Per variabili aleatorie un caso importante è quello della somma di un grande numero di v.a. indipendenti e tutte con la medesima distribuzione di probabilità. L’esempio più semplice è quello della somma di variabili di Bernoulli $X = \sum X_i$. La sua funzione generatrice è semplicemente $(q + pz)^n$. Il valore medio è dato da np quindi conviene considerare la v.a. centrata $X - \langle X \rangle$, a media zero. Tuttavia non è ancora possibile prendere esplicitamente il limite $n \rightarrow \infty$ in quanto la deviazione standard vale nqp e cresce indefinitamente. Si può allora considerare la v.a. normalizzata

$$(21) \quad \eta = \frac{X - \langle X \rangle}{\sqrt{\Delta X}} \equiv \frac{\sum_{i=1}^n X_i - n p}{\sqrt{n p q}}$$

che ha perciò media zero e varianza uno. Calcoliamo la sua funzione generatrice nella forma adatta per una distribuzione continua

$$(22) \quad F_\eta(t) = \mathbb{E}[e^{it\eta}] = e^{-i \frac{npt}{\sqrt{npq}}} \mathbb{E} \left[\exp \left\{ it \frac{X}{\sqrt{npq}} \right\} \right].$$

Adesso però possiamo utilizzare la formula che riduce il calcolo alla potenza n -esima della f.g. per una singola v.a.³⁶:

$$(23) \quad F_\eta(t) = \exp \left\{ -i \frac{npt}{\sqrt{npq}} \right\} \left(q + p \exp \left\{ i \frac{t}{\sqrt{npq}} \right\} \right)^n$$

$$(24) \quad = \left(q \exp \left\{ -i \sqrt{\frac{p}{qn}} t \right\} + p \exp \left\{ i \sqrt{\frac{q}{pn}} t \right\} \right)^n$$

Osserviamo ora che, nel limite $n \rightarrow \infty$, la funzione F_η ammette un semplice sviluppo asintotico che si ottiene sviluppando le funzioni esponenziali:

$$(25) \quad F_\eta = \left(1 - \frac{t^2}{2n} + O(n^{-3/2}) \right)^n \underset{n \rightarrow \infty}{\approx} e^{-\frac{1}{2}t^2}$$

cioè al limite $n \rightarrow \infty$ la f.g. si riduce a una semplice Gaussiana e perciò la distribuzione di probabilità quella di Gauss con media zero e varianza uno (distribuzione normale). Si noti che nel limite *scompare la dipendenza dai parametri p e q* della distribuzione di Bernoulli, un chiaro segnale che nel limite si perde la memoria della distribuzione delle singole variabili aleatorie e che il risultato può essere ben più generale.

Il risultato, noto come teorema di De Moivre-Laplace, rappresenta la formulazione più elementare della legge dei grandi numeri: la somma di n variabili aleatorie statisticamente indipendenti (21) nel limite di n molto grande converge a distribuzione normale. Il risultato è molto più generale: non dipende dalla natura dettagliata della distribuzione di probabilità della singola X_i . Il risultato vale anche, sotto opportune restrizioni, nel caso in cui la distribuzione di X_i possa dipendere da i . Si veda [Roz69]. Il programma `demoivre.m` dimostra attraverso una simulazione lo stesso risultato. Si costruisce una matrice di 0 e 1 distribuiti secondo Bernoulli e la somma sulle colonne corrisponde alla v.a. X . Si analizza l'istogramma e si può controllare di quanto differisce da una distribuzione normale in funzione delle dimensioni dell'esperimento. In Fig.2.21 le linee continue rappresentano la funzione `erf()` e la densità Gaussiana. I risultati della simulazione sono rappresentati dai due istogrammi.

ESERCIZIO 42. Calcolare la distribuzione di probabilità della somma $Z = X + Y$ di due variabili gaussiane indipendenti X e Y con media 0 e varianza σ_X e σ_Y rispettivamente.

SOLUZIONE. La funzione generatrice di X è $F_X(k) = \exp\{-\sigma_X k^2/2\}$ e per Y vale la stessa formula dunque la somma è caratterizzata dalla funzione generatrice

$$F_{X+Y}(k) = F_X(k) F_Y(k) = \exp\left\{-\frac{1}{2}(\sigma_X + \sigma_Y) k^2\right\}$$

³⁶Per questa impostazione del teorema di De Moivre-Laplace si veda [Fis63]

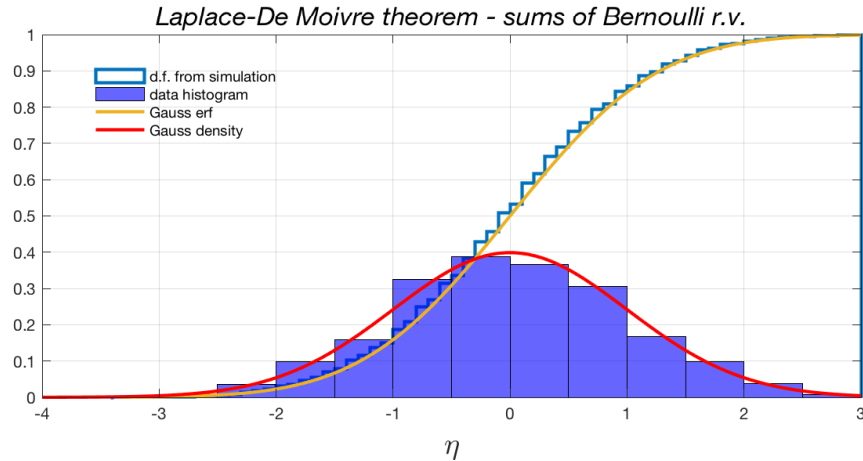


FIGURA 2.21. Funzione di distribuzione e densità per la somma di 500 variabili di Bernoulli.

il che mostra che Z è una variabile gaussiana con media 0 e varianza $\sigma_Z = \sigma_X + \sigma_Y$. In generale la somma di un numero qualunque di variabili gaussiane mutuamente indipendenti $Z = \sum_j X_j$ è una variabile gaussiana con varianza $\Delta Z = \sum_j \sigma_j$. Il caso di variabili gaussiane a media non nulla si affronta in modo analogo. Si può ricordare intuitivamente queste proprietà delle variabili gaussiane pensando che vale per esse un teorema di Pitagora $\langle (X + Y)^2 \rangle = \langle X^2 \rangle + \langle Y^2 \rangle$ che vale nell'ipotesi di indipendenza statistica, cioè $\langle XY \rangle = \langle X \rangle \langle Y \rangle = 0$. ◀

ESERCIZIO 43. Utilizzare `matlab` per dare una dimostrazione empirica della legge dei grandi numeri. Costruire una matrice `x` di numeri casuali (`rand`) di dimensione $n \times N$ e calcolare la somma $X = \text{sum}(x)$; X sarà un vettore di lunghezza N . Calcolare poi $\text{eta} = (X - \langle X \rangle) / \text{std}(X)$. L'istogramma di `eta` deve essere ben riprodotto da una Gaussiana normale. Il comando `histfit(eta, 14, 'normal')` visualizza sia l'istogramma sia la curva gaussiana³⁷. Si può vedere il programma `LLNdemo.m` (Law.Large.Numbers.demo) in cui si sperimenta la somma di variabili variamente distribuite e si verifica la *LLN*. Quest'ultimo programma fa uso di tecniche di generazione di valori casuali distribuiti secondo una arbitraria d.d.p. che illustreremo più avanti (“Metodo degli alias” in 4.2). I codici `matlab` si trovano nel folder `dropbox ~ /Dropbox/MPF2023/Programmi/Matlab/Cap.3`.

2.6.2. Distribuzione di Poisson. Un altro esempio di limite “per grandi numeri” è quello che permette di stimare la distribuzione binomiale nel caso di un grande n e la probabilità

Poisson

³⁷`histfit` fa parte delle funzioni del toolbox di statistica. Nel programma `DeMoivre.m` si utilizza `'fitdist'` per valutare l'intervallo di confidenza dei parametri della distribuzione. Il secondo argomento (14) indica che si vuole ottenere un istogramma con 14 intervalli di campionamento - qualunque altra scelta ragionevole può andare.

vincolata a $np = \lambda = \text{costante}$. Si parte dalla distribuzione

$$P_n(p, q) = \binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k}$$

e nel limite di n grande, utilizzando la formula di Stirling, si può approssimare $\binom{n}{k} \approx n^k/k!$ e si ottiene così la *distribuzione di Poisson*

$$P_k(\lambda) = \frac{\lambda^k}{k!} e^{-\lambda}$$

che rappresenta la probabilità che in una successione di esperimenti ripetuti un gran numero di volte un dato effetto si presenti k volte: cioè $e^{-\lambda}$ è la probabilità che l'effetto non si presenti *mai*, $\lambda e^{-\lambda}$ la probabilità che si presenti una volta, etc. Il parametro λ caratterizza la distribuzione e coincide con il valore medio $\langle k \rangle$. Detta X la variabile aleatoria che segue la distribuzione di Poisson, la funzione generatrice si calcola facilmente $F(z) = \mathcal{M}(z^X) = \exp\{\lambda(z-1)\}$ e da questa si trova immediatamente

$$F'(1) = \lambda = \mathcal{M}(X)$$

$$F''(1) = \lambda^2 = \mathcal{M}(X(X-1)) = \mathcal{M}(X^2) - \lambda$$

$$\mathcal{D}(X) = \mathcal{M}(X^2) - \mathcal{M}(X)^2 = \lambda$$

La distribuzione di Poisson si caratterizza perciò attraverso la proprietà che la deviazione standard $\Delta = \sqrt{\mathcal{D}(X)}$ coincide con la radice quadrata del valor medio. Ne segue che se l'evento descritto dalla distribuzione di Poisson avviene in media N volte con N molto grande, la fluttuazione del risultato cresce “solo” come \sqrt{N} .

ESERCIZIO 44. Verificare la legge dei grandi numeri nel caso della somma di n variabili distribuite secondo la formula di Poisson.

SOLUZIONE. Posto $X = \sum_1^n X_i$, ciascuna delle X_i distribuite secondo la formula di Poisson con media λ , si definisce al solito $\eta = \frac{X - \mathbb{E}[X]}{\sqrt{\text{Var}[X]}}$; si tratta di valutare $\mathbb{E}[e^{it\eta}]$ nel limite di grande n . Si trova

$$\mathbb{E}[e^{it\eta}] = \mathbb{E}\left[\exp\left\{it \frac{\sum (X_i - \lambda)}{\sqrt{n\lambda}}\right\}\right]$$

D'altronde

$$\mathbb{E}[e^{itX_i}] = \sum e^{-\lambda} \frac{\lambda^n}{n!} e^{int} = \exp\{\lambda(e^{it} - 1)\}$$

e perciò, ponendo $t \rightarrow t/\sqrt{n\lambda}$ nell'equazione precedente otteniamo

$$\begin{aligned} \mathbb{E}[e^{it\eta}] &= \mathbb{E}\left[e^{it(X-\lambda)/\sqrt{n\lambda}}\right]^n = e^{-it\sqrt{n\lambda}} \exp\{n\lambda(e^{it/\sqrt{n\lambda}} - 1)\} \\ &= \exp\left\{n\lambda\left(e^{it/\sqrt{n\lambda}} - 1 - it/\sqrt{n\lambda}\right)\right\} \approx e^{n\lambda\left(-\frac{t^2}{2n\lambda}\right)} = e^{-\frac{1}{2}t^2}. \end{aligned}$$



2.6.3. Decadimento radioattivo. La distribuzione di Poisson si presta molto bene a descrivere il decadimento radioattivo, ad es. radio ($Z = 88$) in radon o polonio ($Z = 84$) in piombo, con emissione α . Il principio è che in un intervallo di tempo Δt un numero $\Delta n = \lambda N(t) \Delta t$ di nuclei di radio decadono e il fenomeno si può descrivere come un processo di Bernoulli in cui ogni nucleo ha una probabilità piccola di decadere ma il numero di nuclei è dell'ordine del numero di Avogadro; siamo nelle condizioni in cui Bernoulli converge a Poisson! Si ha pertanto che la probabilità di avere k decadimenti in un intervallo Δt sarà dato da

$$P_k(t) = \frac{a^k}{k!} e^{-a}, \quad a = N(t) = N(0) (1 - e^{-t/\tau})$$

dove $\tau = \lambda^{-1}$ è il *tempo di decadimento*. In media il numero di particelle α emesse in un intervallo di tempo $[0, t]$ sarà $N(t)$ e il numero di atomi di radio superstiti $N(0) e^{-t/\tau}$. L'esperimento di Rutherford e Geiger ³⁸ [RG10] riporta dati sul decadimento α del polonio e li analizza con la distribuzione di Poisson, nota dalla fine del XVIII secolo ma di cui non si fa menzione nell'articolo ³⁹. La tabella seguente è presa dal lavoro del 1910. I dati sono da leggere nel senso che ad es. in 408 casi si sono osservate 5 scintillazioni nell'intervallo di tempo fissato, in 27 casi 9 scintillazioni, etc. ⁴⁰ mentre la terza riga riporta i valori tratti dalla distribuzione di Poisson con $\lambda \approx 3.9$.

n	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
obs	57	203	383	525	532	408	273	139	45	27	10	4	0	1	1
Poisson	54	210	407	525	508	394	254	140	68	29	11	4	1	4	1

e il buon fit con la distribuzione di Poisson si può apprezzare dalla figura 2.22. Per realizzare l'istogramma si è applicato il metodo del *bootstrap* A.3.

Il parametro che entra nella distribuzione si può determinare semplicemente dal valor medio valutato sui dati, oppure fittando la curva teorica ai dati sperimentali (vedi il codice `poissonfit.m`). I dati di Rutherford e Geiger sono raggruppati in solo quattro successivi run e non è possibile estrarre una stima molto precisa dell'errore. Al giovane collaboratore matematico H. Bateman è dovuto lo sviluppo della teoria del processo di Poisson, inclusa nel lavoro già citato (il calcolo di Bateman è riportato in appendice a Pag.135).

ESERCIZIO 45. Nel preparare N panettoni, si immettono nell'impasto n acini di uvetta sultanina. Qual'è la probabilità che uno qualunque dei panettoni contenga almeno un acino?

SOLUZIONE. Considerando gli acini di volume trascurabile rispetto a quello dei panettoni, possiamo considerare che la distribuzione delle uvette sia uniforme. La presenza di una uvetta in un dato panettone aumenterà dunque in media di $1/N$ per ogni chicco immesso nell'impasto. Il problema è quindi equivalente a una successione di n estrazioni di Bernoulli con $p = 1/N$, e quindi $\lambda = np = n/N$ è il parametro di Poisson che caratterizza il limite della distribuzione di Bernoulli per grandi N e n (p è piccolo e $np = \lambda$). La probabilità cercata è quindi data da $1 - P(0) = 1 - \exp\{-np\} = 1 - \exp\{-n/N\}$. ◀

³⁸Si trova copia del lavoro nel folder *Documents*.

³⁹Dall'articolo di Rutherford *et al* si impara qualcosa sulla condizione di lavoro di un Fisico inizio secolo. Giornate intere a contare a occhio nudo i lampi di luce sullo schermo a scintillazione.

⁴⁰I dati sono riportati sul libro di Jeffreys [Jef38]

ESERCIZIO 46. Supponiamo che in una gara di tiro al piattello tra principianti la probabilità di colpire il bersaglio con un singolo colpo sia 10^{-2} . Qualè allora la probabilità di colpire il bersaglio due o più volta in una successione di 200 colpi?

SOLUZIONE. È ancora la distribuzione di Poisson che ci fornisce il risultato: $P_k(\lambda)$ con $\lambda = 200 \times 10^{-2} = 2$; perciò la probabilità di non colpire mai il bersaglio è e^{-2} e per una volta sola $2e^{-2}$. La probabilità di colpire due o più volte sarà dunque $1 - 3e^{-2} \approx 0.59$. ◀

ESERCIZIO 47. La bozza di un libro di 500 pagine contiene 500 errori di stampa. Qualè la probabilità che una data pagina contenga non più di due errori?

SOLUZIONE. Ancora utilizziamo la distribuzione di Poisson con $\lambda = 1$. Ne segue che la probabilità richiesta è $(1 + 1 + 1/2)e^{-1} \approx 0.92$. Perché il risultato sia ragionevole devono valere le ipotesi semplificatrici secondo cui tutte le pagine contengono lo stesso numero di caratteri, il tipografo ha lavorato con la stessa attenzione dalla prima all'ultima pagina, il che può non essere sempre vero... ◀

ESERCIZIO 48 (Distribuzione di Polya). In un contenitore sono presenti w palle bianche e b nere, $b + w = N$. Si estrae a caso una palla e se ne aggiungono $s \geq 0$ dello stesso colore. Questa procedura è ripetuta n volte. Sia X la v.a. che prende il valore k se in queste n estrazioni si ottengono $k(\leq n)$ palle bianche. Determinare la distribuzione di probabilità di X , il valore medio e la dev. standard. (Il caso $s = 0$ coincide con la distribuzione binomiale). Si scriva un programma `matlab` per simulare la v.a. e verificare il risultato.

SOLUZIONE. La distribuzione di Polya è descritta in dettaglio sul libro di Fisz [Fis63], §5.4. La distribuzione di probabilità è data da ⁴¹

$$(26) \quad \mathbb{P}(X = k | n, w, b, s) = \binom{n}{k} \frac{w(w+s) \dots (w+(k-1)s) b(b+s) \dots (b+(n-k-1)s}{(w+b)(w+b+s) \dots (w+b+(n-1)s)}$$

Per il valor medio $\mathbb{E}[X]$ e per la deviazione standard $\text{std}[X]$ si ha

$$(27) \quad \mathbb{E}[X] = np; \quad \text{std}[X]^2 = \Delta(X) = npq \frac{1 + n\alpha}{1 + \alpha}$$

avendo posto $p = w/(w+b)$, $q = b/(w+b)$, $\alpha = s/(w+b)$ per avvicinarsi alla notazione della distribuzione di Bernoulli. Si noti che il valore medio non dipende dal parametro s ! In Fig. 2.23 il risultato di 40 estrazioni con $w = 5$, $b = 4$, $s = 2$. ◀

2.6.4. Teorema del Limite centrale. Il teorema di De Moivre-Laplace può essere esteso a una formulazione molto generale: a parte condizioni tecniche che specificheremo, il risultato è che la somma di un grande numero di variabili aleatorie indipendenti, una volta sottratta del suo valore medio e normalizzata con la sua deviazione standard, risulta distribuita secondo la legge normale di Gauss. L'enunciato preciso ripreso dal testo di Rozanov [Roz69] è il seguente: DEFINIZIONE: Data una successione di variabili aleatorie $\{\xi_k, k = 1, 2, \dots\}$ aventi media finita $\alpha_k = \mathbb{E}[\xi_k]$ e deviazione standard finita $\sigma_k = \sqrt{\text{Var}(\xi_k)}$ si consideri la “somma normalizzata”

$$S_n^* = \frac{S_n - \mathbb{E}[S_n]}{\sqrt{\text{Var}[S_n]}}$$

⁴¹Si noti che la formula è ambigua per $k = 0$ e $k = n$, ma si capisce che è da intendere che il numeratore della frazione vale $b(b+s) \dots (b+(n-1)s)$ per $k = 0$ e $w(w+s) \dots (w+(n-1)s)$ per $k = n$.

dove $S_n \equiv \sum_{i=1}^n \xi_i$. Allora si dice che la successione $\{\xi_k, k = 1, 2, \dots\}$ soddisfa il *teorema limite centrale* se

$$\lim_{n \rightarrow \infty} \mathbb{P}(x' < S_n^* < x'') = \frac{1}{\sqrt{2\pi}} \int_{x'}^{x''} e^{-\frac{1}{2}x^2} dx$$

TEOREMA 2.3. Se la successione di v.a. **indipendenti** $\{\xi_k, k = 1, 2, \dots\}$ aventi medie $\alpha_k = \mathbb{E}[\xi_k]$ e varianze $\sigma_k^2 = \text{Var}(\xi_k)$ finite soddisfa la condizione di Lyapounov

$$(28) \quad \lim_{n \rightarrow \infty} \frac{1}{\mathcal{I}_n^3} \sum_{k=1}^n \mathbb{E}[|\xi_k - \alpha_k|^3] = 0$$

Limite centrale

dove $\mathcal{I}_n^2 \equiv \text{Var}(S_n) = \sum \sigma_k^2$, allora la successione soddisfa il teorema limite centrale ⁴².

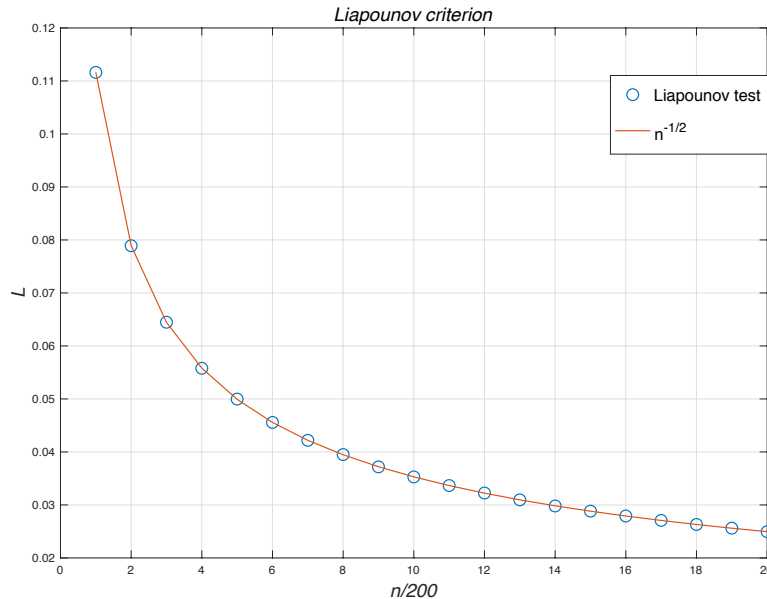
Omettiamo la dimostrazione basata sullo sviluppo di Taylor della funzione generatrice che si può ritrovare sul testo di Rozanov. Osserviamo che il teorema non richiede che le variabili ξ_k abbiano la stessa distribuzione; dunque il teorema del limite centrale caratterizza una situazione generica e giustifica il fatto empirico che la distribuzione di Gauss risulta onnipresente in natura là dove la grandezza fisica misurata riceve contributi da un grande numero di perturbazioni casuali tra loro indipendenti.

ESERCIZIO 49. Dimostrare che nel caso di n variabili distribuite secondo la legge binomiale la condizione di Lyapounov è rispettata.

SOLUZIONE. Nel caso di variabili ugualmente distribuite, quindi con medesimo valore di $\alpha_k = \alpha$ e $\sigma_k = \sigma$ la condizione di Lyapounov è soddisfatta in modo banale. Infatti

$$\frac{1}{\mathcal{I}_n^3} \sum_{k=1}^n \mathbb{E}[|\xi_k - \alpha_k|^3] = \frac{1}{n^{3/2} \sigma^3} n \mathbb{E}[|\xi - \alpha|^3] = \mathcal{O}(1/\sqrt{n}).$$

Possiamo tuttavia toccare con mano quanto affermato realizzando una simulazione delle variabili binomiali con il metodo degli alias ⁴³ e poi calcolando il test di Lyapounov per vari valori di n . La figura seguente presenta il risultato che risulta assolutamente convincente.



L'unico fattore che potrebbe sovvertire il risultato è costituito dal valore di aspettazione $\mathbb{E}[|\xi - \alpha|^3]$, ma nel caso binomiale questo è senz'altro finito. Osserviamo perciò che il criterio di Lyapounov può essere violato nel caso di una distribuzione di probabilità continua su un intervallo infinito e con decrescita all'infinito che renda divergente l'integrale $\int \rho(x) x^3 dx$. ◀

⁴²Secondo un'altra possibile formulazione la tesi vale nel caso in cui l'Eq(2.3) sia valida anche se l'esponente 3 viene sostituito da $2 + \delta$ per un qualche $\delta > 0$.

⁴³Si veda il codice matlab `lyapounov.m` in `dropbox/MPF2023/Programmi/Matlab/Knuth/KRAND`.

2.6.5. Non universalità della distribuzione normale. Posto che il teorema del limite centrale offre condizioni molto generali sulla somma di un grande numero di variabili aleatorie perchè emerga la distribuzione normale, viene spontaneo chiedersi in quali condizioni il limite della somma di v.a. si discosti significativamente dalla distribuzione normale. Nel testo di Whittaker e Robinson ([WR44]) viene riportato il seguente esempio (da Hausdorff): supponiamo che ci sia un numero imprecisato di perturbazioni $\epsilon_1, \epsilon_2, \dots, \epsilon_n, \dots$ che influenzano un dato fenomeno, tutte distribuite secondo la legge

$$(29) \quad \rho(x) = \frac{1}{2} \exp\{-|x|\}$$

L'effetto sia dato dall'accumularsi di un gran numero di v.a. ϵ_n , con intensità decrescente:

$$(30) \quad F = \frac{2}{\pi} (\epsilon_1 + \epsilon_2/3 + \epsilon_3/5 + \dots + \epsilon_n/(2n-1) + \dots)$$

Assumendo che le varie perturbazioni siano statisticamente indipendenti, l'effetto complessivo seguirà la legge di probabilità che possiamo dedurre calcolando la funzione generatrice. Per ϵ_1 abbiamo

$$(31) \quad \Phi_1(t) = \int_{-\infty}^{\infty} e^{itx-|x|} dx = \frac{1}{1+t^2}$$

Per il termine generico $\epsilon_n/(2n-1)$ si ha

$$(32) \quad \Phi_n(t) = \int_{-\infty}^{\infty} e^{itx-(2n-1)|x|} dx = \frac{1}{1 + \frac{t^2}{(2n-1)^2}}$$

Ne segue che la somma delle perturbazioni avrà una distribuzione di probabilità determinabile dalla funzione generatrice *prodotto* delle funzioni generatrici

$$(33) \quad \Phi_{1:N}(t) = \prod_{n=1}^N \left(1 + \frac{t^2}{(2n-1)^2} \right)^{-1}$$

Borel-Cantelli

che nel limite $N \rightarrow \infty$ tende a $1/\cosh(\pi t/2)$, diversa da una Gaussiana. Siamo qui in presenza di una somma di infiniti effetti che si combinano dando una somma convergente diversa dalla Gaussiana. Il termine n -esimo si può considerare "piccolo" al crescere di n in quanto ha media 0 e deviazione standard $1/(2n-1)$. Domanda: la successione $\epsilon_n/(2n-1)$ soddisfa la condizione di Lyapounov? Per una risposta provare a utilizzare il metodo degli alias, estrarre un grande numero di v.a. secondo la distribuzione di Hausdorff e valutare il parametro di Lyapounov.

2.6.6. Successioni di eventi - Lemmi di Borel-Cantelli. Data una successione infinita di eventi $\{A_j \mid j = 1, 2, \dots\}$ ci sono due possibilità: la somma delle probabilità $\sum_j \mathbb{P}(A_j)$ converge a un valore finito oppure diverge. Nel primo caso vale il teorema ⁴⁴ di Borel-Cantelli:

TEOREMA 2.4. Primo lemma di Borel-Cantelli: Se la serie $\sum_{j=1}^{\infty} \mathbb{P}(A_j)$ converge allora con probabilità **uno** (cioè "è certo che") solo un numero finito di eventi della successione può realizzarsi.

D'altro canto

TEOREMA 2.5. Secondo lemma di Borel-Cantelli: se gli infiniti eventi A_j sono *indipendenti* e se la serie $\sum_{j=1}^{\infty} \mathbb{P}(A_j)$ diverge, allora con probabilità **uno** un numero infinito di eventi della successione si realizzerà.

Omettiamo la dimostrazione che si può trovare su [Roz69, Fel68]. Questi due teoremi matematici sono fonte di qualche paradosso divertente. Consideriamo il seguente esempio:

ESERCIZIO 50. Consideriamo una successione di estrazioni a caso secondo Bernoulli: 0 e 1 con la stessa probabilità. Si chiede quale sia la probabilità che nella successione, di lunghezza illimitata, compaia infinite volte la "parola" 101.

⁴⁴Il termine più impiegato dai Matematici è *lemma*, in quanto propedeutico a ulteriori sviluppi.

Il lemma di Borel-Cantelli assicura che 101 appare *infinite volte* nella successione infinita! E dice anche di più: qualunque stringa fatta di 0 e 1 apparirà con certezza nella successione. Ma la stringa si può interpretare in codice Morse, da cui l'apparente paradosso che nella successione infinita con probabilità uno troveremo ⁴⁵ la Divina Commedia... Questo significa solo che dobbiamo utilizzare i due lemmi con cautela. Non è dato sapere quanto sarà lunga l'attesa per trovare anche solo il testo di queste lezioni in codice Morse! Adesso proponiamo un esercizio divertente: utilizzando le funzioni `strfind` e `dec2bin` in `matlab` effettuiamo un esperimento sulla frequenza di 101 in una successione casuale. Il programma costruisce un segnale casuale in codice binario (0,1) lungo N caratteri e cerca quante volte e' contenuta una data stringa "*str*" - secondo il lemma di Borel-Cantelli la stringa si presenta in un numero infinito di volte per $N = \infty$. Il codice completo si trova su `dropbox`. Il nocciolo del programma è costituito da

```
str = dec2bin(randi(2^b)-1,b) % costruire la stringa da cercare lunga 'b' bits
A = dec2bin(randi(2^n2, n1, 1)-1,n2); % il messaggio binario
A = A(:)'; % trasforma A da matrice n1*n2 a matrice a una sola riga
match = strfind(str,A); freq = length(match)/length(A);
```

Il risultato 'match' rappresenta il numero di matches, cioè quante volte la stringa binaria "*str*" si trova nel "messaggio" codificato in *A*. Si trova facilmente che aumentando la lunghezza della stringa *A* il valore di "freq" per *str* = '101' si mantiene vicino a $\approx 1/8$ e dunque una proporzione costante di triplette di caratteri in *A* coincide con '101'. In generale si troverà che la frequenza tende a mantenersi dell'ordine di 2^{-b} , se al posto della stringa binaria '101' ne introduciamo una formata da *b* bits. Ad es. una specie di SOS binario, ossia 111000111, compare in media ogni ≈ 512 gruppi successivi di 9 caratteri. È chiaro allora come il paradosso "della scimmia" si risolva banalmente: la stringa binaria equivalente al testo di queste lezioni è lunga $b = \mathcal{O}(10^6)$ bits e la potenza 2^b è un numero formato da circa un milione di cifre. Appare evidente che il concetto di "successione infinita" nei lemmi di Borel-Cantelli debba essere interpretata con molta cautela. Si trova in Fig.2.24 il risultato di un esperimento con stringhe scelte a caso di *b* bits con valori fino a $b = 15$; la successione è lunga un massimo di 5×10^8 bits. Le linee evidenziate con le stelle corrispondono al valore 2^{-b} . Se da 15 bit si passa a un messaggio più lungo, ad es. "Buon Natale" (≈ 200 bits) questo si troverebbe con una frequenza dell'ordine di 10^{-60} (!), in parole povere in media un match in un tempo paragonabile all'età dell'Universo ($\mathcal{O}(10^{10})$ anni). La battuta, che spesso si sente ripetere, riguardo il fatto che se si lasciasse una scimmia battere sui tasti di un computer ⁴⁶ "prima o poi" ne verrebbe fuori la *Divina Commedia* non considera il fatto che quel "poi" sarebbe a un tempo molti ordini di grandezza superiore all'età dell'Universo (anche a una battitura di un carattere al secondo l'età dell'Universo corrisponde alla speranza matematica di trovare un match con una stringa di circa 60 bits!). Lo stesso esperimento si può effettuare su stringhe di caratteri esadecimali, utilizzando `dec2hex` in `matlab`. In questo caso ovviamente la probabilità di un ciascun match scende da 2^{-b} a 16^{-h} per stringhe di *h* caratteri nell'alfabeto $\{0, 1, 2, \dots, 9, A, B, C, D, E, F\}$ ma la logica è la stessa. In una stringa infinita la probabilità di un match è *uno* ma il tempo di attesa del primo match cresce di conseguenza a valori dell'ordine 16^h e possiamo dimenticarci di trovare mai una stringa lunga più di 14 caratteri. Infatti la vita dell'universo che conosciamo è dell'ordine di 10^{16} misurata in secondi e al ritmo di un carattere al secondo questo corrisponde al tempo minimo per un match su una stringa di 14 caratteri!

2.7. Analisi statistica di un set di dati con matlab

Il problema di generare un insieme di esemplari estratti a caso da una data distribuzione di probabilità trova soluzione nei casi più semplici con il metodo degli alias introdotto alla sezione 4.2 o in casi più complessi che comportano un grande numero di gradi di libertà attraverso le tecniche Montecarlo di cui si discuterà più avanti al Cap. 4. È indubbiamente molto importante anche il problema inverso, cioè *individuare la distribuzione di probabilità da associare a un dato insieme di dati sperimentali*. Per questo esistono varie tecniche che sono sviluppate nei trattati di statistica. Qui ci limitiamo a considerare alcuni strumenti disponibili nei toolbox statistici di `matlab`.

⁴⁵Spesso si è fatto riferimento a questo paradosso attribuendo la successione di bits casuali all'opera di un scimmia che batta sui tasti 0 e 1.

⁴⁶Si veda ad esempio il testo del Feller [Fel68] pag. 202.

histfit.

```
>> histfit(data,nbins,dist)
```

analizza un set di dati (un array di numeri reali) e produce un istogramma di *nbins* bins fittando i dati alla distribuzione “dist”. Per default ‘nbins’ è posta uguale alla radice quadrata del numero di elementi in “data” e “dist” è la distribuzione normale, ma può essere specificata secondo l’elenco che si trova con ‘doc histfit’.

fitdist.

```
>> pd = fitdist(x, dist)
```

ritorna le specifiche secondo la distribuzione “dist” che si adatta meglio ai dati in “x”. Esempio:

```
>> p = poisson(20,5) % Poisson ditribution on 0:19 with mean 5
>> K=kalias(p);      % il metodo degli alias e` descritto piu` avanti
>> x= krand(1000,K)';
>> pd = fitdist(x,'poisson');

Poisson distribution
lambda = 4.988      [4.84958, 5.12642]

`pd` pu`o essere utilizzato per rappresentare la distribuzione di
probabilit`a con il comando:
>> y = pdf(pd, 0:19);
```

Spetta a chi utilizza *fitdist* provare a indovinare la distribuzione più corretta per fittare i dati. La lista delle distribuzioni di probabilità si trova aggiornata sulla documentazione (“doc fitdist”).

fitgmdist. La routine permette di analizzare i dati nell’ipotesi che si tratti di una distribuzione “multimodale”, cioè la combinazione lineare di distribuzioni gaussiane

$$p(x) = \sum A_j \exp\{-(x - x_j)/(2\sigma_j)\}/\sqrt{2\pi\sigma_j}$$

. Esempio:

```
>> opt=statset('MaxIter',1000);
>> x=randn(1000,1);
>> y=3+randn(2000,1); % abbondanza doppia rispetto a x
>> fitgmdist([x;y],2,opt);
Gaussian mixture distribution with 2 components in 1 dimensions
Component 1:
Mixing proportion: 0.661297
Mean:      3.0122
Component 2:
Mixing proportion: 0.338703
Mean:      0.0426
```

2.8. Probabilità quantistica: il teorema di Bell

God does not play dice

Albert Einstein

*It doesn't matter how beautiful your theory is,
it doesn't matter how smart you are. If it
doesn't agree with experiment, it's wrong*

Richard P. Feynman

Bell's theorem

La Meccanica Quantistica è basata sul principio che di ogni processo fisico microscopico si possa formulare una previsione solo in termini di probabilità. Non si tratta dunque, a differenza della Fisica Classica, di una teoria deterministica. Alcuni Fisici molto illustri quali Albert Einstein formularono già negli anni '30 (del XX secolo) l'opinione che la meccanica quantistica fosse incompleta, che cioè ci dovesse essere un livello più profondo, non ancora sperimentabile con le tecniche a disposizione, in cui fossero attivi gradi di libertà che soddisferebbero a leggi deterministiche, ma che alla scala atomica si manifesterebbero solo in modo casuale, similmente a quanto accade in termodinamica statistica, le cui leggi sono manifestazioni macroscopiche di un ordine microscopico perfettamente deterministico. La problematica, abbastanza complessa, si sviluppò a partire dall'articolo di Einstein-Podolski-Rosen [AER35] e trovò una soluzione probabilmente definitiva nel lavoro di J. Bell [Bel64]. In questo lavoro Bell dimostra che, ammettendo l'esistenza di gradi di libertà non accessibili agli strumenti ordinari (filtri polarizzatori, Stern-Gerlach, etc.) soggetti alle leggi probabilistiche secondo il modello classico sviluppato da Poincaré e Kolmogorov, si può dedurre una disuguaglianza cui deve sottostare la funzione di correlazione tra misure relative a due microsistemi che nel passato sono stati in interazione - ad es. due particelle con spin prodotte in uno stato di momento angolare totale nullo. Questa disuguaglianza è violata dalle formule che calcolano la correlazione in accordo con la meccanica quantistica. Questo fatto aprì la possibilità di esperimenti tesi a mettere alla prova la correttezza della MQ e a sondare perciò la possibilità di avere gradi di libertà nascosti. Gli esperimenti condotti sull'arco di trent'anni hanno confermato che la disuguaglianza di Bell è in effetti violata e quindi la MQ è incompatibile con un modello probabilistico classico [AAR82][Xia11]. Si conclude cioè che sebbene le leggi quantistiche siano di natura probabilistica, tuttavia non è possibile utilizzare i concetti di spazio degli eventi e di misura di probabilità che abbiamo introdotto in queste lezioni ⁴⁷. Un caso speciale delle disuguaglianze di Bell si trova su [Wic95], Appendice a cura di W.J. Faris, e si può riprodurre in poche righe con un argomento che si deve a **D. Bohm** che ha riformulato l'EPR in termini di correlazione di spin. Immaginiamo due particelle di spin $\frac{1}{2}\hbar$ prodotte in uno stato a momento angolare totale zero (singoletto) ⁴⁸

$$|0\rangle = \frac{|+-\rangle - |-+\rangle}{\sqrt{2}}$$

essendo $|+-\rangle$ lo stato in cui s_3 è $+\frac{1}{2}\hbar$ per la prima particella e -1 per la seconda ⁴⁹. Lo stato è invariante sotto rotazioni spaziali, quindi la scelta della direzione 3 è puramente convenzionale. Lo stato è inoltre caratterizzato da momento lineare totale nullo (decadimento di una particella scalare a riposo), quindi le due particelle vengono rivelate in posizioni diametralmente opposte rispetto al punto dove avviene il decadimento. Orientiamo gli assi in modo che la direzione dal centro di decadimento ai rivelatori sia l'asse x ; i rivelatori possono essere ruotati nel piano $y-z$. Se entrambi i rivelatori sono orientati in direzione z si ha che se l'osservatore a sinistra \mathcal{O}_1 osserva una successione di valori $+- - + - \dots$ totalmente a caso, l'osservatore a destra \mathcal{O}_2 registrerà valori completamente correlati a quelli di \mathcal{O}_1 e cioè $- + + - - + \dots$. Siamo allora nella condizione di poter prevedere l'esito della

⁴⁷Un racconto del Fisico Colin Bruce[C.98], scritto in perfetto stile ispirato ai romanzi di A. Conan Doyle (Le avventure di Sherlock Holmes) illustra un caso immaginario in cui una sorta di effetto EPR colpisce un concorso a premi tipo "gratta-e-vinci"; sebbene gli accadimenti del racconto siano del tutto impossibili da verificarsi in pratica, cionondimeno il racconto contiene in modo mirabile tutti gli aspetti cruciali del problema.

⁴⁸Il processo si può realizzare con un pione π^0 , particella neutra a spin 0, che decade (raramente) in una coppia elettrone-positrone; il modo di decadimento più frequente è tuttavia quello in due γ (99%). Negli esperimenti realizzati negli anni '70-'80 non sono stati utilizzati mesoni π^0 bensì processi di "cascate atomiche".

⁴⁹Uno stato come questo per i due spin non è fattorizzabile rispetto ai due gradi di libertà e viene denominato "entangled", ossia "allacciato" o "annodato".

misura di \mathcal{O}_1 senza interagire con il suo apparato sulla base della misura da parte di \mathcal{O}_2 , esattamente come prefigurato da EPR [AER35]. Citiamo infatti i passi salienti introduttivi del lavoro di Einstein e collaboratori: *Every element of physical reality must have a counterpart in the physical theory. E più avanti: If, without in any way disturbing a system, we can predict with certainty (i.e. with probability one) the value of a physical quantity, then there exists an element of physical reality corresponding to this physical quantity.* Possiamo allora, in base al risultato precedente sulla correlazione delle misure di spin, asserire che ci deve essere, secondo EPR, un elemento di realtà fisica nell'apparato \mathcal{O}_1 che corrisponde con certezza alla misura dello spin da parte di \mathcal{O}_2 . Se indichiamo con A la misura $+$ di \mathcal{O}_1 , la misura $+$ da parte di \mathcal{O}_2 , che indicheremo con A' , deve avere un corrispettivo \bar{A} nell'apparato di \mathcal{O}_1 . La notazione insiste sul fatto che le due misure possibili $+$ per A e $-$ per \bar{A} sono eventi disgiunti.

Immaginiamo ora che il rivelatore di destra venga ruotato fino a formare un angolo $2\pi/3$ con la direzione iniziale A' (diciamo in posizione B') oppure $-2\pi/3$ (posizione C'). Se misuriamo lo spin della particella a sinistra in direzione A e quello della particella a destra in direzione B' si ha che la probabilità di trovare $+1$ per A e $+1$ per B' è data da $\frac{1}{2} \sin^2(\pi/3) = 3/8$. Infatti per spin $1/2$ la matrice di rotazione intorno all'asse x è

$$U(\alpha) = \begin{pmatrix} \cos(\alpha/2) & i \sin(\alpha/2) \\ i \sin(\alpha/2) & \cos(\alpha/2) \end{pmatrix}.$$

In base alla natura dello stato di singoletto, una misura che dia $+1$ per B' è equivalente alla misura -1 per B ossia \bar{B} . Questo fatto in termini classici si può riesprimere dicendo allora che l'evento congiunto $A = \{\sigma_A = +1\}$ e $B' = \{\sigma_{B'} = +1\} = \{\sigma_B = -1\}$ ha probabilità $3/8$, che scaturisce dal fattore $1/2$ dalla forma di $|0\rangle$ e dal fattore $3/4$ dalla matrice di rotazione $\sin^2 \pi/3$. Notare che non è possibile misurare contemporaneamente lo stesso spin in due direzioni diverse, tuttavia ci si può affidare al fatto che $B' = 1$ (a destra) avviene con probabilità uno quando $B = -1$ (a sinistra) e quindi devono essere considerati *il medesimo evento* dal punto di vista di una teoria basata sulla probabilità classica (secondo EPR corrispondono allo stesso elemento di realtà fisica). Concludiamo allora che $\mathbb{P}(A \cap \bar{B}) = 3/8$. Se consideriamo il setup dell'esperimento ruotato di $2\pi/3$ concludiamo per simmetria che $\mathbb{P}(B \cap \bar{C}) = \mathbb{P}(C \cap \bar{A}) = 3/8$. Tuttavia gli eventi $A \cap \bar{B}$, $B \cap \bar{C}$ e $C \cap \bar{A}$ sono *mutuamente disgiunti* e quindi la probabilità della loro unione si calcola sommando le probabilità (in generale si applica il principio di inclusione/esclusione). *Ma la somma delle probabilità dà $3 \times 3/8 > 1$!* Dunque le previsioni basate sulla meccanica quantistica sono incompatibili con uno schema probabilistico classico. Il teorema generale, sempre riferito a correlazione di spin si trova su [Bel64].

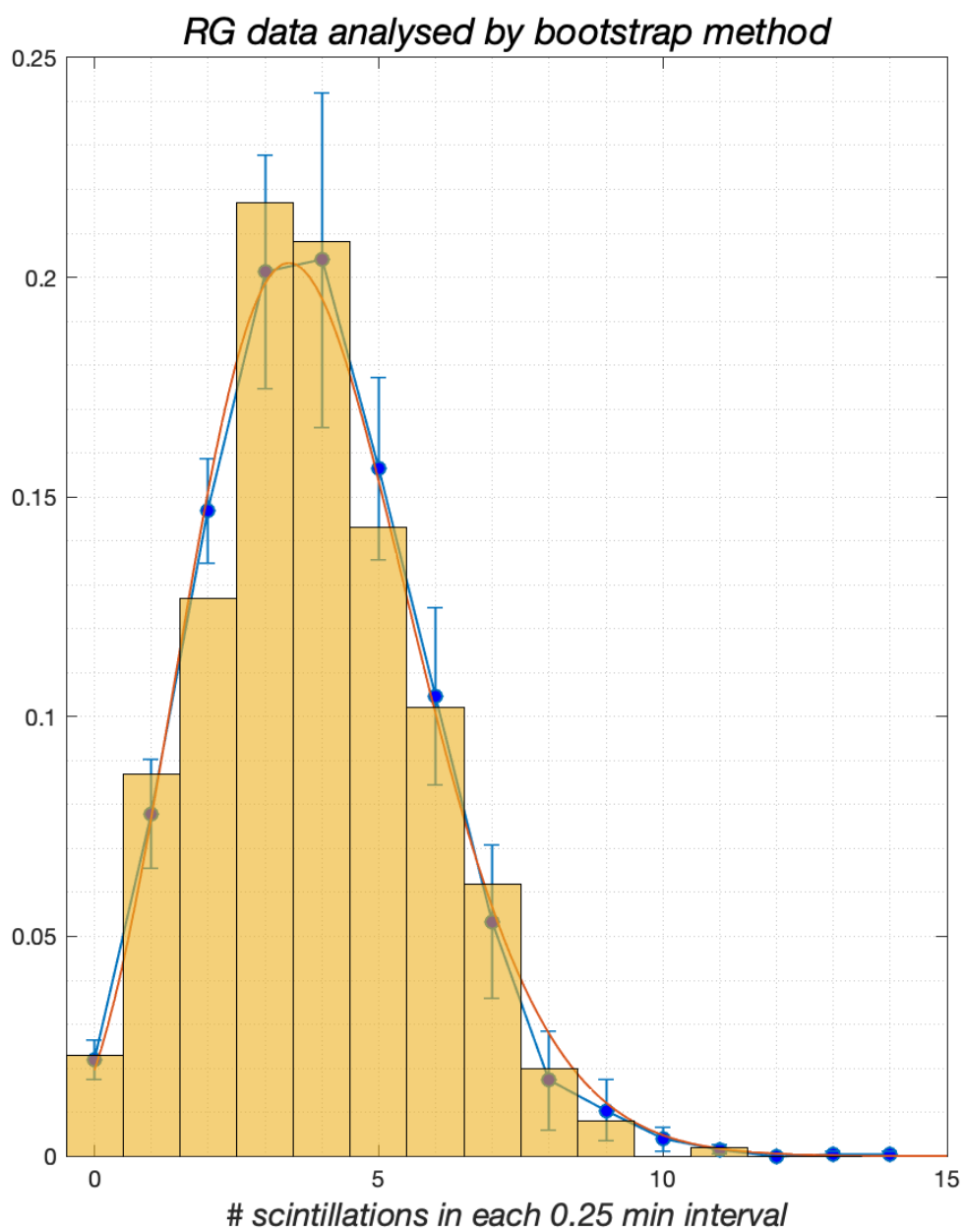


FIGURA 2.22. Analisi dei dati di Rutherford e Geiger (linea con errorbars) in termini di distribuzione di Poisson (linea rossa). Sovrainposto un istogramma ottenuto utilizzando il metodo *bootstrap* (si veda l'appendice A.3).

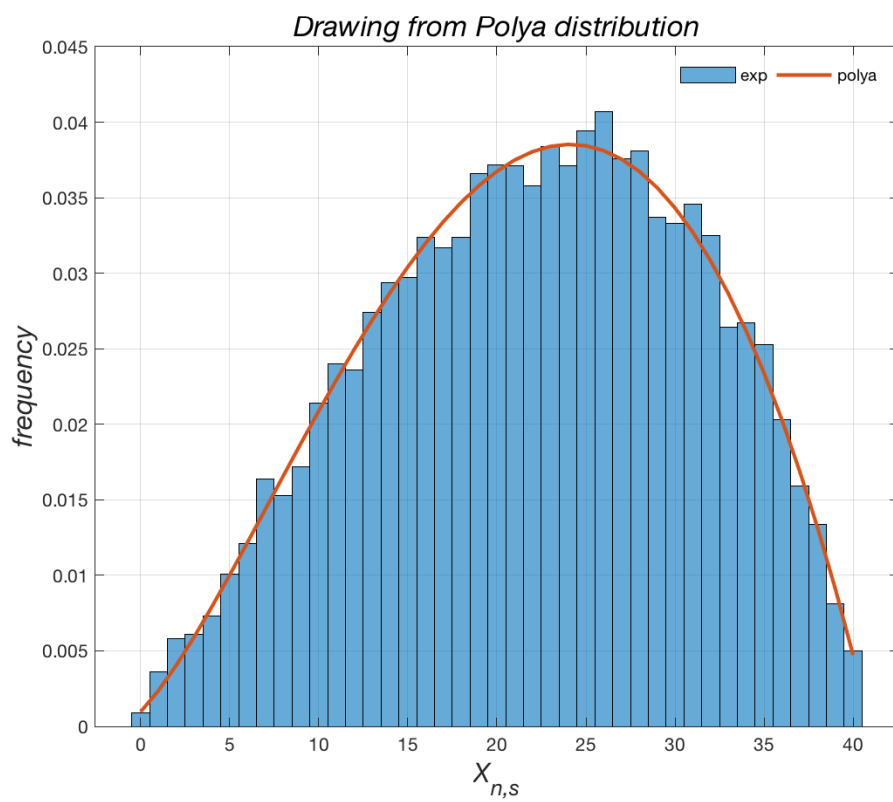


FIGURA 2.23. La distribuzione di Polya (rosso) e il risultato di un esperimento numerico (blu) ($n=40$, $w=5$, $b=4$, $s=2$).

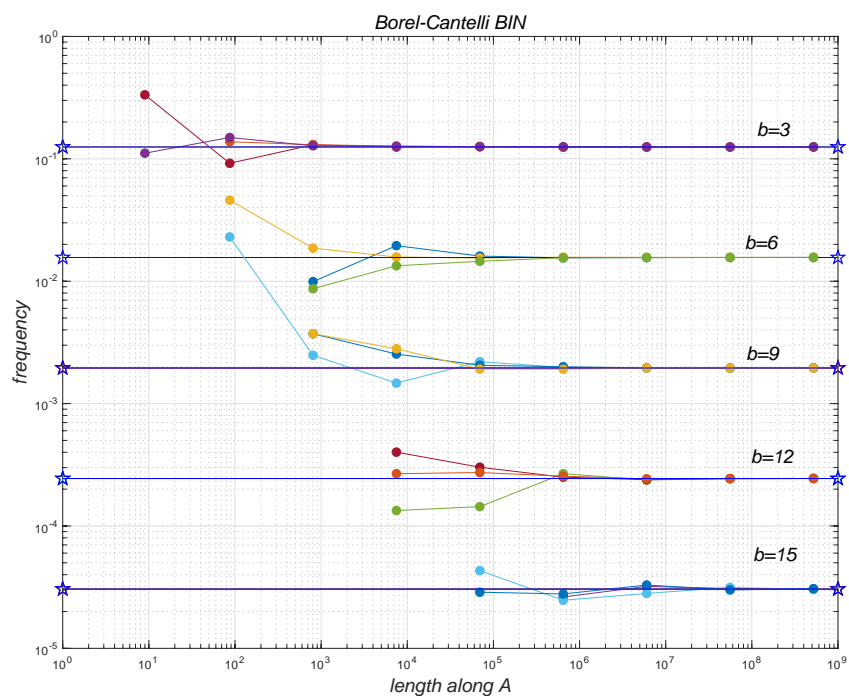


FIGURA 2.24. Match di una stringa binaria in un messaggio binario A.

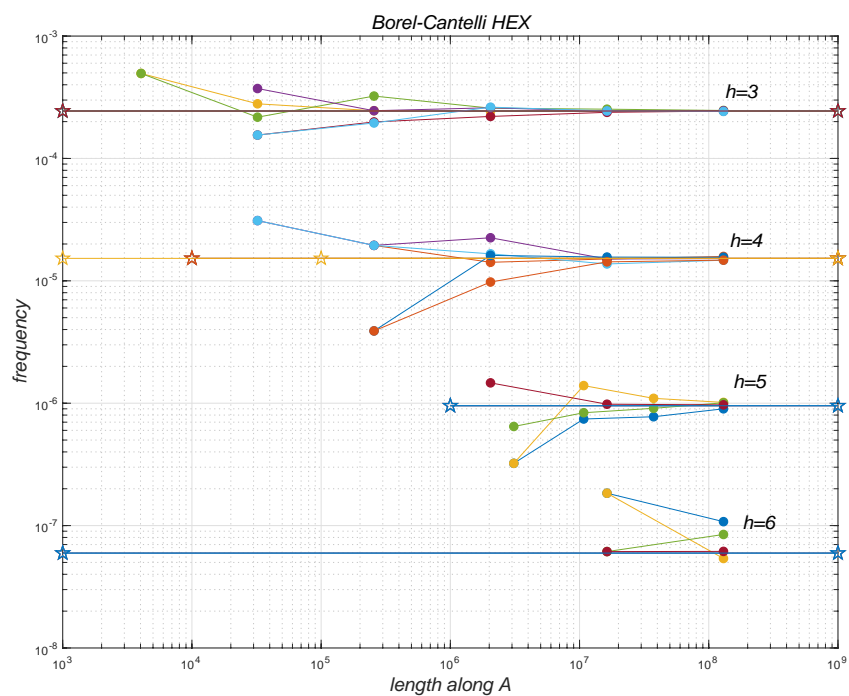


FIGURA 2.25. Come la figura precedente ma con stringhe esadecimali.

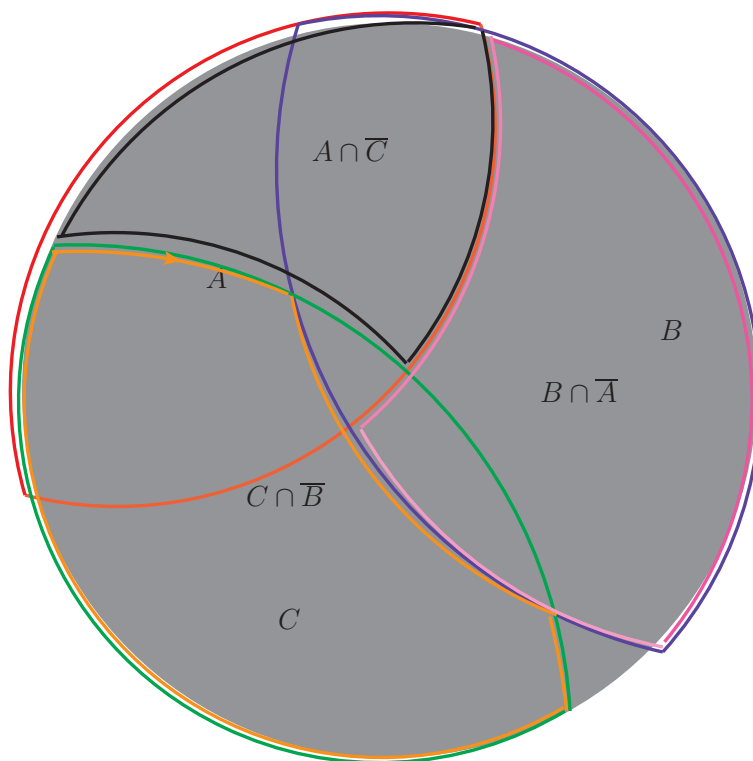


FIGURA 2.26. Lo schema degli eventi (A, B, C) . A è individuato da un contorno rosso, B da un contorno viola e C verde. Le intersezioni $A \cap \overline{C}$ e $C \cap \overline{B}$ (contorno nero e arancione) sono ovviamente disgiunte. Lo stesso vale per le altre intersezioni che entrano nell'argomento di questa sezione.

of lanthanum is $7/2$, hence the nuclear magnetic moment as determined by this analysis is 2.5 nuclear magnetons. This is in fair agreement with the value 2.8 nuclear magnetons determined from La III hyperfine structures by the writer and N. S. Grace.⁹

⁹ M. F. Crawford and N. S. Grace, Phys. Rev. **47**, 536 (1935).

This investigation was carried out under the supervision of Professor G. Breit, and I wish to thank him for the invaluable advice and assistance so freely given. I also take this opportunity to acknowledge the award of a Fellowship by the Royal Society of Canada, and to thank the University of Wisconsin and the Department of Physics for the privilege of working here.

MAY 15, 1935

PHYSICAL REVIEW

VOLUME 47

Can Quantum-Mechanical Description of Physical Reality Be Considered Complete?

A. EINSTEIN, B. PODOLSKY AND N. ROSEN, *Institute for Advanced Study, Princeton, New Jersey*

(Received March 25, 1935)

In a complete theory there is an element corresponding to each element of reality. A sufficient condition for the reality of a physical quantity is the possibility of predicting it with certainty, without disturbing the system. In quantum mechanics in the case of two physical quantities described by non-commuting operators, the knowledge of one precludes the knowledge of the other. Then either (1) the description of reality given by the wave function in

quantum mechanics is not complete or (2) these two quantities cannot have simultaneous reality. Consideration of the problem of making predictions concerning a system on the basis of measurements made on another system that had previously interacted with it leads to the result that if (1) is false then (2) is also false. One is thus led to conclude that the description of reality as given by a wave function is not complete.

1.

ANY serious consideration of a physical theory must take into account the distinction between the objective reality, which is independent of any theory, and the physical concepts with which the theory operates. These concepts are intended to correspond with the objective reality, and by means of these concepts we picture this reality to ourselves.

In attempting to judge the success of a physical theory, we may ask ourselves two questions: (1) "Is the theory correct?" and (2) "Is the description given by the theory complete?" It is only in the case in which positive answers may be given to both of these questions, that the concepts of the theory may be said to be satisfactory. The correctness of the theory is judged by the degree of agreement between the conclusions of the theory and human experience. This experience, which alone enables us to make inferences about reality, in physics takes the form of experiment and measurement. It is the second question that we wish to consider here, as applied to quantum mechanics.

Whatever the meaning assigned to the term *complete*, the following requirement for a complete theory seems to be a necessary one: *every element of the physical reality must have a counterpart in the physical theory*. We shall call this the condition of completeness. The second question is thus easily answered, as soon as we are able to decide what are the elements of the physical reality.

The elements of the physical reality cannot be determined by *a priori* philosophical considerations, but must be found by an appeal to results of experiments and measurements. A comprehensive definition of reality is, however, unnecessary for our purpose. We shall be satisfied with the following criterion, which we regard as reasonable. *If, without in any way disturbing a system, we can predict with certainty (i.e., with probability equal to unity) the value of a physical quantity, then there exists an element of physical reality corresponding to this physical quantity*. It seems to us that this criterion, while far from exhausting all possible ways of recognizing a physical reality, at least provides us with one

FIGURA 2.27. Il frontespizio del lavoro di Einstein e collaboratori del 1935.

CAPITOLO 3

Processi di Markov

3.1. Il modello di Ehrenfest

Iniziamo a studiare le probabilità in un contesto *dinamico*. Partiamo da un esempio classico, il modello introdotto nel lontano 1907 [EE07] da Tatjana e Paul Ehrenfest ¹. Ci sono due urne che contengono complessivamente N palle, ciascuna identificata da un numero progressivo. Con cadenza regolare si estrae un numero a caso tra uno e N , e la palla che corrisponde al numero estratto passa dall'urna in cui si trova all'altra. Fissata la condizione iniziale (ad es. tutte le palle sono concentrate in una stessa urna), si tratta di determinare la probabilità $p(n, t)$ di trovare nella prima urna n palle all'istante t . Qualitativamente si può stimare che il numero di palle tenderà ad oscillare casualmente intorno alla distribuzione uniforme ($N/2$ in ciascuna urna). Si può pensare al modello come una stilizzazione del caso di un gas contenuto in un recipiente diviso in due parti comunicanti attraverso una piccola apertura. Dalla termodinamica sappiamo che la pressione all'equilibrio è la stessa nei due volumi comunicanti. Ci possiamo aspettare perciò che il vettore di probabilità $p(n, t)$ tenderà per t grande ad essere concentrato intorno al valore $n = N/2$ con una dispersione $\Delta n/N$ percentualmente tanto più piccola quanto più grande è N . Studieremo in dettaglio il modello più avanti, in quanto la trattazione matematica abbastanza complessa distoglierebbe dallo sviluppo dei concetti di base (si veda l'App.A.6). Nel frattempo si può utilizzare, a livello di video-gioco, il programma `ehrenplot.m` che visualizza l'evoluzione del vettore di probabilità secondo la seguente Eq.(37) a partire dalla situazione in cui le particelle sono inizialmente tutte nella stessa urna. Come si può vedere, inizialmente il flusso di particelle è preponderante verso l'altra urna inizialmente vuota, per poi tendere ad equilibrarsi (pressione uniforme). Empiricamente si trova anche $\Delta n = \sqrt{N}/2$.

3.1.1. La distribuzione asintotica nel modello di Ehrenfest. Anche senza procedere all'analisi dettagliata, che esponiamo più avanti in appendice, si può ricavare la distribuzione di equilibrio, raggiunta dopo un grande numero di estrazioni a caso, con un semplice argomento. Immaginiamo di avere un numero N di monete, pari al numero delle palline, recanti sulle due facce "D" e "S" e un numero da 1 a N . Lanciamo tutte le monete e ogni pallina vada a collocarsi nell'urna a destra o a sinistra in dipendenza dell'uscita della sua moneta. È ovvio che il numero di palline che si collocano a sinistra avrà la stessa distribuzione di probabilità del numero di risultati "S" nel lancio di N monete (o di una moneta N volte...). Questa d.d.p. l'abbiamo già calcolata con il metodo della funzione generatrice: la somma di N v.a. di Bernoulli simmetriche (prob=1/2) ha una d.d.p. data dai coefficienti del polinomio $\left(\frac{1+x}{2}\right)^N$, cioè la distribuzione *binomiale*. Questo approccio non rende conto della dinamica del modello, che tende alla distribuzione binomiale attraverso un processo simile a quello di un gas posto in due recipienti comunicanti, ma è in linea con l'approccio statistico in cui si prevede la distribuzione di Gibbs come distribuzione all'equilibrio. La simulazione del modello rende in più conto di alcune caratteristiche di non-equilibrio, ad esempio del tempo di ricorrenza.

3.2. Catene di Markov

Il modello di Ehrenfest è rappresentativo di un'intera classe di modelli, denominati *catene di Markov*. L'aspetto essenziale del modello è costituito dal fatto che per una situazione nota all'istante t il futuro non dipende da quello che è successo prima di t . Il processo è determinato dalla funzione $P(n, t+1|m, t)$ che rappresenta la probabilità di trovare lo stato n al tempo $t+1$ posto che lo stato sia m all'istante precedente.

¹Su questo modello Mark Kac osserva: "It is probably one of the most instructive models in the whole of Physics and although merely an exaple of a finite Markov chain, it is of considerable independent interest" [Kac59].

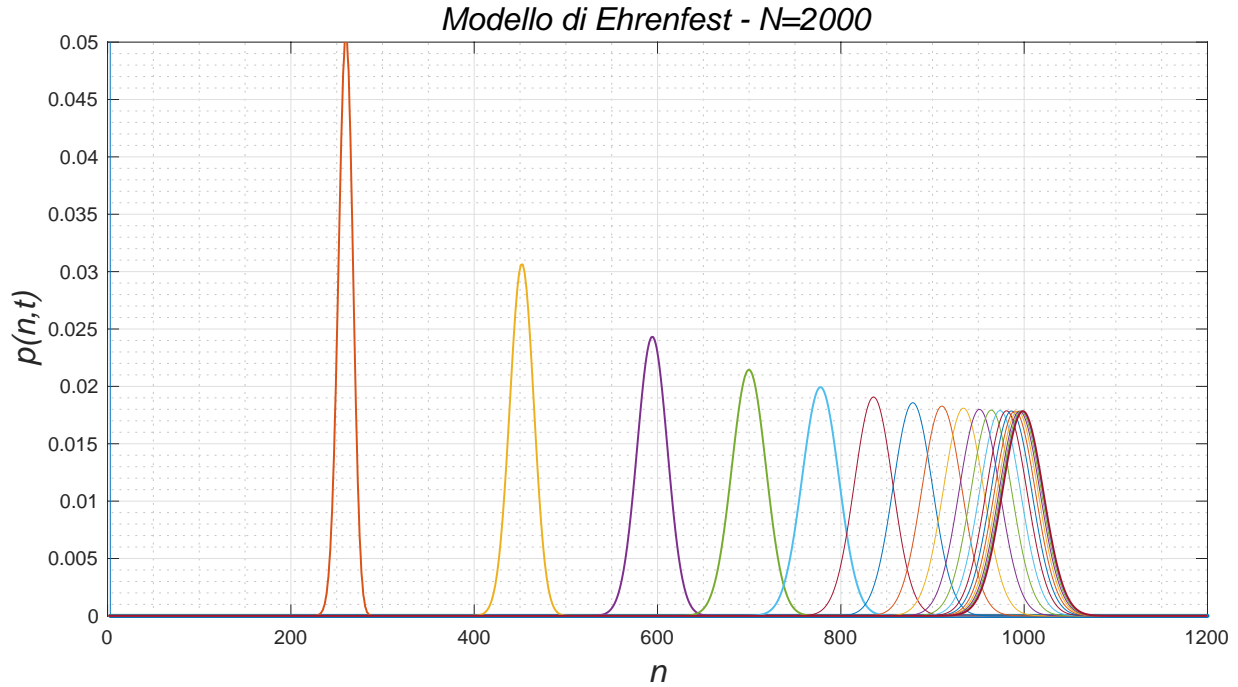


FIGURA 3.1. Evoluzione nel tempo del vettore di probabilità nel modello di Ehrenfest

Questa funzione è detta *probabilità di transizione*. Per il modello di Ehrenfest si trova

$$(34) \quad P(n, t+1|m, t) = \begin{cases} \frac{m}{N} & n = m-1 \ (m \geq 1) \\ 1 - \frac{m}{N} & n = m+1 \ (m < N) \\ 0 & \text{negli altri casi} \end{cases}.$$

Il processo stocastico è poi determinato univocamente specificando lo stato iniziale (ad es. tutte le palle in una singola urna) oppure una distribuzione di probabilità nel caso non si conosca esattamente il numero di palle ma solo la probabilità $p = \{p_n, n = 0, 1, \dots, N\}$ di trovarne n nella prima urna. Si può rappresentare la probabilità di transizione con una matrice ² di dimensione $(N+1) \times (N+1)$, $M_{nm} = P(n, t+1|m, t)$. Per fissare le idee, se $N = 5$ si trova

$$(35) \quad M = \begin{pmatrix} 0 & 1/5 & 0 & 0 & 0 & 0 & (1 \rightarrow 0) \\ 1 & 0 & 2/5 & 0 & 0 & 0 & (0 \rightarrow 1, 2 \rightarrow 1) \\ 0 & 4/5 & 0 & 3/5 & 0 & 0 & (1 \rightarrow 2, 3 \rightarrow 2) \\ 0 & 0 & 3/5 & 0 & 4/5 & 0 & (2 \rightarrow 3, 4 \rightarrow 3) \\ 0 & 0 & 0 & 2/5 & 0 & 1 & (3 \rightarrow 4, 5 \rightarrow 4) \\ 0 & 0 & 0 & 0 & 1/5 & 0 & (4 \rightarrow 5) \end{pmatrix}$$

Si nota che la matrice è caratterizzata dal fatto che ogni elemento di matrice è non-negativo (deve rappresentare una probabilità) e inoltre la somma degli elementi di matrice su ogni colonna vale esattamente uno. Questo vale in generale: ogni matrice che rappresenta la probabilità di transizione di una catena di Markov deve soddisfare le condizioni

$$(36) \quad M_{ij} \geq 0, \quad \sum_i M_{ij} = 1.$$

e viene chiamata **matrice stocastica**. Se consideriamo la matrice trasposta M^T la condizione (36) implica che il vettore con tutte componenti uguali è autovettore con autovalore uno. Dunque $\det(1 - M^T) = 0$, ma

²Per maggiore chiarezza, dove necessario, si potrà utilizzare la notazione $M_{nm} \equiv P(m \rightarrow n)$. Qualche autore preferisce la definizione $M_{mn} \equiv P(m \rightarrow n)$ che implica di prendere la matrice trasposta. Caso per caso bisogna porre attenzione alla convenzione adottata.

il determinante di una matrice è uguale a quello della trasposta, dunque anche M ammette l'autovalore uno. Inoltre si può vedere che tutti gli autovalori di M devono essere localizzati nel disco di raggio uno. Ciò discende dal teorema di Gershgorin [QRF00], riportato in Appendice.

La matrice M permette di calcolare il vettore di probabilità ad ogni tempo t (intero) secondo la semplice formula

$$(37) \quad p(t) = M^t p(0).$$

La prima proprietà matematica del processo $t \rightarrow p$ da richiedere per consistenza è che $p(t)$ sia una corretta distribuzione di probabilità, dunque tutte le componenti del vettore devono mantenersi non-negative e la loro somma deve essere costante e uguale a uno. La matrice di transizione M è tale da assicurare che queste proprietà siano soddisfatte. Infatti gli elementi di matrice di M sono non-negativi e quindi $p(t+1) \geq 0$ se $p(t) \geq 0$. Inoltre

$$\sum_n p(t+1)_n = \sum_n \sum_m M_{nm} p(t)_m = \sum_m \left(\sum_n M_{nm} \right) p(t)_m = \sum_m p(t)_m = 1$$

3.2.1. Analisi spettrale. Ammettiamo ora che M sia *diagonalizzabile*³. Allora vale la decomposizione spettrale

$$M = \sum_{\lambda \in \sigma(M)} \lambda \Pi_\lambda$$

e quindi

$$M^t = \sum_{\lambda \in \sigma(M)} \lambda^t \Pi_\lambda.$$

Per tempi molto grandi gli autovalori in modulo inferiori a uno decadono esponenzialmente e sopravvivono solo quelli in modulo uguali a uno. L'autovalore uno esiste sempre; caso per caso possono esistere altri autovalori $e^{i\phi}$. Si avrà allora

$$p(t) \underset{t \rightarrow \infty}{\sim} \Pi_1 p(0) + \sum_{\phi} e^{i\phi t} \Pi_{e^{i\phi}} p(0).$$

I termini oscillanti si mediano a zero e quindi il processo tende al limite rappresentato dall'autovettore appartenente all'autovalore uno, che rappresenta la *distribuzione stazionaria*. Sotto condizioni più restrittive gli autovalori di M sono interni al cerchio unitario, ad eccezione del singolo autovalore uno, e quindi la convergenza del processo è assicurata indipendentemente dalla media temporale. Queste condizioni sono enunciate nel teorema di *Perron-Frobenius* riportato in appendice (si veda [Var62, Gan59]).

ESERCIZIO 51. Due urne contengono inizialmente N_A particelle di tipo A e N_B di tipo B distribuite a caso. Le particelle sono numerate da 1 a $N = N_A + N_B$ come nel modello di Ehrenfest e si procede nello stesso modo: si estrae un numero a caso e la particella con il numero estratto cambia urna. Come varia la probabilità di avere n_A particelle di tipo A nella prima urna e n_B del secondo tipo?

SOLUZIONE. Si tratta della miscela di due gas non interagenti. La simulazione si realizza con il programma `Ehrenfest2types.m`. ◀

ESERCIZIO 52 (Una variante del modello di Ehrenfest). Nel testo di Rozanov [Roz69] troviamo una variante del modello di due urne. Sono dati due contenitori e N palline bianche e N nere. Le palline vengono disposte a caso nei due contenitori, in modo che ci saranno n palline bianche e $N - n$ nere nel primo contenitore e n palline nere e $N - n$ bianche nel secondo. Il processo consiste nel pescare a caso una pallina dal primo contenitore e una dal secondo e scambiarle. Si chiede: il processo tende all'equilibrio? Se sì quale sarà la distribuzione di probabilità p_j per il numero di palline bianche nel primo contenitore quando il processo ha raggiunto l'equilibrio?

³Quello che è noto in generale sullo spettro di una matrice stocastica irriducibile è che gli autovalori di modulo uno si identificano con le radici dell'unità per un qualche intero r e sono semplici. Per gli autovalori di modulo inferiore a uno non si ha una descrizione così precisa (si veda [Ono84], 2nd ed., App.5.5, disponibile online).

SOLUZIONE. Indichiamo con $p^{(t)}(j)$ la probabilità che si trovino j palline bianche nel primo contenitore dopo t estrazioni. Il numero di palline può aumentare o diminuire di una unità oppure restare invariato. Le rispettive proprietà di transizione sono date da

$$\begin{aligned} P(n \rightarrow n-1) &= \left(\frac{n}{N}\right)^2 \\ P(n \rightarrow n) &= 2\frac{n}{N}\frac{N-n}{N} \\ P(n \rightarrow n+1) &= \left(\frac{N-n}{N}\right)^2 \end{aligned}$$

infatti il primo valore indicato corrisponde all'estrazione di una bianca dal primo contenitore e di una nera dal secondo; il secondo valore corrisponde all'estrazione di una bianca da entrambi i contenitori oppure due nere etc. Il processo è dunque governato dalla matrice stocastica definita da

$$(38) \quad M = \begin{pmatrix} 0 & \varepsilon^2 & 0 & 0 & 0 & \dots & 0 \\ 1 & 2\varepsilon(1-\varepsilon) & (2\varepsilon)^2 & 0 & 0 & \dots & 0 \\ 0 & (1-\varepsilon)^2 & 4\varepsilon(1-2\varepsilon) & (3\varepsilon)^2 & 0 & \dots & 0 \\ 0 & 0 & (1-2\varepsilon)^2 & 6\varepsilon(1-3\varepsilon) & (4\varepsilon)^2 & \dots & 0 \\ 0 & 0 & 0 & (1-3\varepsilon)^2 & 8\varepsilon(1-4\varepsilon) & (5\varepsilon)^2 & \dots \\ \vdots & \vdots & \vdots & \ddots & \ddots & \ddots & \dots \end{pmatrix}$$

dove $\varepsilon = 1/N$. La matrice può essere codificata facilmente in Matlab e si può dunque simulare il processo e studiare l'evoluzione della distribuzione di probabilità. Questo è lasciato come esercizio. Un altro approccio è basato sulla trasformazione della legge di evoluzione del vettore di probabilità in un'equazione differenziale e conseguente ricerca della distribuzione asintotica. Questo è fatto per il modello di Ehrenfest nella Sez.A.6 e a questa derivazione ci si può ispirare per il modello che stiamo studiando. Si ottiene in questo modo la soluzione esatta per la distribuzione asintotica che è data da

$$p_j = \binom{2N}{N}^{-1} \binom{N}{j}^2$$

nota come distribuzione ipergeometrica in quanto la sua funzione generatrice è data dalla funzione ipergeometrica ${}_2F_1(-N, -N, 1, z)$. Nello stesso testo di Rozanov già citato si mostra che la distribuzione è normalizzata correttamente. ◀

ESERCIZIO 53 (Passeggiata aleatoria (random walk)). Consideriamo un reticolo nel piano costituito da tutti i punti con coordinate cartesiane intere, $x = (n_1, n_2)$. Ad ogni istante una particella puntiforme salta a caso dalla sua posizione ad una vicina con uguale probabilità in ognuna delle quattro direzioni. Sia $P_t(n_1, n_2)$ la probabilità di trovare la particella nel punto (n_1, n_2) all'istante t . Si può determinare la funzione \mathbb{P} tenendo conto della relazione di ricorrenza

$$P_{t+1}(n_1, n_2) = \frac{1}{4} (P_t(n_1+1, n_2) + P_t(n_1-1, n_2) + P_t(n_1, n_2+1) + P_t(n_1, n_2-1))$$

Si ha allora

$$\begin{aligned} P_{t+1}(n_1, n_2) - P_t(n_1, n_2) &= \frac{1}{4} (P_t(n_1+1, n_2) + P_t(n_1-1, n_2) + \\ &\quad P_t(n_1, n_2+1) + P_t(n_1, n_2-1) - 4P_t(n_1, n_2)) \end{aligned}$$

e il membro di destra è l'analogo alle differenze finite dell'operatore di Laplace. Se introduciamo unità di lunghezza e di tempo rispetto a cui $t = T/\tau$ e $x = (X, Y)/a$ la formula precedente diventa

$$\begin{aligned} (P_{T+\tau}(X, Y) - P_T(X, Y))/\tau &= \frac{a^2}{4\tau} (P_T(X+a, Y) + P_T(X-a, Y) + \\ &\quad P_T(X, Y+a) + P_T(X, Y-a) - 4P_T(X, Y))/a^2 \end{aligned}$$

Passiamo al limite per $\alpha \rightarrow 0$ e $\tau \rightarrow 0$ in modo però che il rapporto $\alpha^2/\tau = 2D$ si mantenga costante, allora l'equazione converge a

$$\frac{\partial P_T(X, Y)}{\partial T} = \frac{D}{2} \Delta P(X, Y)$$

dove Δ è l'operatore di Laplace. Il modello di random walk rappresenta perciò la discretizzazione del processo di diffusione, ovvero del *moto Browniano* come descritto da Einstein nel 1905 - si veda [Kac46]. Determinare i valori medi $\mathbb{E}[X]$ e $\mathbb{E}[Y]$ in funzione del tempo.

3.2.2. Ritorno al punto di partenza. Consideriamo una catena di Markov su N stati, caratterizzato dalla probabilità di transizione M_{ij} . Si suppone che tutti gli stati possano essere raggiunti prima o poi, cioè che M sia *irriducibile*⁴. Ci si chiede con quale probabilità partendo dallo stato j si tornerà a j dopo un numero qualunque di passi. Inoltre ha interesse sapere quanto tempo in media deve passare perchè ciò avvenga. Sia $f_j(t)$ la probabilità che il processo torni *per la prima volta in* j dopo t passi. Allora si avrà

$$(39) \quad \begin{aligned} (M^t)_{jj} &= f_j(t) + f_j(t-1) M_{jj} + f_j(t-2) (M^2)_{jj} \\ &+ \dots f_j(t-k) (M^k)_{jj} + \dots + f_j(1) (M^{t-1})_{jj} \\ &= \sum_{k=0}^{t-1} f_j(t-k) (M^k)_{jj} \end{aligned}$$

che si giustifica considerando che lo stato può tornare per la prima volta in k passi per poi tornare di nuovo (senza restrizioni) nei $t-k$ passi residui.

L'equazione permette di calcolare f_j dalla conoscenza delle potenze di M , un problema puramente algebrico. Dato che l'equazione è data da una *convoluzione* è facile risolverla introducendo le funzioni generatrici

$$(40) \quad \mathcal{M}_j(s) = \sum_{t=1}^{\infty} (M^t)_{jj} s^t$$

$$(41) \quad \mathcal{F}_j(s) = \sum_{t=1}^{\infty} f_j(t) s^t.$$

Dall'Eq.(39) si trova allora

$$\mathcal{M}_j(s) = \mathcal{F}_j(s) + \mathcal{F}_j(s) \mathcal{M}_j(s)$$

ossia

$$(42) \quad \mathcal{F}_j(s) = \frac{\mathcal{M}_j(s)}{1 + \mathcal{M}_j(s)}.$$

Dalla soluzione possiamo verificare che $\mathcal{F}_j(1)$, ossia la probabilità totale di tornare in j , risulta i) minore di uno se la serie $\mathcal{M}_j(s)$ è convergente per $s \rightarrow 1$, ii) uguale a uno se la stessa serie diverge. Inoltre il tempo medio di ritorno, ossia $\tau_j = \sum_{t=1}^{\infty} t f_j(t)$ risulta dato da $\tau_j = \frac{d\mathcal{F}_j}{ds}|_{s=1}$. Sia u_j la distribuzione stazionaria, soluzione di $Mu = u$. Allora si dimostra che vale il risultato semplicissimo (vedi Appendice A.7)

$$(43) \quad \tau_j = 1/u_j.$$

Nel caso del modello di Ehrenfest, per il quale $u_j = 2^{-N} \binom{N}{j}$, il tempo medio di ritorno è allora 2^N per lo stato in cui tutte le particelle sono inizialmente collocate in una sola urna, mentre risulta approssimativamente $\tau \approx \sqrt{\pi N/2}$ per $j = N/2$ che indica fluttuazioni intorno all'equilibrio dell'ordine di \sqrt{N} .

Se non avessimo a disposizione l'analisi matematica esatta che ci offre la formula per il tempo medio di ritorno dovremmo ricorrere a una simulazione. Il programma `EhrenfestRT.m` che si può trovare su

~/dropbox/MPF2023/Programmi/Matlab/Cap.4/Ehrenfest/

considera un numero grande a piacere di "storie" del processo di Ehrenfest e per un dato stato target misura gli intervalli di tempo tra gli istanti in cui il processo passa per lo stesso stato. Il programma mostra un istogramma della storia del numero di particelle in una delle urne confrontando la statistica con la distribuzione di equilibrio e ritorna il valore medio e la deviazione standard degli intervalli. Il valore di confronto comporta il calcolo del coefficiente binomiale $\binom{N}{k}$ e quindi non si può far crescere N a valori troppo grandi, e in ogni caso non

⁴Questo significa in altre parole che nello spazio degli stati del processo non esistono sottoinsiemi da cui il processo non possa uscire.

potremmo attendere il verificarsi del ritorno al punto di partenza se $N \gtrsim 50$ se non vogliamo stare giorni e giorni davanti al nostro computer. Si osserveranno forti fluttuazioni sul tempo di ritorno. Come esercizio impegnativo si suggerisce di studiare la distribuzione statistica del tempo di ritorno. Al momento ne conosciamo solo il valor medio. Il codice `EhrenfestRT.m` amplia l'analisi statistica del modello e permette di ricavare maggiori informazioni sulla distribuzione dei tempi di ritorno. Si veda anche più avanti a Pag.(139).

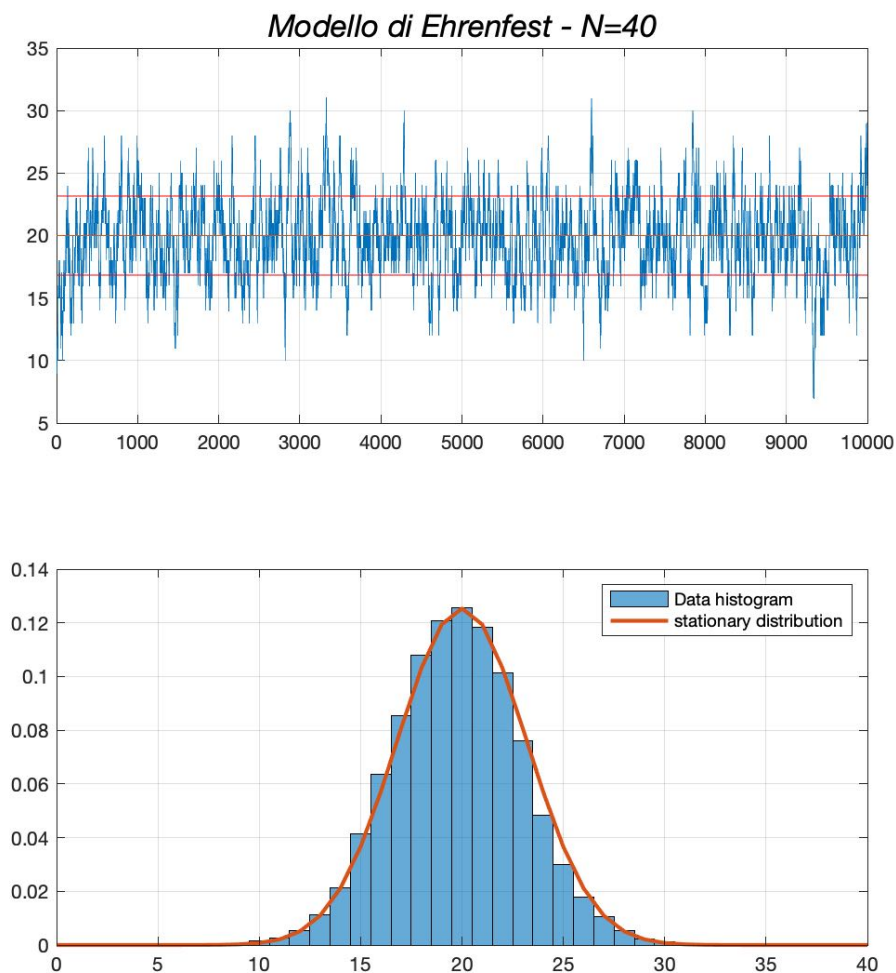


FIGURA 3.2. Simulazione del modello di Ehrenfest per 40 particelle. Da confrontare con il diagramma a pag.74 del testo di Kac.

ESERCIZIO 54. Utilizzando il programma `EhrenfestRT.m` si osserva che i tempi di ritorno presentano una mediana molto inferiore alla media e la “moda”, cioè il valore più frequente è il più piccolo valore possibile (due). Calcolare la probabilità di una transizione $n \rightarrow n+1 \rightarrow n$ e $n \rightarrow n-1 \rightarrow n$ per un dato valore di N (il numero totale di molecole) e spiegare questo fenomeno.

SOLUZIONE. Se si parte da n molecole su N nella prima urna, con probabilità $n/N \times (N-n+1)/N$ il numero passa a $n-1$ per tornare subito a n oppure con probabilità $(N-n)/N \times (n+1)/N$ aumenta a $n+1$ per poi tornare a n . Dunque le traiettorie con ritorno immediato al valore n avvengono con probabilità $(N+2Nn-2n^2)/N^2 = 2x(1-x) + 1/N$ per $x = n/N$. In condizioni di equilibrio la transizione $N/2 \rightarrow N/2$ in due passi avviene più della metà delle volte. ◀

ESEMPIO 6 (*Random walk*). Si consideri il random walk in una sola dimensione. Il sistema comporta *infiniti* stati e quindi l'analisi non può basarsi su quanto detto finora, in particolare i teoremi di Perron e Frobenius non sono applicabili. Dunque è necessario studiare il problema con metodi *ad hoc*. Si ha $M_{nm} = \frac{1}{2}(\delta_{n,m+1} + \delta_{m,n+1})$. Il calcolo della potenza t -esima di M si risolve con la trasformata di Fourier. Infatti se consideriamo il vettore $\phi(k)_n = \exp\{ink\}$ si verifica immediatamente che

$$M_{nm} \phi(k)_m = \cos k \phi(k)_n$$

Un qualunque vettore u_n si può rappresentare attraverso lo sviluppo di Fourier:

$$\begin{aligned} u_n &= \int_{-\pi}^{\pi} c(k) \phi(k)_n \frac{dk}{2\pi} \\ c(k) &= \sum_{n=-\infty}^{\infty} \phi(-k)_n u_n \end{aligned}$$

e nel caso che si parta dal vettore iniziale $u_n = \delta_{n,j}$, per un dato j , si avrà $c(k) = \phi(-k)_j$. Troviamo allora che l'applicazione di M è data semplicemente da

$$(M^t u)_n = \int_{-\pi}^{\pi} \exp\{i k (n - j)\} (\cos k)^t \frac{dk}{2\pi}$$

e quindi per l'elemento $j \rightarrow j$ vale l'espressione

$$(M^t)_{jj} = \int_{-\pi}^{\pi} (\cos k)^t \frac{dk}{2\pi}.$$

La funzione generatrice $\mathcal{M}_j(s)$, dall'Eq.(40), è quindi data da (possiamo eliminare l'indice j in quanto il risultato è indipendente dal sito, grazie alla simmetria di traslazione del reticolo)

$$(44) \quad \mathcal{M}(s) = \int_{-\pi}^{\pi} \frac{s \cos k}{1 - s \cos k} \frac{dk}{2\pi}$$

Il calcolo offre ⁵

$$(45) \quad \mathcal{M}(s) = \frac{1}{\sqrt{1-s^2}} - 1, \quad \mathcal{F}(s) = 1 - \sqrt{1-s^2}.$$

Si ottiene allora che i) il *camminatore* ritorna con probabilità uno al punto di partenza ($\mathcal{F}(1) = 1$) e ii) il tempo medio di ritorno è *infinito* (dato che la derivata di \mathcal{F} in $s = 1$ diverge), un risultato dovuto a G. Polya. *La stessa proprietà vale per il random walk nel piano, mentre in dimensione tre o superiore la probabilità di ritorno risulta inferiore a uno* (si veda [ID89]). Si può infatti vedere facilmente che l'integrale (44) diventa

$$(46) \quad \mathcal{M}(s) = \int_{-\pi}^{\pi} \frac{dk_1}{2\pi} \dots \int_{-\pi}^{\pi} \frac{dk_n}{2\pi} \frac{1}{1 - \frac{s}{n} \sum_j \cos k_i} - 1.$$

e la convergenza o divergenza dell'integrale per k piccolo (in gergo *nell'infrarosso*) si può decidere anche senza calcolarlo esplicitamente. Ponendo $s = 1$, l'integrando presenta un polo del secondo ordine a $k = 0$. Se $n \geq 3$ l'integrale è *finito*, il che implica che il camminatore ha solo una probabilità finita di ritornare al punto di partenza. Il calcolo preciso della probabilità di ritorno passa attraverso il calcolo dell'integrale nell'Eq.(46) che si può trovare in [MI77]. Per un calcolo rapido approssimato si può utilizzare Vegas⁶ o Mathematica e si trova per il valore dell'integrale (in 3 dimensioni) $\mathcal{M}(1) \approx 0.51639$ da confrontare con il risultato esatto (si veda [ID89])

$$P = \sqrt{6}/(32\pi^3) \Gamma(\frac{1}{24}) \Gamma(\frac{5}{24}) \Gamma(\frac{7}{24}) \Gamma(\frac{11}{24}) - 1 = 0.51638605915198...$$

⁵Si veda su [GR65] l'Eq.3.613.1. In alternativa, per i più esperti, si consiglia di utilizzare la tecnica dei residui una volta trasformata la variabile di integrazione con $z \equiv \exp\{ik\}$.

⁶Il codice Vegas `ZuckerWS.c` contiene il calcolo dell'integrale in dimensione 3.

La probabilità cercata è perciò secondo l'Eq.(42), $P = \frac{\mathcal{M}(1)}{1+\mathcal{M}(1)} \approx 0.3405$ (per una probabilità quattro cifre sono fin troppe...)⁷.

ESERCIZIO 55. Random walk su un cerchio. Immaginiamo di restringere gli stati della camminata aleatoria dell'esempio precedente a un numero finito N con la condizione che i primi vicini del nodo N -esimo siano $N-1$ e 1 . Si dice che si impongono condizioni al contorno di periodicità e il modello equivale a un random walk su un cerchio discreto con nodi ad angoli $2j\pi/N$, $j = 0, 1, \dots, N-1$. Calcolare il tempo medio di ritorno su un nodo qualunque.

SOLUZIONE. La matrice di transizione è simile a quella del random walk in una sola dimensione troncata a dimensione $N \times N$. La condizione di periodicità aggiunge un elemento di matrice $1/2$ nella prima riga in posizione $(1, N)$ e nell'ultima in posizione $(N, 1)$, dunque

$$P = \begin{pmatrix} 0 & 1/2 & 0 & 0 & \dots & 0 & 1/2 \\ 1/2 & 0 & 1/2 & 0 & \dots & 0 & 0 \\ 0 & 1/2 & 0 & 1/2 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \dots & \vdots \\ \vdots & \vdots & \dots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \dots & \dots & 1/2 & 0 & 1/2 \\ 1/2 & 0 & \dots & \dots & 0 & 1/2 & 0 \end{pmatrix}$$

Si può procedere in modo analitico dettagliato, determinare lo spettro di P , il proiettore sull'autovalore 1 e così determinare il tempo medio di ritorno. Tuttavia in questo caso c'è una scorciatoia fornita dalla simmetria. Tutti i nodi sono equivalenti trovandosi sui vertici di un poligono regolare di N lati. Quindi all'equilibrio tutti i nodi avranno la stessa probabilità, ossia il vettore che soddisfa $P \mathbf{u}_0 = \mathbf{u}_0$ sarà dato da

$$\mathbf{u}_0 = (1, 1, \dots, 1)/N$$

come è immediato verificare, e il tempo medio di ritorno risulta uguale a N . La via analitica più dettagliata consente di ottenere risultati più fini - tipo qual è la probabilità di trovare il camminatore sul nodo j dopo t passi se è partito da 1 ? Come nel caso dell'esercizio precedente gli autovettori sono ancora dati da

$$\mathbf{u}_j^{(k)} = \exp \{ i k j \} .$$

L'equazione agli autovalori $P \mathbf{u}^{(k)} = \lambda \mathbf{u}^{(k)}$ determina $\lambda = \cos(k)$ come in precedenza, ma la prima e l'ultima riga richiedono di restringere i valori di k in modo che $e^{i k N} = 1$ e quindi $k N = 2\pi(0, 1, 2, \dots, N-1)$. L'autovalore 1 corrisponde a $k = 0$ e l'autovettore è quel vettore \mathbf{u}_0 che abbiamo già individuato.

Nota. Osserviamo che il modello a numero finito N di nodi confluisc, dal punto di vista del risultato, in quello del RW su reticolo infinito (il tempo medio di ritorno diverge per $N \rightarrow \infty$). Inoltre il risultato che discende dalla simmetria è anche in linea con le nostre aspettative per via che il RW costituisce una discretizzazione dell'equazione del calore e sappiamo che all'equilibrio la temperatura deve uniformarsi e tendere a un valore finito in un mezzo conduttore di taglia finita, in questo caso un anello.



ESERCIZIO 56. Determinare il tempo medio di ritorno per la catena di Markov con matrice di transizione

$$P_{ij} = \begin{cases} a & \text{if } i - j \equiv \pm 1 \pmod{N} \\ b & \text{if } i - j \equiv \pm 2 \pmod{N} \\ 0 & \text{altrimenti} \end{cases}$$

con $a + b = 1/2$. Si tratta di un random walk sulle radici N -esime dell'unità come nell'esercizio precedente ma ciascun salto può portare a una distanza di uno o due con probabilità a e b .

⁷Sull'articolo di Zucker [I.J11] si trova un'ampio resoconto sugli sviluppi del calcolo di integrali importanti per i random walk su vari reticoli. Il commento che viene spontaneo è che ci sono voluti decenni per arrivare al prodotto finito in tutta la sua eleganza e tuttavia se oggi ci accontentiamo di un accurato risultato numerico possiamo ottenerlo in meno di un secondo!

SOLUZIONE. La simmetria del problema è la stessa di quella riscontrata in precedenza, la simmetria del poligono regolare di N lati, un gruppo indicato con Z_N . La distribuzione di equilibrio è allora ancora la stessa e il tempo medio di ritorno non cambia ($\tau = N$). ◀

ESEMPIO 7 (Il problema della coda alla biglietteria). Il problema affrontato al Cap. 2.1 senza ancora disporre di strumenti adeguati si può ricondurre al problema del random walk. Il numero di biglietti da 5 euro presenti in cassa varia di ± 1 esattamente come nel random walk in una dimensione. Si tratta di valutare la percentuale di cammini che ritornano all'origine senza mai passare per valori negativi. Si è visto che ad ogni cammino di questo tipo corrisponde un cammino con due passi in più tale da non passare mai per lo zero prima dell'ultimo passo. Quindi possiamo contare i cammini che ci interessano attraverso un problema di “primo ritorno” all'origine. Abbiamo visto che la probabilità di ritornare per la prima volta all'origine dopo t passi ha come funzione generatrice $\mathcal{F}(s)$, Eq.(45)

$$(47) \quad \mathcal{F}(s) = 1 - \sqrt{1 - s^2} = 2 \sum_{n=0}^{\infty} \binom{2n}{n} (s/2)^{2n+2} / (n+1)$$

da cui si deduce che la probabilità di tornare per la prima volta all'origine dopo $2(n+1)$ passi è uguale a $2^{-2n-1} \binom{2n}{n} / (n+1)$. Tenendo conto che il ritorno all'origine può avvenire restando sempre a valori positivi oppure negativi, e che ci sono in totale 2^{2n+2} cammini lunghi $2n+2$, il numero di cammini che si mantengono a valori positivi prima di tornare all'origine è $\binom{2n}{n} / (n+1)$ e questo coincide, per quanto detto nell'Esempio 16, al numero di cammini lunghi $2n$ che non scendono mai sotto lo zero. Si noterà che ci sono diversi modi di impostare il problema della coda: se sappiamo a priori che il numero di biglietti da 5 e da 10 euro sono uguali, allora la probabilità è da conteggiare limitatamente ai cammini che comunque partono e arrivano a zero in $2n$ passi; in questo caso la probabilità di “non-attesa” è appunto $1/(n+1)$. Se invece non abbiamo alcuna informazione sulla distribuzione dei biglietti il numero di cammini possibili è dato da 2^{2n} . Dimostrare che in questo caso la probabilità di non essere costretti ad attendere, per mancanza di resto, diminuisce di un ulteriore fattore $2^{-2n} \binom{2n}{n} \approx 1/\sqrt{\pi n} (1 - \frac{1}{8n} + \frac{1}{128n^2} + O(n^{-3}))$.

ESERCIZIO 57. Il problema della pila di libri. Sul tavolo di una biblioteca c'è costantemente una pila di libri di nuova acquisizione non ammessi al prestito ma solo alla consultazione rapida [Roz69]. I libri sono numerati da 1 a N e per ciascuno si conosce il ranking nella classifica delle nuove pubblicazioni che regola la probabilità che il libro venga consultato. Di regola un libro viene estratto dalla pila e alla fine della consultazione collocato in cima alla pila. Se la probabilità è inversamente proporzionale al ranking si chiede: dopo che un gran numero di frequentatori ha visitato la biblioteca e consultato la pila delle novità qual'è la probabilità di trovare il j -esimo libro in cima alla pila? Stessa domanda per il posto appena sotto alla cima della pila?

SOLUZIONE. Si tratta di impostare la catena di Markov che corrisponde a questa situazione. Un modo che si rivela troppo complicato corrisponde a modellizzare il processo come un random walk sulle permutazioni di N oggetti; se a un certo istante la pila è caratterizzata dalla permutazione $\{b_{i_1}, b_{i_2}, \dots, b_{i_N}\}$ e se si estrae il numero i_k con probabilità p_{i_k} la nuova pila sarà data da $\{b_{i_k}, b_{i'_1}, \dots, b_{i'_{N-1}}\}$ dove la permutazione $\{b_{i'_1}, \dots, b_{i'_{N-1}}\}$ coincide con quella precedente da cui sia stato rimosso b_{i_k} . Il problema di caratterizzare la probabilità dell'intera permutazione che definisce la pila si presenta però alquanto complesso (vedi oltre). Se invece ci si limita a chiedersi qual'è la probabilità di avere un qualunque libro in cima alla pila senza riguardo alla disposizione dei rimanenti, allora il calcolo è banale! Infatti all' n -esimo passo, indipendentemente dall'ordinamento della pila, con probabilità p_k viene portato in cima il k -esimo libro e quindi la probabilità di trovarcelo è ovviamente p_k . Diversa è la considerazione riguardo il secondo, il terzo etc. posto nella pila. A questa domanda si può dare una risposta attraverso una simulazione. Assegnata la probabilità $\{p_k | k = 1 \dots N\}$ l'estrazione di un numero intero a caso si può agevolmente realizzare utilizzando l'algoritmo degli alias. Se $(b_{i_1}, b_{i_2}, \dots, b_{i_N})$ è la permutazione della pila a un certo istante e se si estrae un intero che corrisponde a b_{i_j} allora semplicemente la permutazione diventa quella con b_{i_j} spostato al primo posto. Questo si codifica facilmente in `matlab`, come riportato nel file `bookpile.m`, utilizzando il comando `circshift`. Il programma permette di valutare la probabilità di trovare il libro b_k come r -esimo dall'alto della pila. Dalla simulazione si può congetturare che la probabilità di trovare il k -esimo libro al secondo posto nella pila sia data approssimativamente da $\propto p_k(1 - p_k)$, almeno per N grande.

Torniamo al problema originale e cerchiamo la soluzione senza utilizzare una simulazione. Il problema è costituito dal mappare l'insieme delle permutazioni S_N , detto gruppo simmetrico, su \mathbb{Z} in modo da definire la

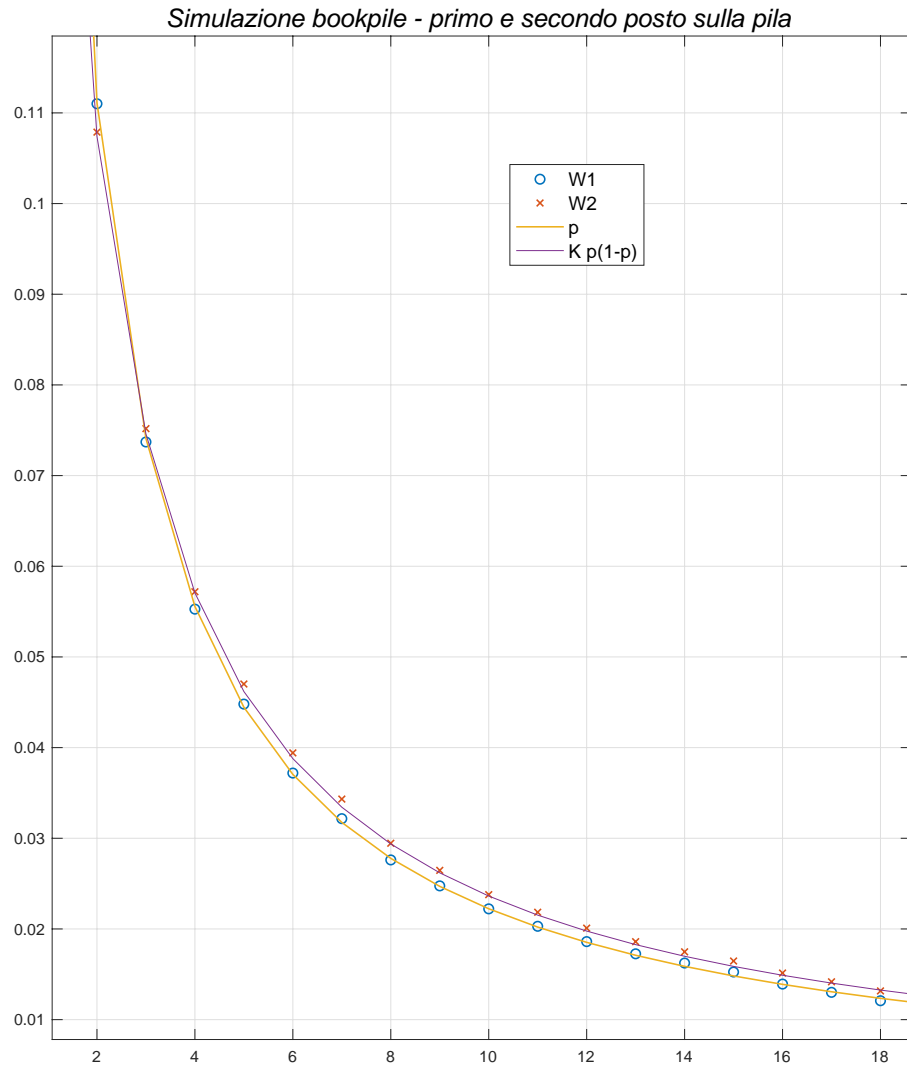


FIGURA 3.3. Un esperimento con una pila di 50 libri.

matrice stocastica che regola il processo. Infatti se estraendo ad esempio 3 la permutazione 12345 viene mappata su 31245 a queste due permutazioni devono essere associati due indici (i, j) della matrice $P_{j \rightarrow i}$. La sorgente più promettente dove trovare l'algoritmo adatto è il testo di Knuth [Knu98]. Infatti al vol.II si trova l'algoritmo di Lehmer che tradotto in `matlab` è come segue:

Kindex.m

```
function f = Kindex(p,t)
% algoritmo di Lehmer (Knuth) per la mappa S_t --> 1:(t!)
% Associa un indice numerico in modo univoco ad ogni permutazione di
% "t" simboli
% L'operazione e' invertibile con "iKindex.m"
% MPF-2020
% P1:
f = 0;
% P2:
while(t>1),
    p = p(1:t);
    [m,s] = max(p);
    f = f*t + s - 1;
% P3:
```

```

    tmp = p(s);
    p(s) = p(t);
    p(t) = tmp;
% P4:
    t = t-1;
end
f=f+1;          % not in Knuth's, but matlab's indices are >0

```

```

function p = iKindex(f,t)
% map 1:(t!) --> S_t
% from Knuth, T.A.C.P. vol.II
% Inverse map of Kindex.m
f = f-1;
p = 1:t;
for r=2:t,
    s = mod(f,r);
    f = floor(f/r);
    tmp = p(r);
    p(r) = p(s+1);
    p(s+1) = tmp;
end

```

Combinando lo stesso metodo utilizzato nella simulazione numerica in cui si sposta il libro b_k , appena scelto con probabilità p_k , al primo posto della pila, le due permutazioni, iniziale e modificata, vengono associate a due indici della matrice attraverso "Kindex" e nel corrispondente elemento di matrice⁸ si inserisce il valore p_k . Ad es. nel caso $N = 3$ la matrice risulta essere

$$M = \begin{bmatrix} p_3 & 0 & 0 & p_3 & 0 & p_3 \\ 0 & p_3 & p_3 & 0 & p_3 & 0 \\ p_2 & p_2 & p_2 & 0 & 0 & 0 \\ p_1 & p_1 & 0 & p_1 & 0 & 0 \\ 0 & 0 & 0 & p_2 & p_2 & p_2 \\ 0 & 0 & p_1 & 0 & p_1 & p_1 \end{bmatrix}$$

e la corrispondenza tra S_3 e $(1:6)$ come data da Kindex.m è la seguente

$$\begin{aligned} (3 \ 1 \ 2) &\rightarrow 1 \\ (3 \ 2 \ 1) &\rightarrow 2 \\ (2 \ 3 \ 1) &\rightarrow 3 \\ (1 \ 3 \ 2) &\rightarrow 4 \\ (2 \ 1 \ 3) &\rightarrow 5 \\ (1 \ 2 \ 3) &\rightarrow 6 \end{aligned}$$

La logica è semplice: si arriva in $(3,1,2)$ se viene estratto il 3 con probabilità p_3 a partire da $(3,1,2), (1,3,2)$ o $(1,2,3)$ e questo fissa la prima riga. Le altre righe sono determinate in modo analogo. È evidente che per $N \geq 4$ la costruzione di M senza l'aiuto del computer diventa un compito piuttosto laborioso. La conoscenza di M ci permette di calcolare esattamente la distribuzione stazionaria come autovettore di M con autovalore 1 e a questo punto si può valutare la probabilità di trovare un determinato libro in cima alla pila o al secondo posto, etc. In realtà il programma così congegnato non può raggiungere valori alti di N dato che la routine `perms` che costruisce l'elenco delle permutazioni si rifiuta di procedere oltre $N = 11$, in quanto per $N = 12$ l'occupazione in memoria raggiungerebbe i 43 GBy solo per l'elenco delle permutazioni. Anche disponendo di un supercomputer e di impostare il problema su memoria distribuita etc. etc. il gioco si fermerebbe poco oltre. Quello che si può fare è di controllare per N piccolo che il risultato esatto venga ben riprodotto dalla simulazione e questa può poi lanciarsi a N grandi, come della figura precedente Fig. 3.3. Per $N = 9$ il programma esatto e la simulazione sono confrontati facilmente e il risultato è riportato in Fig. 3.4.



⁸La matrice stocastica M del processo è opportunamente definita come matrice sparsa: per ognuna delle $N!$ righe sono diversi da zero solo N elementi e perciò l'occupazione di memoria si abbatte di un fattore $(N-1)!$. Già per $N=8$ la matrice P piena richiederebbe 13GB in memoria, mentre la nel caso sia definita sparsa occupa meno di 3 MB!

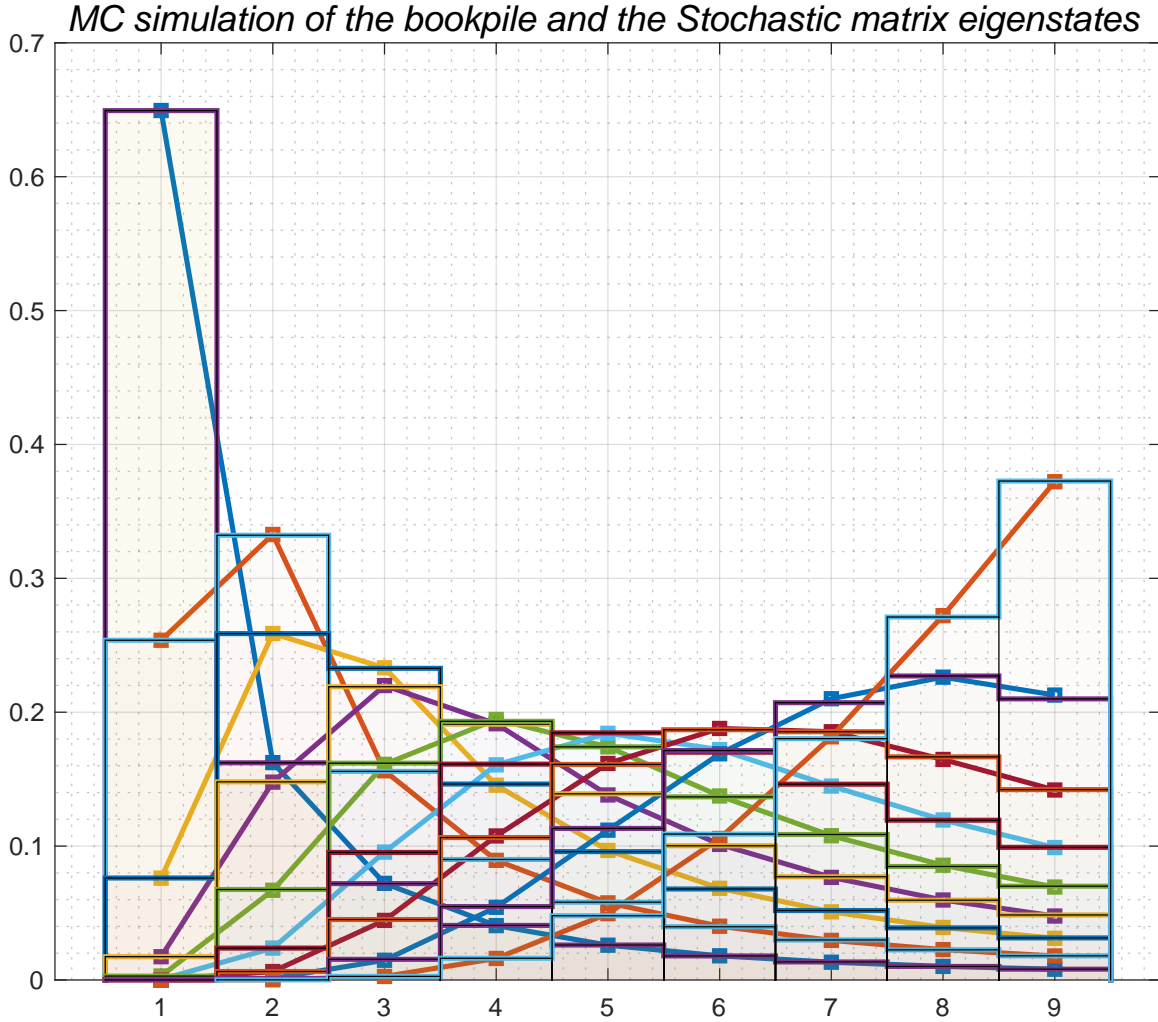


FIGURA 3.4. Confronto tra calcolo esatto (linee continue) e simulazione (barre) per la pila di 9 libri.

3.3. Processi di Markov in tempo continuo

Studiamo ora un'altra varietà di processi aleatori. Sempre limitandoci a sistemi con un numero finito di stati possibili, esistono casi in cui la probabilità $p_n(t)$ dipende da un parametro continuo t . Si pensi ad esempio a un sistema che può trovarsi in un insieme di stati A_1, A_2, \dots, A_N di cui si conosce il ritmo di transizione $\lambda_{i \rightarrow j}$ dall'uno all'altro; la probabilità $p_n(t)$ di trovare il sistema nello stato n al tempo t soddisfa allora l'equazione differenziale

$$(48) \quad \frac{dp_n}{dt} = \sum_m \Lambda_{nm} p_m(t).$$

dove gli elementi di matrice Λ_{nm} per $n \neq m$ rappresentano il ritmo di transizione dallo stato m a quello n mentre gli elementi diagonali Λ_{nn} danno il ritmo di transizione dallo stato n a tutti gli altri. Affinché questa

equazione conservi la probabilità totale uguale a uno, la matrice Λ dovrà soddisfare il vincolo

$$(49) \quad 0 = \sum_n \frac{dp_n}{dt} = \sum_{n,m} \Lambda_{nm} p_m(t).$$

che è automaticamente soddisfatto se $\sum_n \Lambda_{nm} = 0$, cioè se la somma degli elementi di ogni colonna risulta nulla. Si avrà quindi che gli elementi fuori diagonale Λ_{nm} sono positivi in quanto rappresentano una probabilità di transizione nell'unità di tempo, mentre i termini diagonali devono saldare il bilancio della colonna e quindi saranno negativi. Un esempio solubile esattamente è il seguente.

ESERCIZIO 58 (si veda [Gin00]). Si studi il caso di un sistema a tre stati A, B, C la cui funzione di probabilità soddisfa l'equazione di evoluzione (α, β e γ sono costanti positive)

$$\begin{cases} dp_A/dt = -\alpha p_A + \gamma p_C \\ dp_B/dt = \alpha p_A - \beta p_B \\ dp_C/dt = \beta p_B - \gamma p_C \end{cases}$$

e se ne trovi, se esiste, la soluzione asintotica per tempi grandi. Se non si riesce a trovare la soluzione, si esplori il problema per qualche valore particolare dei parametri attraverso una simulazione numerica (e proprio come ultima risorsa, prima di esaminare la soluzione, si veda il codice `aug_problem.m` su `dropbox`).

SOLUZIONE. *Lo spettro della matrice*

$$A = \begin{pmatrix} -\alpha & 0 & \gamma \\ \alpha & -\beta & 0 \\ 0 & \beta & -\gamma \end{pmatrix}$$

è dato dalle radici dell'equazione $(\lambda + \alpha)(\lambda + \beta)(\lambda + \gamma) = \alpha\beta\gamma$, il che implica che una radice è zero e le rimanenti due hanno parte reale negativa. Il processo converge dunque alla distribuzione asintotica

$$p_A = \mathcal{N}\beta\gamma, \quad p_B = \mathcal{N}\gamma\alpha, \quad p_C = \mathcal{N}\alpha\beta$$

che rappresenta l'autovettore $A\mathbf{p} = 0$. \mathcal{N} è fissato dalla normalizzazione a uno della probabilità totale. Per visualizzare l'evoluzione del vettore di probabilità si può utilizzare il codice `aug_problem.m` su `dropbox/matlab/Cap.2/rw`.



3.3.1. Proprietà di convergenza. Si tratta di controllare le proprietà spettrali della matrice Λ ; infatti la soluzione $\mathbf{p}(t) = \exp(t\Lambda)\mathbf{p}(0)$ tenderà a un *unico* limite $\mathbf{p}(\infty)$ se sono soddisfatte due condizioni: i) esiste un unico autovettore corrispondente all'autovalore zero e ii) gli altri autovalori hanno parte reale negativa. Ora, la condizione che la somma sulle colonne sia nulla implica immediatamente che lo zero appartiene allo spettro (il vettore $(1, 1, \dots, 1)$ è autovettore sinistro), ma in generale esisteranno più autovettori, a meno che Λ non soddisfi l'ulteriore proprietà di essere *irriducibile* (in modo analogo alle matrici stocastiche del Cap. 3). Per la seconda condizione, si può concludere del tutto in generale, grazie al teorema di Gershgorin (vedi A.5) che lo spettro è contenuto nell'unione dei cerchi

$$\mathcal{C}_i = \{z \in \mathbb{C} \mid |z - \lambda_{ii}| \leq \sum_{j \neq i} |\lambda_{ji}| = |\lambda_{ii}|\}$$

Dunque ognuno dei cerchi \mathcal{C}_i ha centro sull'asse reale negativo e passa per l'origine. In conclusione, detto \bar{h} il massimo valore assoluto degli elementi sulla diagonale di Λ , lo spettro è compreso nel cerchio di raggio \bar{h} centrato in $-\bar{h}$. Se consideriamo l'evoluzione infinitesimale

$$\mathbf{p}(t + \delta t) \sim \mathbf{p}(t) + \delta t \Lambda \mathbf{p}(t)$$

questa è equivalente a una catena di Markov con matrice di transizione

$$M = 1 + \delta t \Lambda$$

a patto di scegliere un passo δt sufficientemente piccolo, e precisamente $\delta t \leq 1/\bar{h}$. Il problema è quindi ricondotto a quello di catene discrete. Nel caso irriducibile, la distribuzione asintotica è perciò sempre data dalla soluzione di $\Lambda \mathbf{p} = 0$.

Si noti che, viceversa, ogni catena di Markov discreta si può trasformare in un processo continuo. Per ogni salto della catena si introduce un certo numero B di pause in cui il processo si arresta e la decisione di arresto sia presa con probabilità $B/(B+1)$: ne segue il processo governato da

$$p_{t+1} = (1-r)p_t + rMp_t$$

avendo introdotto il “parametro di rilassamento” $r = 1/(B+1)$. È chiaro che se facciamo crescere B l’equazione tenderà a rappresentare un’evoluzione continua governata dall’equazione

$$\frac{dp_t}{dt} = (M-1)p_t$$

Si vede che la matrice $M-1$ è una matrice a elementi positivi fuori diagonale e soddisfa la condizione che la somma su ogni colonna si annulla. È quindi del tipo considerato nel paragrafo precedente.

ESERCIZIO 59 (Problema di Ehrenfest continuo). Si prenda il problema delle due urne del Cap. 3, con una variante: l’estrazione a caso non avviene con regolarità ogni τ secondi, piuttosto l’estrazione avviene con probabilità λdt nell’intervallo dt . Siamo dunque nel caso prefigurato, con un parametro di rilassamento

$$p(t+dt) = (1-\lambda dt)p(t) + \lambda dt M_E p(t)$$

dove abbiamo indicato con M_E la matrice stocastica del modello discreto delle due urne. Si ha dunque un processo di Markov in tempo continuo governato dalla matrice $\Lambda = M_E - 1$. Dal risultato dell’App. A-A.6 si può verificare che lo spettro di λ_E è contenuto nel cerchio di raggio uno centrato in -1 .

ESERCIZIO 60 (“Six crosses in a row”). Sul libro di Baldi [P.B98] si trova il quesito: se lanciamo cento volte una moneta qual’è la probabilità che escano sei “croci” di fila? Dobbiamo cioè valutare la probabilità che in una stringa casuale “0110010100101...” lunga cento bits si trovi la sotto-stringa ‘000000’.

SOLUZIONE. Abbiamo già incontrato un problema analogo quando si è parlato del teorema di Borel–Cantelli e sappiamo che per una stringa lunga L bits con L molto grande dobbiamo aspettarci di trovare $\approx L/2^6$ volte la stringa di sei bits uguali⁹. Qui però dobbiamo valutare attentamente il caso di una stringa lunga solo 100 e se è vero che dobbiamo aspettarci un numero $O(1)$ di riscontri non sappiamo ancora il valore esatto. Possiamo procedere (almeno) in due modi: utilizziamo una simulazione oppure applichiamo la teoria delle catene di Markov come mostrato sul libro di Baldi [P.B98]. Una piccola modifica del programma BCbin.m che si trova in matlab/Cap.2/BorelCantelli/HT.m permette di ricavare rapidamente la probabilità $P(\text{“sei croci consecutive”}) \sim .55 \pm 0.005$. Il codice matlab si trova su dropbox ~/MPF2023/Programmi/Matlab/Cap.2/Baldi/HN.m. L’approccio che utilizza una catena di Markov, e che si trova relizzato nello stesso codice matlab, è il seguente. Consideriamo lo spazio degli stati costituito dalle sette stringhe

$$\Omega = \{ 'T', 'TC', 'TCC', 'TCCC', 'TCCCC', 'TCCCCC', 'TCCCCC' \}.$$

Nel corso dei cento lanci stabiliamo che ci troviamo nello stato ‘T’ se l’ultimo tiro ha dato testa, nello stato ‘TC’ se l’ultimo tiro ha dato croce dopo una testa, e così via. Una volta raggiunto lo stato ‘TCCCCC’, che per semplicità conviene indicare con notazione “chimica” TC_6 , non c’è bisogno di continuare essendo raggiunto l’obiettivo delle sei croci consecutive; ciò è codificato dicendo che lo stato “ TC_6 ” non ammette transizioni ad altri stati e quindi il processo è descritto da una matrice di transizione che nell’ultima colonna ha un 1 in posizione 77:

$$M = \begin{pmatrix} 1/2 & 1/2 & 1/2 & 1/2 & 1/2 & 1/2 & 0 \\ 1/2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1/2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1/2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1/2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1/2 & 1 \end{pmatrix}$$

⁹Si racconta che Richard Feynman fosse in grado di ricordare un gran numero di cifre di π . Alla 762-esima cifra si trova una serie di sei 9 consecutivi: è detto il “Feynman poit”. La stima precedente suggerirebbe un’attesa di una tale serie ogni milione di cifre, se π fosse una successione di numeri a caso. Ma in ogni caso argomenti alla Borel–Cantelli non permettono di prevedere con sicurezza il punto esatto in cui si troverà una successione di cifre prescelta.

La sesta colonna, ad esempio, si interpreta dicendo che da TC_5 si può passare a TC_6 o a T con uguale probabilità, e così per le altre colonne: o si ritorna da capo allo stato 1 o si avanza verso lo stato successivo. Se poi si raggiungono 6 croci (TC_6) lì si rimane, in quanto l'obiettivo è stato centrato. Ora, al primo lancio avremo ovviamente T o C con la stessa probabilità, dunque il vettore di probabilità è dato da $(1/2, 1/2, 0, \dots, 0)$ e a questo dobbiamo applicare M per 99 volte fino a ottenere la probabilità cercata che risulta 0.5460936....

Uno dei vantaggi di una simulazione numerica consiste nel valutare abbastanza facilmente le fluttuazioni di un processo. In un test con 500 runs in ognuno dei quali si simulano 100 tiri della moneta per 10^5 volte si ottiene il risultato mostrato nell'istogramma che segue.

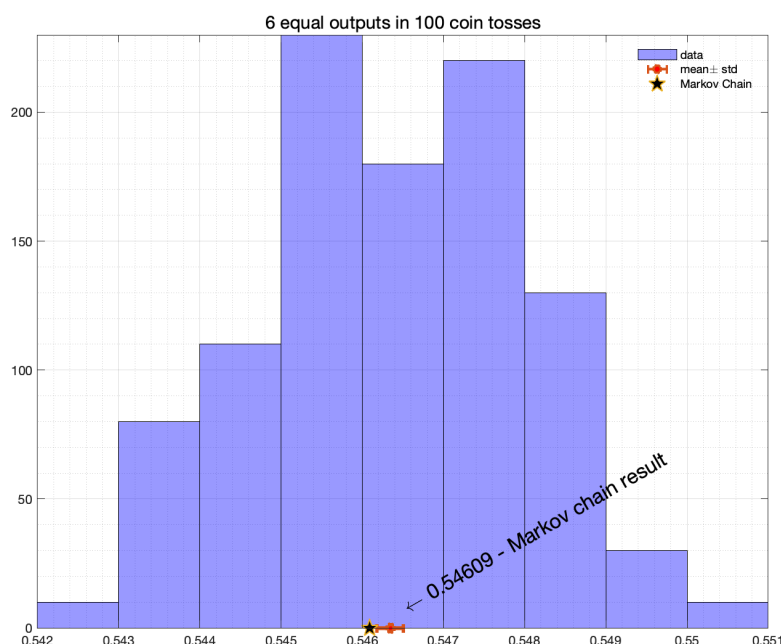


FIGURA 3.5. Sei Croci consecutive - probabilità $\sim .54609$ calcolata con la simulazione MC.

Nella figura successiva, lo stesso calcolo per il caso di otto uscite consecutive su un totale di 400 tiri della moneta. Le probabilità nei due casi differiscono di poco. ◀

ESERCIZIO 61. Varianti sul problema precedente:

- (i) calcolare la probabilità di ottenere una successione di n croci consecutive nel lancio di una moneta per 2^n volte.
- (ii) calcolare la probabilità di ottenere una successione di 4 sei consecutivi nel lancio di un dado per N volte.

ESERCIZIO 62. Si consideri di nuovo il problema di calcolare la probabilità $p(n, N)$ di ottenere almeno una successione di n 'teste' lanciando una moneta per N volte. Si dimostri che la funzione generatrice $G(x, n) = \sum_{j=0}^{\infty} p(n, j) x^j$ è data dall'espressione

$$G(x, n) = \frac{(\frac{1}{2}x - 1)(\frac{1}{2}x)^n}{(x - 1)(1 - x + (\frac{1}{2}x)^{n+1})}$$

Ad esempio nel caso della serie di 6 croci in 100 lanci, la probabilità è data dal coefficiente di x^{100} di $G(x, 6)$ e vale precisamente

$$692255904222999797557597756576 \times 2^{-100} \approx 0.546094...$$

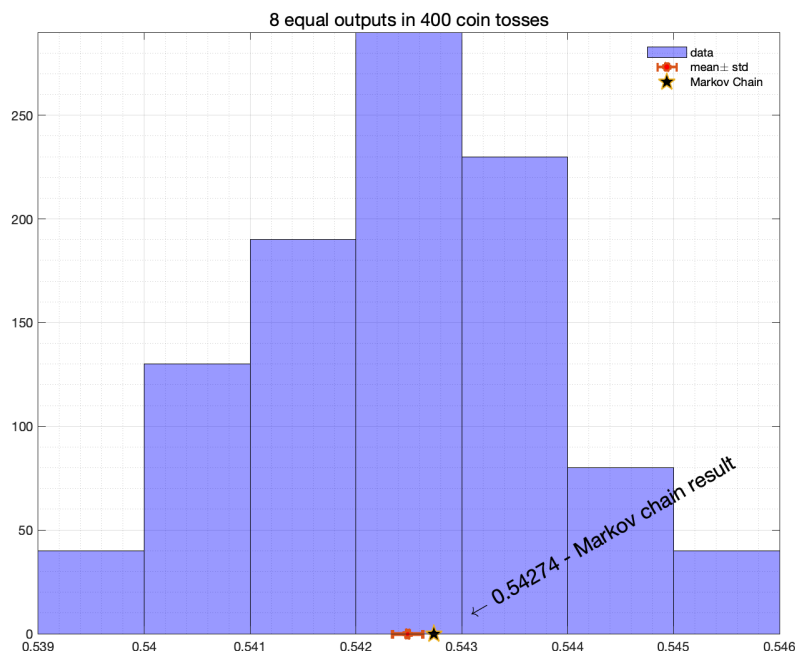


FIGURA 3.6. Otto Croci consecutive in 400 lanci della moneta- probabilità $\sim .5427\ldots$ calcolata con la simulazione MC.

Naturalmente sia il calcolo esatto della potenza 100 di una matrice 7×7 sia il calcolo dello sviluppo di una funzione per ordini molto alti non sono fattibili in un tempo ragionevole se non utilizzando un programma simbolico. Su dropbox/MPF2023/Programmi, sub-folder */Mathematica* si troverà il programma *CoinTosses.nb* per essere utilizzato con Mathematica.

SOLUZIONE. La funzione generatrice è data dall'elemento di matrice $(n+1, 1)$ della serie

$$\sum_N z^N M(n)^N \equiv (1 - zM(n))^{-1}$$

dunque si può calcolare invertendo una matrice di dimensione $n+1$. Per $n=6$ è perfino possibile eseguire il calcolo senza utilizzare il calcolatore... Per curiosità, il risultato dell'enunciato si trova su oeis.org; è sufficiente digitare i primi elementi della successione

$$1, 3, 8, 20, 48, 112, 256, 575, 1275, 2798, \dots$$

che corrisponde ai primi termini dello sviluppo di Taylor per $n=6$ nella variabile $z/2$ per essere indirizzati a <http://oeis.org/A143662>. Le probabilità si ottengono dividendo per la potenza di 2 che è stata introdotta per rendere i coefficienti interi ed essere così accettati da OEIS! ◀

ESERCIZIO 63 (“Six crosses OR six heads in a row”). Una variante del problema, già affrontato a pag.94, secondo un suggerimento di Claudio Destri¹⁰. In cento lanci di moneta qual'è la probabilità di trovare sei teste consecutive oppure sei croci consecutive? Il problema si può affrontare attraverso una simulazione con minime modifiche dal programma *HNnew.m*. È sufficiente cercare entrambe le stringhe di caratteri *CCCCC* e *TTTTT* e registrare il numero di casi in cui la ricerca ha successo almeno per C o T. Naturalmente la probabilità NON coincide con il doppio della probabilità di trovare sei teste (troveremmo un valore maggiore di uno!) ma bisogna applicare la formula $\mathcal{P}(A \cup B) = \mathcal{P}(A) + \mathcal{P}(B) - \mathcal{P}(A \cap B)$, tuttavia il calcolo a priori della probabilità di trovare nella stessa serie di lanci sei teste e sei croci non sembra così agevole. La simulazione porta a un valore di poco superiore a .8, abbastanza alto comunque per convincerci che trovare una serie di uscite uguali non può

¹⁰Claudio Destri, professore di Fisica Teorica all'Università di Milano Bicocca

insospettirci sulla onestà della moneta. Anche per questo problema si può impostare il calcolo attraverso una catena di Markov. Si tratta di trovare lo spazio degli stati in analogia con quanto fatto in precedenza. Una possibilità (non l'unica) è quella di introdurre l'insieme di stati $\Omega = \{C, T, C_2, T_2, C_3, T_3, C_4, T_4, C_5, T_5, C_6, T_6\}$, definendo per semplicità $C_n = CC \dots CC$ (n volte). La convenzione è che da una stringa C_n si passa a C_{n+1} se esce C oppure si passa a T nel caso contrario. La matrice di transizione della catena sarà allora

$$M = \begin{pmatrix} 0 & 1/2 & 0 & 1/2 & 0 & 1/2 & 0 & 1/2 & 0 & 1/2 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 & 1/2 & 0 & 1/2 & 0 & 1/2 & 0 & 0 & 0 \\ 1/2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1/2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1/2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1/2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1/2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1/2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1/2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1/2 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1/2 & 0 & 1 \end{pmatrix}$$

Le righe e le colonne sono indicizzate da Ω . Le ultime due righe corrispondono ad avere raggiunto 6 croci e 6 teste rispettivamente. Applicando M al vettore colonna $u = (1/2, 1/2, 0, 0, \dots, 0)'$ per 99 volte si troveranno negli elementi 11 e 12 del vettore risultante le probabilità di avere raggiunto sei croci o 6 teste, dunque la somma rappresenta la probabilità cercata. Si veda `HN.m` che ammette variazioni sul numero di lanci e sul numero di risultati uguali consecutivi.

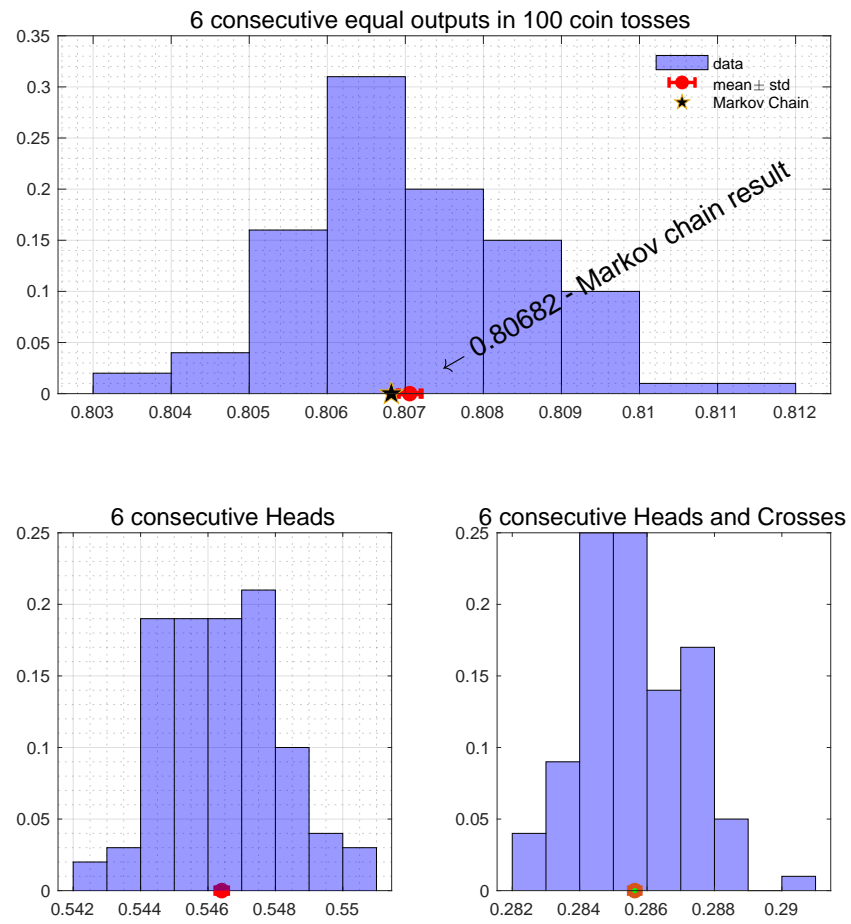


FIGURA 3.7. Es.63: “Six crosses and six heads in a row”.

CAPITOLO 4

Il metodo Monte Carlo

Everything should be made as simple as possible, but not simpler

Albert Einstein

Montecarlo is an extremely bad method; it should be used only when all alternative methods are worse

A. D. Sokal, Cargese Lectures 1996.

Con il termine Montecarlo si intende una vasta categoria di tecniche e algoritmi che permettono di simulare numericamente un processo fisico (totalmente o parzialmente) governato dal caso. L'interesse può risiedere nel processo stesso (ad es. si vuole studiare la dinamica dei neutroni in un reattore) oppure il processo in sè è fittizio, non rappresenta cioè una situazione fisica reale, ma il risultato dell'analisi statistica coincide con una qualche grandezza collegata, questa sì, alla fisica. Ad es. lo studio di un sistema di meccanica statistica in *quattro* dimensioni, pur non corrispondendo a un sistema fisico reale, permette di calcolare grandezze che hanno un'interpretazione in termini di masse o ampiezze di transizione in teorie di campo quantistiche. Le tecniche Montecarlo nascono con i primi calcolatori elettronici (vedi [Sch17], cap.21), un nome che ricorre molto di frequente è quello di un collaboratore di **Enrico Fermi** e **John Von Neumann**: *N. Metropolis* (si veda a proposito anche [Mal03]). Si deve a Mark Kac l'intuizione secondo cui lo studio di processi stocastici attraverso la simulazione numerica può fornire metodi di calcolo per la meccanica quantistica.

Un problema tipico che si presenta è quello di generare successioni di stati individuati da coordinate x_1, x_2, \dots, x_n e che sono distribuiti a caso secondo una probabilità

$$P(x \in \mathcal{S}) = \frac{1}{Z(\beta)} \int_{x \in \mathcal{S}} e^{-\beta H[x]} [dx].$$

Qui $Z(\beta)$, detta la *funzione di partizione*, assicura la normalizzazione della distribuzione, cioè $Z(\beta) = \int e^{-\beta H[x]} [dx]$, l'integrale essendo esteso a tutto lo spazio delle fasi. Questo problema è tipico per la meccanica statistica (classica). L'approccio Montecarlo rappresenta un rovesciamento completo della logica che sta alla base della meccanica statistica alla Boltzmann-Gibbs: la strategia di meccanica statistica traduce il problema dinamico troppo complesso per essere studiato analiticamente per via del numero enorme di gradi di libertà (ma bastano tre corpi per impedire una soluzione analitica delle equazioni di Newton) nel calcolo di un integrale nello spazio delle fasi

$$(50) \quad \langle f(q, p) \rangle = \frac{1}{Z(\beta)} \int d^n q d^n p f(q, p) e^{-\beta H(q, p)}.$$

Secondo l'ipotesi "ergodica" questa media coincide con quella che sarebbe la media temporale

$$(51) \quad \overline{f(q, p)} = \lim_{T \rightarrow \infty} \frac{1}{T} \int f(q(t), p(t)) dt$$

calcolata lungo l'evoluzione temporale del sistema meccanico. La strategia dell'approccio Montecarlo al contrario riconduce il calcolo dell'integrale dell'Eq.(50) alla media temporale dell'Eq.(51) dove però (e qui sta l'idea vincente) non si è obbligati a risolvere per via numerica le equazioni di Newton, *ma si può liberamente scegliere un qualunque processo dinamico caratterizzato unicamente dall'ammettere la distribuzione di Gibbs come distribuzione di probabilità stazionaria*, quella che viene raggiunta lasciando evolvere il sistema per tempi lunghi - dunque si è liberi di costruire un processo dinamico anche di tipo Markoffiano e la scelta si orienterà verso il

processo che converge più rapidamente alla distribuzione stazionaria. Si tratta quindi di generare una successione di stati che risultino distribuiti secondo la probabilità fissata dalla distribuzione di Gibbs. Ciò permette di valutare le varie medie statistiche (energia media, magnetizzazione, etc.). Lo stesso schema è applicabile in fisica quantistica (si veda [Reb83]) nella cosiddetta “formulazione Euclidea”, di cui vedremo un primo esempio più avanti. Data l’importanza di disporre di sistemi di calcolo affidabili che permettano di risolvere problemi di questo genere (ad es. in teoria della *interazioni forti*) si sono sviluppate varie tecniche che vanno sotto il nome di *Metropolis*, *Heat-bath*, *Langevin*, *Hybrid-MC*. Questi argomenti interessano più che altro lo specialista. In queste lezioni ci limitiamo a illustrare alcuni aspetti della problematica. Si consiglia di consultare le lezioni di A. Pelissetto [Pel94] e di Sokal [Sok97] per approfondire il tema.

4.1. Come realizzare successioni casuali

Il problema ridotto all’osso è il seguente. Data una lista di eventi $\Omega = \{\varepsilon_1, \varepsilon_2, \dots, \varepsilon_N\}$ con probabilità p_1, p_2, \dots, p_N , come si fa a generare una successione di eventi che siano *mutuamente indipendenti* e distribuiti secondo la probabilità $\mathbb{P}(\varepsilon_i) = p_i$? L’ingrediente essenziale per ogni algoritmo di simulazione è costituito da un generatore di numeri (pseudo)-random distribuiti uniformemente nell’intervallo $[0, 1)$. Ammetteremo di disporre di questo ingrediente, presente in qualunque libreria matematica, con l’avvertenza che non tutti i generatori hanno le stesse qualità. Non esistono in realtà generatori *perfetti*, che soddisfino cioè alla condizione di uniformità e indipendenza statistica. Si veda l’Appendice. Ogni generatore (ad es. `rand()`, `drand48()` in C o analoghi in Fortran, `rand()` in Matlab, etc.) produce una successione $r_1, r_2, \dots, r_n, \dots$ che innanzitutto è data da numeri in virgola mobile a precisione finita (usualmente 64 bit ossia 15 cifre) che quindi non possono riprodurre se non un sottoinsieme dei reali. Ma la questione più grave è che, inevitabilmente, i numeri generati sono *correlati*. Molta cura si è posta nel disegno di algoritmi che diminuiscano al di sotto di limiti accettabili questa correlazione. Si veda [Knu98], Vol.II. Ammettiamo dunque di avere un buon generatore¹ `rand()`, con le precauzioni del caso. Si tratta ora di trasformare la successione $\{r_i | i = 1, 2, \dots\}$ nella successione di eventi

$$r_1, r_2, \dots, r_n, \dots \rightarrow \varepsilon_{i_1}, \varepsilon_{i_2}, \dots, \varepsilon_{i_k}, \dots$$

in modo da realizzare un campionamento (una simulazione) di (Ω, \mathbb{P}) . In genere serviranno *molte* estrazioni di numeri casuali per costruire un esemplare in Ω (si pensi ad es. a una distribuzione di punti in \mathcal{R}^d , serviranno allora almeno d chiamate a `rand()`).

La prima idea è quella (poco efficiente) di organizzare una cascata di diramazioni del tipo

```
r = rand;
if r < p_1
    return 1
else if r < p_1+p_2
    return 2
else if r < p_1+p_2+p_3
    return 3
.....
end if
```

A parte la goffaggine dello schema, appare evidente che il metodo risulta poco efficiente. Sarà necessario in media effettuare $\mathcal{O}(N)$ controlli logici per estrarre un singolo risultato e a poco giova mettere gli eventi in ordine decrescente di probabilità. Una formulazione più compatta è la seguente, ma è poco più di un miglioramento estetico:

```
q = cumsum(p)-p; % q=[0,p_1,p_1+p_2,...,p_1+p_2+...+p_{N-1}]
x = zeros(Ns,1); % Ns e' il numero di samples estratti
for j = 1:Ns
    r = rand;
    x(j) = length(find(q<r));
end
```

¹Una scelta molto popolare è quella del “Mersenne twister” già citato nell’introduzione.

La logica del programma è evidente: se il numero estratto a caso r è tale che solo la prima componente di q è minore di r allora x vale 1 e ciò avviene con probabilità p_1 ; se le prime due componenti sono minori di r allora $x = 2$ e ciò avviene con probabilità p_2 in quanto $p_1 < r < p_1 + p_2$, etc. La semplicità dell’istruzione `find(q < r)` non deve trarre in inganno - ci sono in realtà N operazioni di confronto e il programma è relativamente lento, anche se con un moderno processore il tempo di esecuzione per $N_s = 10^6$, $N = 100$ è dell’ordine di 2 secondi.

Questo metodo può risultare utile nel caso di distribuzioni di probabilità continue (si vedano esempi sulle lezioni di Pelissetto [Pel94] Sec.2). L’idea è la seguente: sia $\Phi(x) = \mathbb{P}(X < x)$ la funzione di distribuzione per la v.a. X . Φ è una funzione monotona crescente, dunque esiste la funzione inversa $\Phi^{-1}(u)$. Se U è estratto a caso uniformemente nell’intervallo $[0, 1)$ allora $\Phi^{-1}(U)$ risulta distribuito secondo la f.d.d. Φ . Infatti

$$\mathbb{P}(X < x) = \mathbb{P}(\Phi^{-1}(U) < x) = \mathbb{P}(U < \Phi(x)) = \Phi(x)$$

La possibilità di invertire esplicitamente per via analitica la f.d.d. si presenta tuttavia raramente. Alcuni esempi si trovano sulle lezioni di Pelissetto.

ESERCIZIO 64. Realizzare un algoritmo per estrarre numeri random distribuiti secondo la funzione di distribuzione $\Phi(x) = \frac{1}{2}(1 + \tanh(x))$.

ESERCIZIO 65. Realizzare un algoritmo per estrarre numeri random distribuiti secondo la funzione di distribuzione $\Phi(x) = 1/2 + \arctan(x)/\pi$.

ESERCIZIO 66. Realizzare un algoritmo per estrarre numeri random distribuiti secondo la funzione di distribuzione $\Phi(x) = \exp(-x)$, $x > 0$.

ESERCIZIO 67. Si consideri una variabile aleatoria che ha valori nel piano di coordinate (x, y) . Entrambe le coordinate hanno una distribuzione normale. Introducendo coordinate polari nel piano mostrare come si possa costruire un algoritmo per generare esemplari (x, y) secondo la distribuzione normale sfruttando il codice dell’esercizio precedente.

Alias method

4.2. Il metodo degli “alias”

Descriviamo ora un algoritmo che permette di estrarre un esemplare a caso con un numero di operazioni $O(1)$, il che abbassa drasticamente il tempo di esecuzione (nell’esempio in esame il tempo di calcolo può essere inferiore per due ordini di grandezza).

ESEMPIO 8 (*Il problema dei portalettere*). All’ufficio postale si è accumulato un gran numero L di lettere indirizzate a N diversi destinatari, con $L/N = \ell$. Il Direttore per affrontare l’emergenza assume tante persone per portare a N il numero di portalettere, pensando così di risolvere facilmente il problema dello smaltimento. A complicare le cose tuttavia succede, com’è naturale, che le lettere non sono equamente distribuite: ℓ_1 sono indirizzate a \mathcal{I}_1 , ℓ_2 a \mathcal{I}_2 , etc., con $\sum_j \ell_j = L$. Inoltre per accordi sindacali ogni portalettere deve al massimo visitare **due indirizzi** e la quantità di lettere deve essere rigorosamente la stessa per tutti. Come fa il direttore ad assegnare le buste rispettando i vincoli sindacali?

SOLUZIONE. Il metodo consiste nel suddividere ricorsivamente le buste secondo un principio semplice: organizziamo gli indirizzi in ordine crescente di volume di posta, cioè ammettiamo che $\ell_1 \leq \ell_2 \leq \dots \leq \ell_N$. Assegniamo ℓ_1 buste al portalettere Π_1 . Se fosse $\ell_1 = \ell = L/N$ saremmo nel caso banale, in cui tutti gli indirizzi hanno lo stesso numero di buste - ma abbiamo escluso questo caso in quanto non pone problema di sorta. Dunque abbiamo $\ell_1 < \ell$. Allora Π_1 può prendersi anche ulteriori $\ell - \ell_1$ buste, e la scelta giusta è quella di prenderle con l’indirizzo \mathcal{I}_N , quello con il numero maggiore di buste. A questo punto Π_1 è al completo e può partire verso i destinatari \mathcal{I}_1 (con ℓ_1 lettere) e \mathcal{I}_N (con $\ell - \ell_1$). Restano $N - 1$ portalettere e $(N - 1)\ell$ buste recanti $(N - 1)$ indirizzi (perché il primo indirizzo è ormai completamente soddisfatto). Ma allora, una volta rimessi in ordine gli indirizzi secondo ℓ_k crescenti (si tratterà di mettere al posto giusto $\ell'_N \equiv \ell_N - (\ell - \ell_1)$), ci ritroviamo al problema iniziale con N diminuito di uno. Si applica la stessa strategia fino ad esaurire gli indirizzi (e i portalettere). ◀

ESEMPIO 9. Un’altra versione del problema dei portalettere è la seguente ([Knu98], sec.3.4.1 Ex.7). Ci sono N cubi di k colori diversi, n_1 di colore C_1 , n_2 di colore C_2 , etc. fino a n_k di colore C_k . Ci sono anche k scatole identiche, in ciascuna delle quali possono stare esattamente n cubi, con $n_1 + n_2 + \dots + n_k = nk$.

Quindi tutti i cubi si trovano posto nelle scatole senza lasciare vuoti. Tuttavia si deve fare in modo che ogni scatola contenga cubi di non più di due colori diversi. La corrispondenza ovvia è “cubi” \leftrightarrow “lettere”, “colori” \leftrightarrow “indirizzi” e infine “scatole” \leftrightarrow “portalettere”.

È evidente che l’idea si può tradurre in un metodo per estrarre efficientemente un valore $i \in \{1, 2, \dots, N\}$ con probabilità $p_1 \leq p_2 \leq \dots \leq p_N$. Il problema si mappa su quello dei portalettere con l’equivalenza $p_i = \ell_i/L$ e gli indirizzi I_j sono gli eventi con assegnate probabilità $\{p_j\}$. Al “portalettere” Π_1 spettano ℓ_1 lettere per \mathcal{I}_1 che corrisponde dunque a una probabilità p_1 . Se viene estratto il numero 1 dovrà ritornare I_1 con probabilità² $P_1 \equiv N p_1$ mentre con probabilità $1 - P_1$ si ritorna l’“alias” I_N . Questo è l’inizio della tabella Alla seconda

$P_1 = N p_1$
$Y_1 = I_N$

TABELLA 1. lookup table del metodo degli alias

iterazione si procede allo stesso modo definendo $P_2 = N p_2$ e Y_2 coincide con l’evento che ha al momento maggiore probabilità tenendo conto che a I_N si è assegnata la nuova probabilità $p'_N = p_N - \frac{1}{N}(1 - P_1) = p_N - (\frac{1}{N} - p_1)$. Procedendo ricorsivamente si costruiscono così le tabelle $\{P_j, Y_j \mid j = 1, N\}$

P_1	P_2	...	P_N
Y_1	Y_2	...	Y_N

Per estrarre a caso dall’insieme $\{\mathcal{I}_j\}$ con probabilità $\{p_j\}$ si procede in questo modo:

- (i) si estrae un numero intero j a caso in modo uniforme tra 1 e N ($j = \text{randi}(N)$)
- (ii) si estrae un numero reale r dalla distribuzione uniforme tra 0 e 1 ($r = \text{rand}()$)
- (iii) se $r < P_j$ si sceglie l’output I_j
- (iv) altrimenti si sceglie Y_j (l’altro evento detto l’*alias* di I_j).

Dalla relazione $\ell'_j + \sum_k (\ell - \ell'_k) \delta_{Y_k, j} = \ell_j$, che esprime il fatto che tutte le lettere arrivano a destinazione segue la correttezza del calcolo delle probabilità

$$\begin{aligned} \ell'_j/L + \sum_k (\ell - \ell'_k)/L \delta_{Y_k, j} = \\ \frac{1}{N} (P_j + \sum_k (1 - P_k) \delta_{Y_k, j}) = \ell_j/L \equiv p_j. \end{aligned}$$

Il metodo fu introdotto da A. J. Walker [Wal77] ed è noto come *metodo degli alias*. Si deve soprattutto a D. Knuth la sua diffusione ([Knu98, BS99])³.

I codici `matlab` sono presenti su dropbox `~/MPF2023/Programmi/Matlab/Knuth` dove si troverà la necessaria documentazione per utilizzare i codici (please read...Readme!). Si può cominciare utilizzando il programma `Kdemo.m` che offre una scelta di distribuzioni di probabilità. Studiando il testo ci si può così familiarizzare con l’utilizzo di `kalias.m`, `Krand.m` e `Krandc.m` che implementano il metodo degli alias. Si noti che una volta costruite le tabelle di probabilità e degli alias $\{P, Y\}$ l’estrazione di un campione costa soltanto la generazione di un intero random j , di un reale random r e un confronto $r < P_j$, e questo indipendentemente dalla dimensione dell’insieme N . Non stupisce che il metodo sia molto “popolare”!

Il metodo si può utilizzare anche per il caso di distribuzioni continue. A questo scopo si suddivide il campo di definizione in tanti canali (*bins*) in modo che la scelta del canale viene effettuata con il metodo degli alias e all’interno del canale si approssima la distribuzione con una funzione lineare per la quale esiste un algoritmo veloce. Il lavoro addizionale costituito dalla costruzione delle tabelle è ampiamente ripagato dalla velocità dell’algoritmo (a patto che si lavori in assetto costante, cioè che la distribuzione di probabilità sia sempre la stessa!).

²infatti viene estratto 1 con probabilità $1/N$.

³Un’implementazione alternativa del metodo si trova in [M.D91]

4.2.1. Un’applicazione elementare. Vediamo cosa ci offre il metodo degli alias in un problema elementare già affrontato, il lancio di due dadi. Sappiamo che la probabilità p_j dell’uscita di j come somma del valore dei due dadi si ottiene come convoluzione $p \star p$ con $p = (1, 1, 1, 1, 1, 1)/6$ e precisamente

$j \rightarrow$	2	3	4	5	6	7	8	9	10	11	12
$p \rightarrow$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

Utilizzando `kalias.m` troviamo le tabelle di alias

$j \rightarrow$	2	3	4	5	6	7	8	9	10	11	12
$Y \rightarrow$	7	6	7	5	5	5	6	5	7	9	8
$P \rightarrow$	$\frac{11}{36}$	$\frac{11}{18}$	$\frac{11}{12}$	1	$\frac{35}{36}$	$\frac{35}{36}$	$\frac{5}{6}$	$\frac{5}{6}$	$\frac{11}{12}$	$\frac{11}{18}$	$\frac{11}{36}$

Questa tabella si legge così: estraendo un intero a caso tra 2 e 12, $j = 5$ uscirà con probabilità $1/11$ e dato che il corrispondente $P(4)$ vale 1 il valore 5 viene accettato senz’altro; a questo dobbiamo aggiungere il contributo dei tre “alias” uguali a 5 e cioè $j = 6, 7$, con probabilità $1 - P$ e cioè $(1 - 35/36)/11$ e $j = 9$ con probabilità $(1 - 5/6)/11$. Quindi in totale 5 viene estratto $(1 + 2/36 + 1/6)/11 = 1/9 = 4/36$. Analogamente per il 6: si ha la probabilità $(35/36 + (1 - 11/18) + (1 - 5/6))/11 = 5/36$. In questo caso *il metodo non è conveniente* in quanto dobbiamo estrarre un intero e un numero reale a caso mentre due dadi costano solo due interi a caso. Ma immaginiamo il caso di n dadi: invece di estrarre n interi a caso risulta conveniente il metodo degli alias. Infatti, una volta costruite le tabelle (Y, P) si tratta sempre di estrarre solo un intero e un reale a caso. Ovviamente il vantaggio cresce con la complessità del problema e bisogna anche tenere conto del lavoro aggiuntivo della costruzione delle tabelle; questo aumenta in termini di tempo in modo approssimativamente proporzionale al quadrato di n ma va fatto una sola volta per preparare l’esperimento.

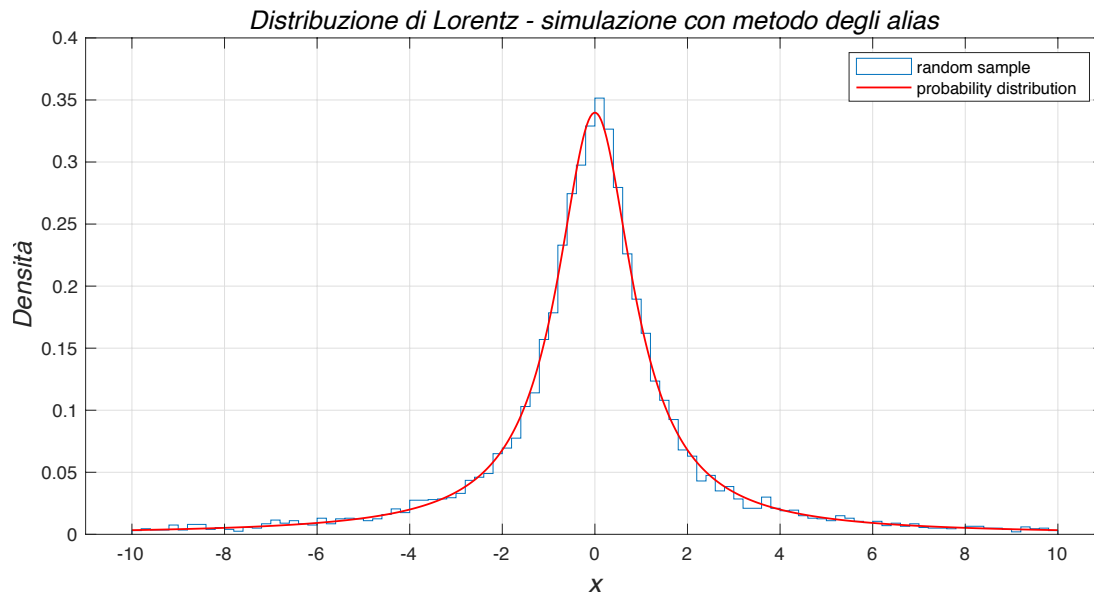


FIGURA 4.1. $O(10^5)$ campioni estratti dalle distribuzioni di Lorentz

ESERCIZIO 68. Utilizzare il metodo degli alias per simulare il lancio di n dadi per N volte. Ogni lancio “costa” un intero e un reale estratti a caso mentre il metodo diretto richiederebbe di estrarre n interi a caso.

SOLUZIONE. Il codice matlab `manydices.m` calcola la probabilità come convoluzioni successive $p \circ p \circ \dots \circ p$ (n volte) e costruisce le tabelle $\{P_j, Y_j\}$ utilizzando `kalias.m`. Il codice `krand.m` estrae successivamente il risultato del lancio di n dadi. In Fig.4.3 per confronto si riporta l’istogramma per un milione di lanci di venti dadi con il plot della distribuzione p .



ESERCIZIO 69. Realizzare un algoritmo per estrarre numeri reali dalla densità di probabilità $p(x) = \alpha x$, $0 \leq x \leq X_{\max}$. Stesso problema per estrarre numeri interi con probabilità discreta $p_n = \beta n$, $n \leq N_{\max}$.

SOLUZIONE. Dato che il valor medio $\mathbb{E}(x)$ è dato da

$$\mathbb{E}(x) = \int_0^{X_{\max}} f(x) x dx / (X_{\max}^2/2)$$

è sufficiente cambiare variabile $\xi = x^2$ per ottenere

$$\mathbb{E}(x) = \int_0^{X_{\max}^2} f(\sqrt{\xi}) d\xi / X_{\max}^2$$

il che ci permette di rappresentare x come la radice quadrata di una variabile distribuita uniformemente sull’intervallo $[0, X_{\max}^2]$. Dunque una semplice chiamata `x = f(1e3, 100)` con

```
f = @(N, xmax) xmax*sqrt(rand(N,1));
```

restituisce mille numeri distribuiti tra 0 e 100 secondo una densità lineare. In questo caso, essendoci un codice ad hoc, il metodo degli alias non è competitivo (si veda `kdemo.m`). Da Knuth [Knu98] Vol.II, sec.3.4.1.B si ottiene un’altra possibilità ancora più efficiente in quanto evita il “dispendioso” utilizzo della funzione `sqrt()`:

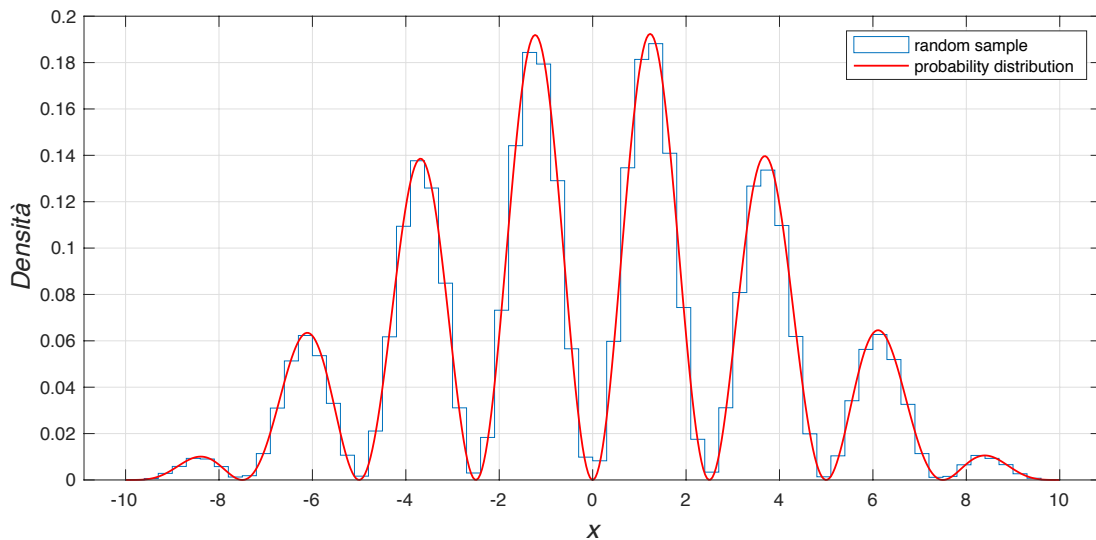


FIGURA 4.2. $O(10^5)$ campioni estratti da una figura di interferenza elettronica

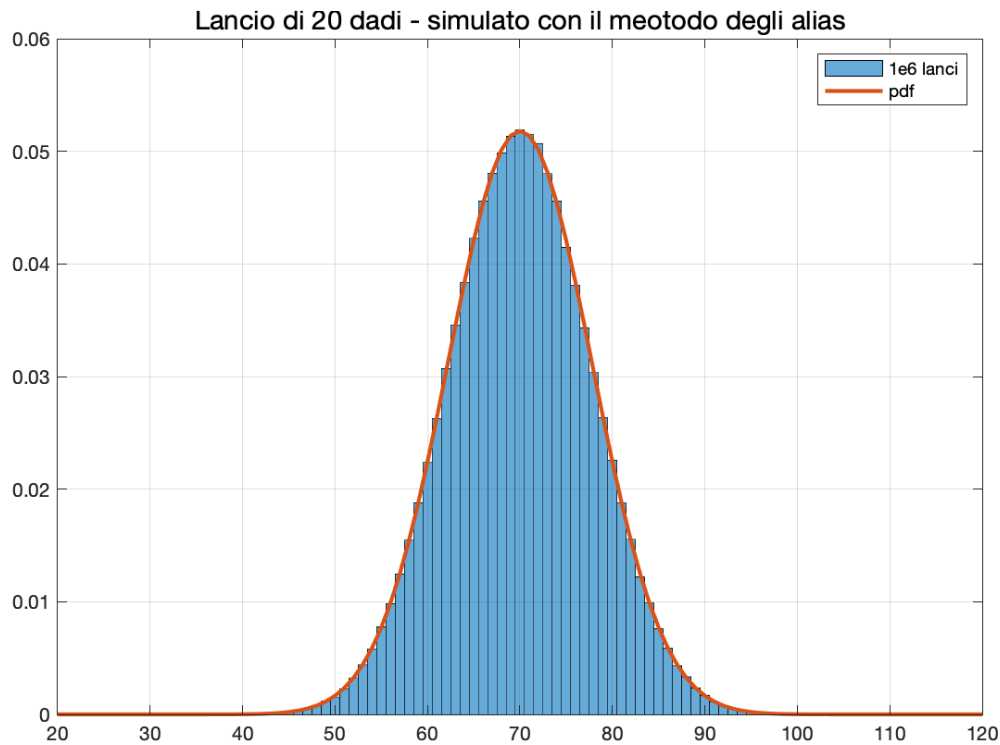


FIGURA 4.3. Metodo degli alias applicato al lancio di molti dadi da gioco

```
function X = lpdf(N, xmax)
r = rand(2,N);
X = xmax*max(r);
```

oppure con una “funzione anonima”

```
X = @(N, L) L*max(rand(2,N));
```

Dunque ogni valore a caso X costa solo due `rand()` e un confronto - più o meno come il metodo degli alias, ma non ci sono neanche da costruire le tabelle... Per dimostrare la validità dell’algoritmo è sufficiente calcolare

$$\mathbb{P}(X < x) = \int_0^1 dx \int_0^x dx dy + \int_0^1 dy \int_0^y dx dy = x^2$$

da cui si ricava la densità $\rho(x) = 2x$. In generale se consideriamo un punto a caso nell’ipercubo a d dimensioni, la probabilità di $\max(x_1, x_2, \dots, x_d) < x$ è data da

$$\begin{aligned} \mathbb{P}(\max(\mathbf{x}) < x) &= \int_0^1 \int_0^{x_1} \dots \int_0^{x_1} \prod dx_i + \text{altre } d-1 \text{ direzioni} = \\ &= x^d / d * d = x^d \end{aligned}$$

il fattore d essendo dovuto al fatto che il valore massimo può essere assunto da una qualunque delle d coordinate. Questo allora corrisponde alla densità $\rho(x) = d x^{d-1}$. ◀

4.3. Montecarlo dinamico

Per le applicazioni recenti del Montecarlo interessanti la Fisica Teorica, il problema centrale consiste nel generare una successione casuale di configurazioni di fase per un sistema di meccanica statistica distribuite secondo la formula di Gibbs. In sostanza si utilizza una catena di Markov irriducibile e, sperabilmente, con buone proprietà di convergenza, per campionare i punti $\mathbf{x} \in \mathcal{S}$ nello spazio delle fasi \mathcal{S} e determinare così una stima dell'integrale dell'Eq.(50). Un algoritmo MC specifico consiste in una catena di Markov con probabilità di transizione $P_{\mathbf{x} \rightarrow \mathbf{x}'}$ costruita in modo tale che la distribuzione di equilibrio ⁴ $\pi(\mathbf{x}') = \sum_{\mathbf{x}} \pi(\mathbf{x}) P_{\mathbf{x} \rightarrow \mathbf{x}'}$ sia data da $\pi(\mathbf{x}) = \exp\{-\beta H[\mathbf{x}]\}/Z(\beta)$. Concentriamo l'attenzione su un particolare problema di meccanica statistica, quello di sistemi di spin distribuiti su un reticolo spaziale. Un tipico esempio è il modello di Heisenberg: in ogni punto di un reticolo $\mathbf{x} \in \mathbb{Z}^3$ è definito un vettore unitario $\mathbf{n}(\mathbf{x})$ che rappresenta la direzione del momento magnetico $\boldsymbol{\mu} = \mu \mathbf{n}$. L'energia di interazione tra due dipoli magnetici decresce molto velocemente con la distanza, quindi in una prima schematizzazione si considera che ogni magnete elementare interagisca solo con i magneti più vicini (i *primi vicini*) e l'energia totale assume perciò la forma

$$(52) \quad H(\boldsymbol{\mu}) = -J \sum_{\langle \mathbf{x}, \mathbf{y} \rangle} \boldsymbol{\mu}(\mathbf{x}) \cdot \boldsymbol{\mu}(\mathbf{y})$$

essendo J una costante e la notazione $\langle \mathbf{x}, \mathbf{y} \rangle$ significa che la somma va estesa alle coppie di primi vicini. In presenza di un campo magnetico esterno \mathbf{B} all'energia va aggiunto anche un termine $-\sum_{\mathbf{x}} \mathbf{B} \cdot \boldsymbol{\mu}(\mathbf{x})$. Ora il valore di una grandezza fisica, quale la *magnetizzazione* esprimibile attraverso un integrale del tipo

$$(53) \quad \mathbb{E}(\boldsymbol{\mu}(\bar{\mathbf{x}})) = Z(\beta)^{-1} \int \dots \int \prod_{\mathbf{x} \in \mathbb{Z}^3} d\boldsymbol{\mu}(\mathbf{x}) e^{-\beta H(\boldsymbol{\mu})} \boldsymbol{\mu}(\bar{\mathbf{x}}) .$$

indicando con $\bar{\mathbf{x}}$ un qualunque punto nel reticolo; conviene poi mediare su $\bar{\mathbf{x}}$ prendendo la somma su tutto il reticolo $\mathcal{M} = \mathbb{E}(\sum_{\mathbf{x}} \boldsymbol{\mu}(\mathbf{x}))$. In generale si desidera valutare il valor medio di una qualunque grandezza fisica che sia una funzione definita nello spazio delle fasi,

$$\mathbb{E}(\mathcal{O}(\boldsymbol{\mu})) = Z(\beta)^{-1} \prod_{\mathbf{x} \in \mathbb{Z}^3} \int d\boldsymbol{\mu}(\mathbf{x}) e^{-\beta H(\boldsymbol{\mu})} \mathcal{O}(\boldsymbol{\mu})$$

In casi speciali, per i modelli più semplici, si è riusciti a trovare metodi analitici per calcolare integrali pur così complessi ⁵. Tuttavia, nella gran parte dei modelli interessanti, la via della simulazione Montecarlo rappresenta l'unica risorsa. Sono stati concepiti vari metodi per realizzare questo obiettivo. Forse il più semplice concettualmente è quello del cosiddetto “heat bath”. Il metodo consiste nel visitare sequenzialmente tutti i punti del reticolo (in gergo *effettuare una sweep*) e di volta in volta determinare $\boldsymbol{\mu}(\mathbf{x})$ come se fosse l'unica variabile dinamica (tutte le altre variabili si tengono temporaneamente congelate). Dato che $\boldsymbol{\mu}(\mathbf{x})$ interagisce solo con un numero limitato di primi vicini, lo spazio di probabilità è costituito dai punti della sfera con densità di probabilità proporzionale a

$$\exp \left\{ -\beta J \boldsymbol{\mu}(\mathbf{x}) \cdot \sum_{k=1}^6 \boldsymbol{\mu}(\mathbf{x}_k) - \mathbf{B} \cdot \boldsymbol{\mu}(\mathbf{x}) \right\}$$

dove \mathbf{x}_k indica i primi vicini di \mathbf{x} . A titolo di esempio, prendiamo un modello più semplice che basterà a chiarire l'idea. Il modello di **Ising** prevede in ogni punto di un reticolo spaziale una variabile a due soli valori ± 1 , essenzialmente una restrizione del modello di Heisenberg in cui si vincoli ogni magnete ad avere la stessa direzione in modo che sopravviva come variabile solo il verso del vettore di momento magnetico. Se consideriamo di più il caso di un reticolo piano ⁶ il modello si riduce al seguente

$$(54) \quad \langle \mathcal{O}(\boldsymbol{\mu}) \rangle = Z(\beta)^{-1} \prod_{\mathbf{x} \in \mathbb{Z}^2} \sum_{\boldsymbol{\mu}(\mathbf{x}) = \pm 1} e^{\beta J \sum_{\langle \mathbf{x}, \mathbf{y} \rangle} \boldsymbol{\mu}(\mathbf{x}) \boldsymbol{\mu}(\mathbf{y})} \mathcal{O}(\boldsymbol{\mu}) .$$

⁴Bisogna notare che la definizione della matrice stocastica $p_{i \rightarrow j}$ corrisponde alla trasposta di quella utilizzata in precedenza, per cui la matrice va applicata *da destra* al vettore di probabilità e la normalizzazione è data da $\sum_j p_{i \rightarrow j} = 1$.

⁵Per il modello di Ising la fonte più completa è il testo di B. McCoy e T.T. Wu [BW14]

⁶Il modello di Ising in due dimensioni è solubile esattamente [Ons44]. Presenta un punto critico a $\beta J = \ln(1 + \sqrt{2})/2$. Uno studio completo del modello si trova sul libro già citato di B. McCoy e T.T. Wu.

Per un qualunque spin $\mu(\mathbf{x})$ i primi vicini sono quattro e per una data configurazione si possono riscontrare le situazioni seguenti

$$\begin{array}{cccccccccccc} \begin{array}{c} + \\ \square \\ + \end{array} & \begin{array}{c} + \\ \square \\ - \end{array} & \begin{array}{c} + \\ \square \\ + \end{array} & \begin{array}{c} - \\ \square \\ + \end{array} & \begin{array}{c} + \\ \square \\ + \end{array} & \begin{array}{c} + \\ \square \\ - \end{array} & \begin{array}{c} + \\ \square \\ - \end{array} & \begin{array}{c} - \\ \square \\ + \end{array} & \begin{array}{c} - \\ \square \\ + \end{array} & \begin{array}{c} + \\ \square \\ + \end{array} & \dots \end{array}$$

più tutte le altre simili per un totale di 2^4 . La probabilità di $\mu(\mathbf{x})$ dipende solo dalla somma degli spin primi vicini; posto $J = 1$ e definita $s(\mathbf{x}) = \sum_{\mathbf{y}} \mu(\mathbf{y}) \delta_{\|\mathbf{x}-\mathbf{y}\|,1}$ la probabilità sarà data semplicemente da

$$\mathbb{P}(\mu(\mathbf{x}) = \pm 1) = \frac{e^{\pm \beta s(\mathbf{x})}}{e^{\beta s(\mathbf{x})} + e^{-\beta s(\mathbf{x})}}.$$

La somma $s(\mathbf{x})$ può assumere solo cinque valori differenti, ed è quindi molto facile e rapido assegnare a $\mu(\mathbf{x})$ un valore in accordo con questa formula. Si applica questa ricetta a tutti gli spin successivamente e poi si ricomincia. Ovviamente è necessario limitare il sistema a un numero finito di punti individuando una regione finita del piano. In gergo questo si dice “fissare il cut-off infrarosso”. Non è consigliabile effettuare la sweep punto per punto, ma conviene agire su un sottoreticolo di variabili disaccoppiate, ad esempio tutte quelle di posto pari (somma delle coordinate) e in seguito quelle di posto dispari. Il problema si presta perciò a essere affrontato *in parallelo*. Un grande numero di sweeps sarà necessario per generare una configurazione del modello non troppo correlata alla iniziale. Infatti certe configurazioni locali tendono a mantenersi stabili: per β grande gli spin tendono a essere allineati e si formano dei domini di spin uguali che impiegano molti passi per essere modificati. Esiste perciò un tempo caratteristico per ogni modello che è necessario attendere prima che la correlazione sia diminuita di un fattore accettabile - si considera cioè che la correlazione avrà una dipendenza dal “tempo” della catena di Markov del tipo $C(t, t') = \exp(-|t - t'|/\tau)$. Tanto più piccolo è τ , il *tempo di autocorrelazione*, tanto più efficiente sarà il processo di generazione di configurazioni statisticamente indipendenti. Molto studio è stato messo nella analisi e nello sviluppo di algoritmi nel tentativo di abbattere il tempo di autocorrelazione. È un buon investimento, perchè un algoritmo che migliori l'efficienza rappresenta un risparmio di tempo e/o di denaro. Per chi sia interessato a sviluppi recenti si consiglia di consultare i Proceedings delle conferenze di Teorie di Gauge su Reticolo, che escono con cadenza annuale, in particolare [Sok91]. Qualche risultato di esperimenti Montecarlo sul modello sono riportati nelle Fig.(4.8) e seguenti alla fine del capitolo.

4.3.1. Bilancio dettagliato e l'algoritmo di Metropolis. Un metodo molto utilizzato per realizzare un processo di Markov con distribuzione di probabilità stazionaria preassegnata $P(\omega) = Z^{-1} \exp\{-\beta S(\omega)\}$ è quello legato al nome di **Nicholas Metropolis**, collaboratore di *Enrico Fermi*, di **Stan Ulam**, **J. Von Neumann** e **Arianna Rosenbluth**⁷ a Los Alamos per il progetto Manhattan⁸. Il metodo è basato su una strategia di “mosse”, cioè modifiche nella configurazione di un sistema, assoggettate a una decisione di accettazione-rigetto in base alla variazione del valore dell'azione S . Consideriamo un sistema dinamico i cui gradi di libertà siano legati ai nodi di un reticolo cubico $\mathbf{x} = \mathbf{n}\mathbf{a}$, con $\mathbf{n} \in V \subset \mathbb{Z}^d$ (un volume finito) e l'azione abbia una struttura “locale”, cioè i gradi di libertà siano rappresentati da campi $\phi(\mathbf{x})$ con accoppiamenti solo tra primi vicini, ad es.

$$S(\phi) = \sum_{\mathbf{n}} \left(U(\phi(\mathbf{n})) + \sum_{\mu=1}^d \frac{1}{2} (\phi(\mathbf{n}) - \phi(\mathbf{n} + \mathbf{a}\mathbf{e}_{\mu}) - \phi(\mathbf{n} - \mathbf{a}\mathbf{e}_{\mu}))^2 \right),$$

dove \mathbf{e}_{μ} indicano i vettori di lunghezza unitaria in tutte le d direzioni. Il metodo consiste nel partire da una configurazione di campo arbitraria ϕ_0 e procedere a costruire una successione di configurazioni $\phi_1, \phi_2, \dots, \phi_n, \dots$ secondo la strategia che segue: si visitano successivamente tutti i siti \mathbf{n} del reticolo in V e si propone una mossa $\phi(\mathbf{n}) \rightarrow \phi'(\mathbf{n}) = \phi(\mathbf{n}) + \delta\phi$. La mossa viene accettata se la variazione dell'azione è negativa $\delta S(\phi) < 0$ e cioè la mossa è nella direzione di una configurazione più probabile. Se questa fosse l'unica regola il processo terminerebbe quando si raggiungesse il valore più piccolo dell'azione. Al contrario si decide di accettare la mossa anche quando la variazione dell'azione è positiva ma solo con probabilità $\exp\{-\delta S\}$ mentre si mantiene la configurazione inalterata con probabilità $1 - \exp\{-\delta S\}$.

Il metodo è estremamente generale ed è stato utilizzato in ricerche di teorie di campo quantistiche su reticolo a partire dagli anni '80 del secolo scorso. Ha il pregio della semplicità e della economia di calcolo. Infatti la

⁷Si veda la voce “Rosenbluth” su Wikipedia. A.R. codificò per prima l'algoritmo sul calcolatore MANIAC.

⁸Nel folder `dropbox/MPF2023/Documents/Montecarlo` si trova un articolo di Metropolis su *Los Alamos Science* che descrive il periodo '43-53 con l'avvento dei primi calcolatori elettronici e la nascita del metodo Montecarlo [?]. Si veda anche [Hit03].

variazione della azione comporta un numero molto contenuto di operazioni aritmetiche dal momento che in ogni mossa solo un numero limitato di campi è coinvolto. Il metodo è affetto dal problema della autocorrelazione e per combattere questo difetto nei regimi in cui ci si avvicina a un punto critico sono state individuate delle modifiche all'algoritmo per renderlo più efficiente. Si possono consultare le lezioni di Sokal a questo riguardo.

Conviene sottolineare che la correttezza del metodo di Metropolis è garantita dal principio di “bilancio dettagliato” che adesso illustriamo. Una matrice stocastica $M_{ij} \equiv \{P(j \rightarrow i)\}$ è detta reversibile rispetto alla distribuzione di equilibrio $\pi = \{\pi_j\}$ se vale l'identità

$$\forall(i, j) : \pi_i P(i \rightarrow j) = \pi_j P(j \rightarrow i) .$$

In termini puramente matematici la relazione esprime la condizione che la matrice M' , la trasposta di M , sia simmetrica rispetto alla metrica fissata da π ; ciò significa che se definiamo $\langle u, v \rangle \equiv \sum_i u_i \pi_i v_i$ la condizione di bilancio dettagliato equivale a $\langle u, M'v \rangle = \langle M'u, v \rangle$ che costituisce appunto la condizione per M' di essere una matrice simmetrica. Una catena di Markov con matrice di transizione reversibile rispetto a π ammette la stessa π come distribuzione di equilibrio, infatti sommando l'equazione precedente rispetto a i si ottiene

$$\sum_i \pi_i P_{i \rightarrow j} = \pi_j \sum_i P_{j \rightarrow i} = \pi_j$$

ESERCIZIO 70. Dimostrare che la matrice stocastica del modello di Ehrenfest delle due urne con N particelle (si veda la Sez.A.6) soddisfa la condizione di bilancio dettagliato rispetto alla distribuzione stazionaria $\rho_n = 2^{-N} \binom{N}{n}$, $n = \{0, 1, 2, \dots, N\}$.

La proprietà di bilancio dettagliato non è una condizione necessaria affinché π sia la distribuzione stazionaria, ma costituisce una condizione sufficiente, in genere di più facile verifica rispetto alla verifica diretta della stazionarietà di π_i . Il fatto che la condizione non sia necessaria risulta ovvia quando si consideri che una matrice simmetrica, rispetto a qualunque definizione della metrica, presenta uno spettro di autovalori tutti reali, e questo non è vero in generale per le matrici stocastiche. Quello che interessa qui è però che la condizione vale per l'algoritmo di Metropolis, e questo dimostra che l'algoritmo ammette la distribuzione voluta come distribuzione stazionaria. Indicando con $q_{i \rightarrow j}$ la matrice **simmetrica** che rappresenta la “proposta” di modifica dello stato ($q_{i \rightarrow j} = q_{j \rightarrow i}$), si definisce per $i \neq j$

$$p_{i \rightarrow j} = \min\left(1, \frac{\pi_j}{\pi_i}\right) q_{i \rightarrow j} .$$

Detto in parole, lo stato i è cambiato in j con probabilità $q_{i \rightarrow j}$ se $\pi_j \geq \pi_i$ oppure con probabilità $q_{i \rightarrow j} \pi_j / \pi_i$ se $\pi_j < \pi_i$. Quindi, se scegliendo j in base alla proposta $q_{i \rightarrow j}$, si trova $\pi_j < \pi_i$, la transizione è rifiutata con probabilità $1 - \pi_j / \pi_i$. In questo caso lo stato non è modificato *ma deve essere registrato come nuova configurazione*. Ciò è importante, perché se la mossa $i \rightarrow j$ viene rifiutata la successione di eventi/stati sarà \dots, i, i, \dots , e cioè i verrà conteggiato due volte nella statistica finale. La ricetta soddisfa il bilancio dettagliato, infatti, supponendo $\pi_j < \pi_i$ si ha ⁹.

$$\pi_i p_{i \rightarrow j} = \pi_i \frac{\pi_j}{\pi_i} q_{i \rightarrow j} = \pi_j q_{i \rightarrow j} = \pi_j q_{j \rightarrow i} = \pi_j p_{j \rightarrow i}$$

e quindi la ricetta per p definisce una catena di Markov avente π_i come distribuzione di equilibrio. Maggiori dettagli sul testo di Baldi [?] e su quello di Berg [Ber04] oppure su [DSC98], di cui si può trovare copia in [dropbox/MPF2023/Documents/Montecarlo](https://www.dropbox.com/sh/MPF2023/Documents/Montecarlo).

Come generalizzazione del metodo si può anche utilizzare una funzione $q_{i \rightarrow j}$ asimmetrica, nel qual caso lo step di accettazione/rigetto deve essere modificato come segue:

$$\text{if } \text{rand}() < \inf\left(1, \frac{\pi_j q_{j \rightarrow i}}{\pi_i q_{i \rightarrow j}}\right) \text{ then accept.}$$

Si verifica infatti che il bilancio dettagliato è ancora soddisfatto. Si può pensare all'algoritmo di Metropolis come una correzione al volo di una proposta $q_{i \rightarrow j}$ che *non rispetti il bilancio dettagliato*. Infatti se la mossa dettata da $q_{i \rightarrow j}$ rispettasse il bilancio dettagliato avremmo che la mossa verrebbe sempre accettata.

⁹Si noti che la relazione è simmetrica nello scambio $i \leftrightarrow j$ e quindi non è necessario considerare l'altro caso $\pi_i < \pi_j$.

ESERCIZIO 71. Codificare l'algoritmo di Metropolis per costruire una successione di interi distribuiti secondo a) la distribuzione di Poisson oppure b) secondo il modello di Ehrenfest (distribuzione binomiale), utilizzando una mossa di Metropolis non necessariamente simmetrica. Verificare tuttavia che la percentuale di accettazione è ottimale con la proposta simmetrica ($\approx .8$).

SOLUZIONE. Si veda il codice `matlab` `metroPoisson.m` e `metroEhrenf.m` nel folder `dropbox/matlab/Cap.5`.



4.3.2. Il modello di Ising in una dimensione. Il modello di Ising in una dimensione è caratterizzato dalla funzione di partizione ¹⁰

$$\begin{aligned} Z(\beta, H) &= \sum_{s_1=(-1,1)} \sum_{s_2=(-1,1)} \dots \sum_{s_N=(-1,1)} \prod_j e^{\beta J s_j s_{j+1} + \beta B s_j} \\ &= \sum_{s_1} \sum_{s_2} \dots \sum_{s_N} \langle s_1 | P | s_2 \rangle \langle s_2 | P | s_3 \rangle \dots \langle s_N | P | s_1 \rangle. \end{aligned}$$

Si può calcolare Z analiticamente [HW41] identificando le somme sugli spin con le somme sugli indici nel prodotto di matrici 2×2 . Esplicitamente si ha

$$P = \begin{pmatrix} e^{\beta(J+B)} & e^{-\beta J} \\ e^{-\beta J} & e^{\beta(J-B)} \end{pmatrix}$$

e dunque la funzione di partizione si calcola prendendo la traccia $Z = \text{Tr}(P^N)$. La magnetizzazione si calcola poi prendendo la derivata logaritmica rispetto a B . Per calcolare la potenza N -esima di P si può trovare lo spettro

$$\det(P - \lambda \mathbb{I}) = \lambda^2 - 2\lambda e^{\beta J} \cosh(\beta B) + e^{2\beta J} - e^{-2\beta J} = 0$$

e quindi

$$\begin{aligned} Z(\beta) &= e^{N\beta J} \left(\cosh(\beta B) + \sqrt{e^{-4\beta J} + \sinh^2(\beta B)} \right)^N \\ &+ e^{N\beta J} \left(\cosh(\beta B) - \sqrt{e^{-4\beta J} + \sinh^2(\beta B)} \right)^N \\ &= e^{N\beta J} \lambda_+^N \left(1 + \frac{\lambda_-^N}{\lambda_+^N} \right) \end{aligned}$$

avendo indicato con λ_+ l'autovalore più grande. Nel limite $N \rightarrow \infty$ si trova allora

$$Z(\beta) = e^{N\beta J} \lambda_+^N.$$

La magnetizzazione media si ottiene calcolando la derivata logaritmica rispetto a B della $Z(\beta)$. Si trova facilmente

$$(55) \quad M = \frac{1}{N\beta} \frac{\partial}{\partial B} \ln(Z(\beta)) = \frac{\sinh(\beta B)}{\sqrt{e^{-4\beta J} + \sinh^2(\beta B)}}$$

M risulta funzione analitica di β , J e B per tutti i valori reali dei parametri e dunque il modello non presenta transizioni di fase.

4.3.3. Il modello di Potts. Una generalizzazione del modello di Ising è costituita da quello introdotto da R.B.Potts [Pot52]. Come nel modello di Ising, in ogni nodo di un reticolo piano è definito uno *spin* rappresentato da un vettore di lunghezza unitaria e orientato in un certo numero q di possibili direzioni. L'energia di interazione tra spin è data in modo molto schematico da $-J$ nel caso di spin primi vicini paralleli e 0 altrimenti. È presente anche un campo magnetico esterno B che contribuisce all'energia un termine $-\vec{B} \cdot \vec{s}$. Se $q = 2$ il modello coincide con quello di Ising, mentre per $q > 2$ la molteplicità degli stati modifica quantitativamente e qualitativamente le proprietà statistiche del modello. Per $q = 3$ si può identificare il modello con quello di spin orientati in tre direzioni ad angolo $2\pi/3$ (il simbolo della Mercedes!): in questo caso l'energia $-J \vec{s}_1 \cdot \vec{s}_2$ assume valori $(-J, J/2, J/2)$ per angoli $(0, 2\pi/3, -2\pi/3)$ e dato che il peso statistico non cambia ridefinendo lo zero dell'energia questa forma dell'energia coincide con quella del Potts a $q = 3$, riscalando β di un fattore $2/3$. Ne segue che il modello

¹⁰Adottiamo condizioni al contorno di tipo periodico, cioè il sito N ha come primi vicini $N-1$ e 1 . Gli spin si trovano perciò su N punti equidistanti su un cerchio.

di Potts ammette la simmetria Z_3 delle radici cubiche dell'unità. In 2 dimensioni il modello presenta una transizione di fase per $\beta = \ln(1 + \sqrt{q})$: per $q \leq 4$ si ha una transizione di secondo ordine mentre la transizione è del primo ordine per $q > 4$. In dimensione superiore la transizione è del primo ordine già per $q \geq 3$. Le proprietà del modello si esplorano attraverso un Montecarlo che basato sugli stessi codici del modello di Ising con poche modifiche. Nelle Fig. 4.11 4.12 a pag.128 sono riportati risultati della simulazione sul modello di Potts in due o tre dimensioni e $q = 3, 6$. Per informazioni dettagliate sul modello di Potts si possono consultare [J.82, Wu82, HKW78, WY76].

Una formulazione meno schematica dell'interazione tra spin è data dal modello Z_N in cui gli spin sono orientati in N direzioni corrispondenti alle radici N -esime dell'unità e l'interazione è data da $-J \cos(\theta_i - \theta_j)$. Per $N = 2$ il modello coincide con quello di Ising, per $N = 3$ è equivalente al Potts a tre stati e presenta una transizione di fase a $\beta = \frac{2}{3} \log(1 + \sqrt{3}) \approx 0.67$. Per $N = 4$ è noto che la transizione avviene a un valore di β doppio rispetto all'Ising in 2D. Per $N \geq 5$ la struttura di fase è più complessa. Si veda [AJ82]. Tutti i modelli qui considerati sono affrontabili con il medesimo codice Montecarlo semplicemente variando la forma dell'interazione.

4.3.4. Esperimenti numerici - modello di Ising. Le idee generali riguardo gli algoritmi Montecarlo sono facilmente tradotti in codici di calcolo numerico. Abbiamo come consuetudine utilizzato il linguaggio `matlab` che permette di realizzare molto facilmente e in modo trasparente gli algoritmi che ci interessano. Prima di esaminare il più semplice modello di meccanica statistica non banale e di cui si conosce la soluzione esatta (Ising in due dimensioni) possiamo fare un esercizio di riscaldamento codificando un Metropolis per il modello di Ising in una dimensione di cui conosciamo la soluzione esatta per la magnetizzazione.

ESERCIZIO 72. Realizzare un codice Montecarlo per il modello di Ising su un reticolo mono-dimensionale, i.e.

$$\exp\{-\beta S(J, B)\} \equiv \exp\left\{\beta J \sum_{j=1}^N s_j s_{j+1} + \beta B \sum_{j=1}^N s_j\right\}$$

dove J rappresenta la forza dell'accoppiamento tra gli spin e B è il campo magnetico esterno. Con la simulazione numerica si calcoli la *magnetizzazione* in funzione di H :

$$M = \left\langle \frac{1}{N} \sum s_j \right\rangle.$$



Il reticolo è definito da $n = 1, 2, \dots, N$ e consideriamo condizioni al contorno periodiche. Conviene prendere N pari e sud dividere il reticolo nel sottoreticolo dei posti dispari $1, 3, 5 \dots N - 1$ e in quello dei posti pari. La mossa di Metropolis sarà allora applicata successivamente ai due sottoreticoli, nel programma 'odd' e 'even'. I primi vicini dei siti dispari sono dati da `even` e da `even_nnb = circshift(even,1)` e quelli dei siti pari da `odd` e da `odd_nnb=circshift(odd,-1)`, cosicché ad es. si hanno le relazioni di prossimità come nella tabella che mostra per ogni 'odd' i suoi primi vicini immediatamente sopra e sotto:

even	2	4	6	8	...
odd	1	3	5	7	...
even_nnb	N	2	4	6	...

mentre per gli 'even' si trova

odd_nnb	3	5	7	9	...
even	2	4	6	8	...
odd	1	3	5	7	...

La mossa di Metropolis per gli 'odd' è semplicemente

```
r = rand(1,N);
accept = r < exp(-2*beta.*(even+even_nnb+B).*odd); % Metropolis
odd(accept) = -odd(accept); % Inversione degli spin
odd_nnb = circshift(odd,-1);
```

e quella per i pari è del tutto simile. In seguito alla mossa anche l'array dei primi vicini deve essere aggiornato. Il programma confronta il calcolo della magnetizzazione attraverso il Metropolis con la formula esatta dell'(55) (si veda la Fig.4.4).

Per il modello in due dimensioni si considera un reticolo regolare $\Lambda = \{s_{n_1, n_2} \mid |(n_1, n_2) \in \mathbb{Z}^2\}$ e ogni variabile di spin s_i assume due possibili valori ± 1 . È necessario introdurre una regione limitata del reticolo che abbiamo scelto come $0 \leq n_i \leq N - 1$. Si organizza la simulazione fissando

- (i) N , tipicamente compreso tra 128 e 1024;
- (ii) β , la temperatura inversa (.2 - .6);
- (iii) B , il campo magnetico esterno parallelo agli spin;
- (iv) N_{sweeps} , il numero di *sweeps* (1000-5000)
- (v) **therm**, il numero di sweeps preliminari perchè gli spin raggiungano una situazione di equilibrio termico;
- (vi) **skip**, il numero di sweeps che passano tra una misura e l'altra delle osservabili (energia, magnetizzazione).

Risulta conveniente, anche dal punto di vista della semplicità di programmazione, suddividere il reticolo in quattro sottoreticoli $\Lambda_{ee}, \Lambda_{eo}, \Lambda_{oe}$ e Λ_{oo} così definiti $\Lambda_{ee} = \{s_{n_1, n_2} \mid n_1 \text{ \& } n_2 \text{ pari}\}$, $\Lambda_{eo} = \{s_{n_1, n_2} \mid n_1 \text{ pari, } n_2 \text{ dispari}\}$, $\Lambda_{oe} = \{s_{n_1, n_2} \mid n_1 \text{ dispari, } n_2 \text{ pari}\}$, $\Lambda_{oo} = \{s_{n_1, n_2} \mid n_1 \text{ \& } n_2 \text{ dispari}\}$.

L'interazione tra gli spin è a primi vicini e quindi è possibile applicare la mossa di Heat Bath o di Metropolis successivamente a ciascun sottoreticolo Λ in un colpo solo e avendo a disposizione un calcolatore con più processori l'algoritmo può essere realizzato in parallelo. I dettagli si trovano nei codici matlab nel folder

~/Programmi/Matlab/Montecarlo/Metropolis/Ising2

Per questi codici non si può parlare di "calcolo parallelo", ma ci si affida all'efficienza del compilatore di **matlab** per ottimizzare il codice. Per modelli in dimensione superiore a due il reticolo può essere partizionato nello stesso modo pari/dispari e si possono individuare sottoreticoli che possono essere aggiornati con una singola istruzione di matlab. Ad es. nel caso dell'Ising in tre dimensioni conviene definire s_{eee} l'insieme degli spin su siti con coordinate pari, s_{oe} quelli con coordinata 1 dispari e le altre due pari, s_{eoe} quelli con coordinata 2 dispari e le altre due pari, e via dicendo; in totale si avranno 8 sottoreticoli per ognuno dei quali si può procedere all'aggiornamento di Metropolis in una singola istruzione di matlab. Va posta molta cura nella identificazione dei primi vicini. Per questo si consulti il codice **IsingMetrop3D.m**. Ci si potrebbe chiedere perchè sia necessario partizionare il reticolo invece che applicare la mossa di Metropolis all'intero reticolo: il fatto è che se cambiamo segno a tutti gli spin **l'energia tra i primi vicini rimane la stessa e quindi la mossa sarebbe sempre accettata e gli spin continuerebbero a cambiare segno ad ogni passo** senza modificare la configurazione; verrebbe meno così il requisito di irriducibilità della catena di Markov: lo stato del sistema deve potere raggiungere in un tempo opportuno ogni altro stato possibile.

La simulazione del modello di Ising per β vicino al valore critico mostra chiaramente la formazione di ampie zone di spin allineati (domini magnetici) che tendono ad assumere dimensioni sempre più grandi, fino a percolare da un'estremità all'altra del reticolo (si veda ad es. la Fig.4.8). Questi domini sono molto stabili, cioè risulta difficile smontarli con mosse locali quali quelle adottate nel metodo di Metropolis o nell'HeatBath. La conseguenza immediata è che le configurazioni costruite dal Montecarlo tendono a essere strettamente correlate e si perde quindi in statistica (è come se si stesse misurando le osservabili sulla stessa configurazione per molto tempo). Questo è il *problema dell'autocorrelazione* che si è imparato a combattere in vari modi con tecniche che tuttavia non sono universali (si vedano le lezioni di Sokal al riguardo).

La misura dell'energia media per spin e da questa la derivata rispetto alla temperatura $1/\beta$ ci fornisce il calore specifico. Nelle Figg.(4.7,4.9) il risultato dell'esperimento MC (Metropolis) per il calore specifico e la magnetizzazione ¹¹ $\langle \frac{1}{N^2} \sum_{ij} s_{ij} \rangle$. Gli esperimenti sono stati realizzati con un codice **matlab** e tipicamente hanno richiesto $\mathcal{O}(10^4)$ sec. Si noti che il Montecarlo riproduce molto bene la magnetizzazione a grandi valori di beta e che la discrepanza per valori vicini o inferiori al β critico potrebbe essere ridotta prendendo valori ancora

¹¹La magnetizzazione per campo esterno nullo è zero, ma si intende solitamente che si deve calcolare il limite per $B \rightarrow 0^+$ e questo dà un risultato finito. La formula analitica dovuta a Onsager ci dà [BW14] $M = (1 - \sinh(2\beta)^{-4})^{\frac{1}{8}}$ per $\beta > \beta_c = \frac{1}{2} \sinh^{-1}(1) = \frac{1}{2} \ln(1 + \sqrt{2}) \approx 0.4407...$, e 0 altrimenti.

più piccoli del campo esterno e aumentando la dimensione del reticolo, con conseguente rilevante aumento del tempo necessario alla simulazione.

ESERCIZIO 73. Modificare il codice Montecarlo per l'Ising 2 o 3 dimensionale per adattarlo al modello di Potts.

4.3.5. Vegas. Vegas, creato da G. P. Lepage [Lep78], è un algoritmo Montecarlo adatto per integrare funzioni reali definite su N variabili, con N compreso nel range $3, \dots, 20$. Nel folder `~/dropbox/MPF2023/Programmi/Vegas` si trovano alcuni esempi di programmi C che utilizzano l'implementazione di Vegas nella libreria matematica `gsl` (GNU Scientific Library). La documentazione si trova in `dropbox/MPF2023/Documents/Lepage.pdf` per la descrizione del programma, e `GSL.pdf` per tutta la libreria GSL. Gli esempi riguardano

- il calcolo di integrali come

$$\int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} dx dy dz \frac{1}{1 - \cos x \cos y \cos z}$$

o anche

$$\int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} dx dy dz \frac{1}{1 - (\cos x + \cos y + \cos z)/3}$$

di interesse in meccanica statistica [MI77].

- il calcolo dell'integrale

$$\int_{\mathbb{R}^3} d^3x \int_{\mathbb{R}^3} d^3y \frac{u_0(x)^2 u_0(y)^2}{|x - y|}$$

dove $u_0(x) = \exp(-|x|)$ rappresenta la correzione al primo ordine perturbativo per lo stato fondamentale dell'atomo di Elio;

- il calcolo del volume dell'ipersfera in un numero qualunque di dimensioni (funziona bene per dimensione non superiore a 18).

Per tutti questi casi si conosce il valore esatto dell'integrale e Vegas si comporta molto bene in tempi molto ragionevoli. Si consiglia di consultare la documentazione della Gnu Scientific Library se si intende migliorare l'implementazione attuale dei tre programmi.

4.4. Equazione di Langevin

Con il (capo)lavoro di Parisi e Wu [PW81] i metodi delle *equazioni differenziali stocastiche* (EDS) entrano a far parte della cassetta degli attrezzi del fisico interessato alle interazioni forti. In precedenza le EDS erano state assunte a base dell'approccio di Nelson ([Nel67]) alla Meccanica Quantistica con alterni successi. La tecnica delle EDS si era sviluppata già nella prima metà del XX secolo, con applicazioni a vari problemi di Fisica Classica ([Wax54]). Qui daremo solo un accenno, rimandando alla Ref.[PW81] e alla [DH87] chi desideri approfondire l'argomento.

L'idea nella sua formulazione più semplice è quella di studiare un'equazione differenziale (nota come *equazione di Langevin*) del tipo

$$(56) \quad dx(t) = b(x(t)) dt + \sqrt{\sigma(x(t))} dw(t)$$

dove $x(t)$, e $b(x(t))$ sono in generale funzioni a valori in \mathcal{R}^n , $\sigma(x(t))$ è una funzione reale positiva (in genere una costante, la "costante di diffusione") e $w(t)$ è il *moto browniano standard* in n dimensioni. La definizione di quest'ultimo può richiedere una lunga preparazione matematica (vedi [Nel67]), ma per quanto ci interessa si può definire per via puramente assiomatica: assumiamo che $w(0) = (0, 0, \dots, 0)$; allora

- per ogni scelta di istanti $t_1 < t_2 < \dots < t_N$ le v.a. $w(t_1), w(t_2), \dots, w(t_N)$ sono variabili aleatorie mutuamente Gaussiane;
- $\mathbb{E}[w(t)] = 0$
- $\mathbb{E}[w_i(t) w_j(s)] = \delta_{ij} \min(t, s)$.

Ne segue che gli incrementi $w(t_1) - w(t_2)$ e $w(t_3) - w(t_4)$ sono statisticamente indipendenti se gli intervalli (t_1, t_2) e (t_3, t_4) non si sovrappongono, ossia

$$(57) \quad \mathbb{E}[(w(t_2) - w(t_1))(w(t_4) - w(t_3))] = |(t_1, t_2) \cap (t_3, t_4)|$$

intendendo con $|I_1 \cap I_2|$ la misura dell'intersezione dei due segmenti. In particolare si ha

$$(58) \quad \mathbb{E}[|w(t + \tau) - w(t)|^2] = \tau$$

Si deve tenere presente che il processo browniano è tale che se si osserva la posizione a un istante t da quel momento il processo riparte come se le origini delle coordinate fossero centrate in $w(t)$, ossia

$$\mathbb{E}[w(t + \tau)|w(t)] = w(t)$$

Le relazioni precedenti si possono riassumere in una forma di facile utilizzazione:

$$(59) \quad \mathbb{E}\left[\int_0^t f(\tau) dw(\tau) \int_0^{t'} g(\tau') dw(\tau')\right] = \int_0^{\min(t, t')} f(\tau) g(\tau) d\tau$$

che si può ricordare più facilmente nella forma simbolica

$$\mathbb{E}[dw(\tau) dw(\tau')] = \delta(\tau - \tau') d\tau$$

essendo $\delta(\cdot)$ la distribuzione (funzione) di Dirac.

Una nota sulle dimensioni fisiche: dato che dw ha le dimensioni $[T]^{\frac{1}{2}}$ la costante di diffusione ha la dimensione

$$[\sigma] = L^2 T^{-1}$$

Convieni far figurare sempre la sostanza di diffusione nelle equazioni che seguono al fine di mantenere il controllo delle dimensioni fisiche.

4.4.1. Metodi di soluzione dell'equazione di Langevin. Prima di vedere come si può risolvere in pratica un'equazione Eq.(56), osserviamo che, secondo un teorema di cui non daremo qui dimostrazione, il processo $x(t)$ è caratterizzato da una densità di probabilità $P(x, t)$ la cui evoluzione nel tempo è descritta dall'“equazione di Chapman-Kolmogorov” (si veda [PR69], §4, oppure la Sez. A.8)

$$(60) \quad \frac{\partial P(x, t)}{\partial t} = \frac{1}{2} \Delta (\sigma(x) P(x, t)) - \nabla \cdot (b(x) P(x, t))$$

Ora, c'è una classe particolare di possibili funzioni $b(x)$ (denominata come “drift”) che permette di ricavare la soluzione asintotica per t grande: è il caso in cui $b(x)$ è un campo irrotazionale e nel contempo σ è una costante. Si ha infatti che, posto $b(x) = -\nabla \rho(x)$, il membro di destra dell'equazione diventa

$$\begin{aligned} & \frac{1}{2} \Delta (\sigma P(x, t)) + \nabla \cdot (\nabla \rho(x) P(x, t)) \\ &= \nabla \cdot \left(\frac{1}{2} \sigma \nabla P(x, t) + P(x, t) \nabla \rho \right) \end{aligned}$$

e quindi si annulla per $P = \exp\{-2\rho(x)/\sigma\}$. Ma allora è possibile costruire una EDS che ammetta P come soluzione stazionaria scegliendo il drift secondo la ricetta $b = -\nabla \rho$, $\rho = -\frac{1}{2}\sigma \log P$. Il caso che interessa in meccanica statistica e in teoria quantistica è quello in cui

$$P(x) = Z^{-1} \exp\{-\beta H(x)\}$$

e quindi

$$b(x) = -\nabla H(x), \quad \beta = 2/\sigma.$$

Dunque è possibile costruire un processo aleatorio $x(t)$ che è garantito convergere per tempi grandi a qualunque distribuzione stazionaria che si desideri. Questo permette quindi di utilizzare un'EDS come algoritmo di simulazione Montecarlo.

ESERCIZIO 74. Determinare la soluzione dell'equazione differenziale stocastica lineare (processo di Ornstein-Uhlenbeck)

$$dx(t) = -\lambda x(t) dt + dw(t).$$

Per semplicità assumiamo $\sigma = 1$. Si chiede cioè di trovare $\mathbb{E}[x(t)]$ e $\mathbb{E}[x(t)x(t')]$, valor medio e correlazione del processo diffusivo $x(t)$.

SOLUZIONE. L'equazione differenziale è lineare disomogenea, dunque possiamo applicare la soluzione generale $x(t) = x_{\text{omo}}(t) + x_{\text{part}}(t)$, essendo $x_{\text{omo}}(t)$ la soluzione generale dell'equazione omogenea $\dot{x} = -\lambda x$ mentre $x_{\text{part}}(t)$ è una qualunque soluzione particolare dell'equazione inhomogenea. Per x_{omo} si ha ovviamente

$$x_{\text{omo}}(t) = C e^{-\lambda t}$$

e una soluzione particolare si ottiene con il solito ansatz $x(t) = C(t) e^{-\lambda t}$. Troviamo così

$$\dot{C}(t) e^{-\lambda t} dt - \lambda x(t) dt = -\lambda x(t) dt + dw(t)$$

ossia

$$C(t) - C(0) = \int_0^t e^{\lambda \tau} dw(\tau)$$

da cui, scegliendo $C(0) = 0$,

$$x(t) = x(0) e^{-\lambda t} + \int_0^t e^{-\lambda(t-\tau)} dw(\tau)$$

Qui la “costante iniziale” $x(0)$ è una variabile aleatoria da specificare per potere studiare la soluzione. Come prima conclusione possiamo osservare che il processo $x(t)$ è gaussiano, in quanto combinazione lineare di $w(\tau)$, sempreché anche $x(0)$ sia una v.a. gaussiana. Dunque è sufficiente calcolare media e covarianza. Se immaginiamo che il processo parta da una posizione nota, ad es. da $x(0) = 0$, possiamo calcolare facilmente media e covarianza:

$$\mathbb{E}[x(t)] = \int_0^t e^{-\lambda(t-\tau)} \mathbb{E}[dw(\tau)] = 0$$

in quanto gli incrementi del processo Browniano hanno media nulla. Per la correlazione si trova (applicando l'Eq.(59))

$$\begin{aligned} \mathbb{E}[x(t)x(t')] &= \int_0^t e^{-\lambda(t-\tau)} \int_0^{t'} e^{-\lambda(t'-\tau')} \mathbb{E}[dw(\tau)dw(\tau')] \\ &= \int_0^{\min(t,t')} e^{-\lambda(t+\tau')} e^{2\lambda\tau} d\tau \\ &= \frac{1}{2\lambda} \left(e^{-\lambda|t-t'|} - e^{-\lambda(t+t')} \right). \end{aligned}$$

In particolare troviamo che la variabile aleatoria $x(t)$ ha una distribuzione gaussiana con media zero e varianza $\frac{1}{2\lambda} (1 - e^{-2\lambda t})$. Si tratta del processo che descrive una particella Browniana legata da una forza elastica a un punto fisso $x = 0$ ed è noto come processo di Ornstein-Uhlenbeck (si può vedere [Wax54]). ◀

ESERCIZIO 75. Risolvere l'equazione di Langevin nel caso di un drift costante come vale ad esempio nel caso di un moto browniano in campo di gravità (la caduta di un grave con attrito porta a una velocità limite costante). È il caso della sedimentazione di un pulviscolo le cui micro particelle sono soggette sia alla gravità che al moto di agitazione molecolare.

SOLUZIONE. L'equazione stocastica è data da

$$dx(t) = -b dt + \sqrt{\sigma} dw(t)$$

e la soluzione è semplicemente $x(t) = x(0) - bt + \sqrt{\sigma} w(t)$ ma si deve affrontare il problema delle condizioni al contorno. La caduta deve arrestarsi quando la particella raggiunge il fondo - in assenza di questo si avrebbe una caduta infinita che non corrisponde ad alcuna situazione fisica realistica. La soluzione si ottiene nel modo più semplice utilizzando la discretizzazione di tempo e spazio come mostrato da Mark Kac nella raccolta [Wax54, Kac46]. Si introduce un moto tipo random walk monodimensionale ma asimmetrico: ad ogni tempo discreto $n\tau$ la particella si sposta in alto di un passo Δ con probabilità $u = \frac{1}{2}(1 - \beta\Delta)$ o verso il basso con probabilità $d = \frac{1}{2}(1 + \beta\Delta)$. Se la particella raggiunge il fondo dove c'è la barriera, con probabilità 1 torna nel punto $+\Delta$. La matrice della catena di Markov così definita è data da

$$M = \begin{pmatrix} 0 & d & 0 & 0 & 0 & \dots \\ 1 & 0 & d & 0 & 0 & \dots \\ 0 & u & 0 & d & 0 & \dots \\ 0 & 0 & u & 0 & d & \dots \\ \vdots & & & \ddots & \ddots & \ddots \end{pmatrix}$$

Kac analizza l'equazione agli autovalori per M e trova lo spettro; l'autovettore appartenente all'autovalore 1, e cioè la distribuzione stazionaria, è data da $p_n \propto (u/d)^n$ per $n > 1$ e quindi si ha una distribuzione esponenziale che si mantiene come carattere anche andando al limite continuo. Per esplorare in maggiore dettaglio l'evoluzione del processo si può utilizzare `matlab` che permette di costruire la Matrice M anche di grandi dimensioni. Dallo spettro di M si può dedurre il tempo medio per raggiungere l'equilibrio $T \approx 1/|\log(\lambda_{\max})|$ essendo λ_{\max} il più grande autovalore minore di uno. Si verifica facilmente che la distribuzione asintotica ottenuta applicando M molte volte a un qualunque vettore iniziale decade esponenzialmente alla distribuzione di equilibrio, a patto che u sia strettamente minore di d ; infatti nel limite $u \rightarrow d$ il processo tende a diffondere all'infinito e non raggiunge nessuna distribuzione di equilibrio. ◀

ESERCIZIO 76. Calcolare esattamente la distribuzione di equilibrio del problema precedente.

SOLUZIONE. L'equazione $Mx = x$ si traduce in infinite equazioni lineari accoppiate

$$\begin{aligned} dx_1 &= x_0 \\ x_0 + dx_2 &= x_1 \\ ux_1 + dx_3 &= x_2 \\ ux_2 + dx_4 &= x_3 \\ &\dots \dots \dots \\ ux_{n-1} + dx_{n+1} &= x_n \\ &\dots \dots \dots \end{aligned}$$

L'equazione si può risolvere con un ansatz $x_n = \alpha z^n$. Inserendo questa forma nell'equazione generica si trova $uz^{n-1} + dz^{n+1} = z^n$ che impone $dz^2 - z + u = 0$ ossia, ricordando che $u + d = 1$, $z = u/d$ - l'altra soluzione $z = 1$ porta a una soluzione non normalizzabile. Si controlla che le prime due equazioni che riguardano le condizioni al contorno sono soddisfatte con $x_0 = \alpha u$. Si determina poi in modo univoco il valore di α imponendo che $\sum p_n = 1$. ◀

ESERCIZIO 77. Limite continuo del random walk asimmetrico. Come abbiamo visto nell'esercizio a pag. 84 si può prendere il limite per il passo reticolare a che tende a zero e mantenendo fisso il rapporto $a^2/\tau = D$. In questo limite la probabilità di transizione soddisfa un'equazione alle derivate parziali secondo quanto si deduce dalle formule seguenti:

$$(61) \quad P(t + \tau, na) = \frac{1}{2} (1 - \beta a) P(t, na - a) + \frac{1}{2} (1 + \beta a) P(t, na + a)$$

$$\begin{aligned} \frac{P(t + \tau, x) - P(t, x)}{\tau} &= \\ \frac{a^2 (1 - \beta a) P(t, x - a) - 2P(t, x) + (1 + \beta a) P(t, x + a)}{\tau 2a^2} &= \\ \frac{1}{2} D \frac{P(t, x + a) - 2P(t, x) + P(t, x - a)}{a^2} + \beta D \frac{P(t, x + a) - P(t, x - a)}{2a} \end{aligned}$$

Al limite, i rapporti tendono alle derivate parziali e si raggiunge l'equazione di diffusione modificata con un termine del primo ordine che deriva dalla asimmetria tra u e d . Questo termine rappresenta la forza di gravità che agisce sulla particella browniana:

$$(62) \quad \frac{\partial P(t, x)}{\partial t} = \frac{1}{2} D \frac{\partial^2 P(t, x)}{\partial x^2} + \beta D \frac{\partial P(t, x)}{\partial x}.$$

Per la distribuzione di equilibrio $\partial P/\partial t = 0$ si trova

$$(63) \quad 0 = D \frac{\partial}{\partial x} \left(\frac{1}{2} \frac{\partial P(t, x)}{\partial x} + \beta P(t, x) \right)$$

da cui $P_{eq}(x) \propto \exp\{-2\beta x\}$.

4.4.2. Equazione di Langevin e Meccanica Quantistica. Vi sono almeno due contesti in cui le EDS sono utili nello studio della Meccanica Quantistica. Nelson [Nel67] ha mostrato che l'equazione di Schroedinger è matematicamente equivalente ad una famiglia di processi diffusivi, governati da EDS con

$$\sigma \propto \sqrt{\hbar}, \quad b(x, t) = \nabla \Re\{\log(\psi(x, t))\}.$$

Da un punto di vista pratico, questo fatto non comporta grossi benefici, in quanto anche solo per scrivere l'EDS è necessario conoscere la soluzione dell'Equazione di Schroedinger o un suo sistema equivalente di equazioni non lineari. L'approccio di Nelson si comprende facilmente da questo punto di vista: supponiamo che l'Hamiltoniana sia data da

$$H = \frac{1}{2} D^\dagger D, \quad D = e^{-\varphi(x)} \frac{d}{dx} e^{\varphi(x)} = \frac{d}{dx} + \varphi'(x)$$

Ovviamente H ammette la funzione $e^{-\varphi(x)}$ come il suo stato fondamentale con energia zero. Infatti $D e^{-\varphi(x)} = 0$ e $\langle \psi | H \psi \rangle = |D \psi|^2 \geq 0$. Esplicitamente

$$H = -\frac{1}{2} \left(\frac{d}{dx} \right)^2 + \frac{1}{2} (\varphi'(x)^2 - \varphi''(x))$$

dunque siamo in presenza di un operatore di Schroedinger con energia potenziale dipendente da $\varphi(x)$. L'equazione di Schroedinger dipendente dal tempo adottando un tempo immaginario $\tau = t/(i\hbar)$ diventa

$$\frac{\partial}{\partial \tau} \psi(x, \tau) = -H \psi$$

Ora trasformiamo l'equazione ponendo $\psi(x) = e^{-\varphi(x)} \rho(x, \tau)$; si ottiene immediatamente

$$\frac{\partial}{\partial \tau} \rho(x, \tau) = \frac{1}{2} \frac{\partial}{\partial x} \left(\frac{\partial}{\partial x} + 2\varphi'(x) \right) \rho(x, \tau) = \frac{1}{2} \left(\frac{\partial}{\partial x} \right)^2 \rho(x, \tau) - \frac{\partial}{\partial x} (b(x) \rho(x))$$

ossia l'equazione di Schroedinger in tempo immaginario è legata da una trasformazione di similitudine all'equazione di Fokker-Planck di un processo diffusivo con drift dato dalla derivata logaritmica della funzione d'onda dello stato fondamentale¹². La cosa si può sfruttare per valutare il gap di energia attraverso lo studio del processo diffusivo e della sua correlazione $\mathbb{E}(x(\tau)x(\tau'))$, come spiegheremo tra poco¹³. Il grosso difetto di questo schema è costituito dal fatto che per iniziare a impostare il calcolo numerico è necessario conoscere la funzione d'onda dello stato fondamentale e il guadagno consiste nel trovare la stima del gap di energia. Ma lo stato fondamentale *non* è a disposizione nei casi più interessanti, ad esempio nel caso di una molecola semplice quale quella di ammoniaca. Di conseguenza l'approccio di Nelson è ormai largamente dimenticato se non per ricerche relative ai fondamenti della Meccanica Quantistica (si veda ad es. [GGAT78]).

L'altra strategia, di applicazione universale, è invece quella della “quantizzazione stocastica” nel senso di Parisi-Wu [PW81, DH87]. Il nocciolo della questione è dato dal fatto che l'EDS con drift $b(x) = -\nabla H(x)$ permette di valutare medie statistiche rispetto alla misura di Gibbs $\exp(-\beta H(x))$, e inoltre la Meccanica Quantistica nella formulazione Euclidea (tempo immaginario) si riduce alla Meccanica Statistica classica¹⁴. Per l'equazione di Schroedinger di una particella scalare in un grado di libertà e energia potenziale $V(x)$ la funzione $H(x)$ si scrive

$$(64) \quad H(x) = \int_{-\infty}^{\infty} \left(\frac{m}{2} \dot{x}^2 + V(x(t)) \right) dt$$

¹²Per l'oscillatore armonico quantistico il corrispondente processo diffusivo coincide con quello di Ornstein-Uhlenbeck (vedi Es.74)

¹³Per esperienza personale è facile confondersi con il segno della relazione $b = \pm \nabla \log \rho$. Se l'equazione differenziale stocastica è fissata dall'Eq.(56), nel caso della funzione $\psi(x) = \exp\{-\lambda x^2\}$ si ha $b(x) = -\varphi'(x) = -2\lambda x$ e l'equazione esprime l'attrazione verso l'origine $dx = -2\lambda x dt + dw$. Un segno sbagliato darebbe $dx \propto +x$ con conseguente esplosione $x(t) \sim e^{\lambda t}$.

¹⁴Un lavoro fondamentale dove si illustra la corrispondenza nel contesto della teoria dei campi è [FLB75]

e l'equazione di Langevin assume la forma ¹⁵

$$\frac{\partial x(t, \tau)}{\partial \tau} = m \frac{\partial^2 x(t, \tau)}{\partial t^2} - V'(x(t)) + \sqrt{\sigma} \eta(t, \tau)$$

Si deve porre attenzione al fatto che vi sono due “tempi” coinvolti: il t che definisce l'azione del sistema meccanico classico nell'Eq.(64) e il tempo τ del processo diffusivo di Langevin. Quest'ultimo è un tempo “tecnico” dell'algoritmo, né più né meno che l'indice crescente in un processo di Metropolis che scandisce le varie configurazioni. Il processo deve evolvere fino a che la distribuzione di probabilità è abbastanza vicina a quella di equilibrio, che per costruzione è la distribuzione di Gibbs $e^{-\beta H}$. Sta a chi effettua l'esperimento regolare i parametri del programma (tempo di evoluzione di Langevin, passo discreto nella approssimazione numerica dell'equazione differenziale stocastica) in modo da ottimizzare l'approssimazione e il tempo necessario per l'esperimento (questo è il tempo dell'orologio dello sperimentatore).

Per risolvere numericamente l'equazione che governa il processo di Langevin $x(t, \tau)$ si passa in genere ad una approssimazione discreta, introducendo un reticolo $t_n = n a, a = T/N, n = 0, 1, 2, \dots, N$ adottando poi un'algoritmo di soluzione, il più semplice essendo quello di Eulero:

$$(65) \quad x(t_n, \tau + \Delta\tau) \approx x(t_n, \tau) + a h^{-1} \Delta\tau \left(\frac{m}{2} \frac{x(t_{n+1}) - 2x(t_n) + x(t_{n-1}))}{a^2} - V'(x(t_n)) \right) + \sqrt{\sigma \Delta\tau / h} dw(t_n, \tau)$$

Il risultato che si ottiene da questa approssimazione è affetto da un errore sistematico $O(\Delta\tau)$ ¹⁶. Per rimediare a questo errore si può cercare di estrapolare al limite $\Delta\tau \rightarrow 0$ oppure di adottare algoritmi più accurati che sono disponibili anche per le equazioni differenziali stocastiche, il più semplice essendo quello dell'integrazione con il “metodo dei trapezi”.

Una volta impostato l'algoritmo di integrazione numerica dell'equazione di Langevin si può sfruttare la simulazione del processo *per ottenere informazioni sullo spettro di energia e su elementi di matrice quantistici dallo studio di funzioni di correlazione* del processo diffusivo. Ad es. si ha (vedi appendice A.9)

$$(66) \quad \mathbb{E}[x(t)x(s)] = \sum_E |\langle E_0 | q | E \rangle|^2 e^{-(E-E_0)|t-s|/\hbar}$$

dove q è l'osservabile posizione in meccanica quantistica e la somma è estesa allo spettro dell'Hamiltoniana. Studiando quindi l'andamento per $|t-s|$ grande il segnale è dominato dal primo livello eccitato $\mathbb{E}[x(t)x(s)] \approx |\langle E_0 | q | E_1 \rangle|^2 e^{-(E_1-E_0)|t-s|/\hbar}$ e dalla pendenza della curva di correlazione in scala logaritmica si ottiene il gap di energia. Questa tecnica offre la possibilità di stimare le proprietà spettrali di sistemi quantistici complessi per i quali un calcolo diretto attraverso l'equazione di Schroedinger non sarebbe possibile. Ovviamente l'approccio Montecarlo introduce necessariamente degli errori che devono essere valutati e ridotti al minimo. Ma questo riguarda un'altra serie di lezioni. Per ulteriori dettagli sull'argomento si veda l'Appendice A.8 e per iniziare con una realizzazione pratica dell'algoritmo si possono vedere i codici matlab su [dropbox/MPF2023](https://eoinfnr.it) in `~/Matlab/Montecarlo/Langevin`. Su <https://eoinfnr.it> si possono poi trovare le relazioni finali dei partecipanti al corso tra cui un rapporto sull'applicazione dell'equazione di Langevin/Parisi. In Fig.4.5 il risultato di un esperimento che utilizza la quantizzazione stocastica nel caso dell'oscillatore anarmonico $V = \frac{1}{2} m \omega^2 x^2 + g x^4$. Si trova la stima del gap per valori del passo temporale $\tau = 0.02, 0.01, \dots, 0.00125$.

4.4.3. Un esperimento nel caso più semplice: LangHO. L'esperimento più semplice che possiamo concepire per familiarizzarci con il metodo di Langevin-Parisi-Wu consiste nello studiare il gap di energia dell'oscillatore armonico quantistico. Il codice si trova su [dropbox](https://eoinfnr.it) in `~/Matlab/Montecarlo/Langevin`.

Per mettere a punto l'esperimento è necessario fissare i parametri fisici (costante di Planck, massa, frequenza dell'oscillatore armonico) e alcuni parametri tecnici dell'algoritmo:

- (i) L , l'intervallo finito in t per $x(t, \tau)$; in teoria dovrebbe essere $L = \infty$, ma sul computer ciò non è permesso;
- (ii) N , il numero di intervalli in cui suddividere L : L/N è detto “passo reticolare”;

¹⁵Si ricordi che la *derivata funzionale* da applicare è la stessa che conduce alle equazioni di Lagrange a partire dall'azione $S = \int dt L(\dot{x}, x)$, cioè $\delta S / \delta x(t) \equiv \partial L / \partial x - \frac{d}{dt} (\partial L / \partial \dot{x})$.

¹⁶La ragione è che il termine diffusivo Browniano $dw(t)$ è di ordine \sqrt{dt} .

- (iii) *T*: il tempo di evoluzione dell'equazione di Langevin; deve essere abbastanza lungo per raggiungere la distribuzione di equilibrio;
- (iv) *steps*: in numero di passi discreti in cui è suddiviso *T*;
- (v) *skip*: la misura di correlazione sarà presa ogni “skip” steps in quanto l'evoluzione di Langevin è continua e le configurazioni devono essere decorrelate prima di accumulare statistica;
- (vi) *therm*: lunghezza del ciclo di termalizzazione; si immagina che in capo a questo tempo la distribuzione sia già vicina all'equilibrio
- (vii) *exps*: il numero di esperimenti che si effettuano al fine di valutare le fluttuazioni statistiche

Il programma `LangH0.m` realizza un esperimento completo di simulazione di Langevin per calcolare il gap di energia nel caso dell'oscillatore armonico, problema che è ovviamente accademico ma serve a fissare le modalità di programmazione. L'output è tipicamente quello della figura 4.5 che corrisponde ai dati di default che si trovano nel programma e forniscono la stima $E_1 - E_0 = 0.991 \pm 0.006$.

4.4.4. Il calcolo della correlazione $\langle x(t)x(s) \rangle$. In un lavoro di Parisi del 1981 [Par81] si trova una tecnica molto efficiente per il calcolo delle correlazioni. Si tratta di sfruttare un trucco ben noto nella quantizzazione dei campi attraverso i path integrals. Una modifica della funzione di energia nel fattore di Gibbs (a $\beta = 1$)

$$(67) \quad \exp\{-H(x)\} \rightarrow \exp\{-H(x) + \lambda x_j\}$$

per un valore fissato *j* permette di esprimere la correlazione connessa $\langle x_i x_j \rangle_c$ come

$$(68) \quad \langle x_i x_j \rangle_c = \frac{d}{d\lambda} \langle x_i \rangle|_{\lambda=0}$$

L'equazione di Langevin contiene un termine dipendente da λ e viene naturale sviluppare la soluzione $x_i(t, \lambda) = x_i^{(0)}(t) + \lambda x_i^{(1)}(t)$ di modo che l'equazione diviene un sistema di EDS accoppiate. Indicando con $b_i(x) = -\frac{\partial H}{\partial x_i}$ il drift, si troverà

$$(69) \quad \begin{aligned} dx_i^{(0)}(t) &= b_i(x)dt + dw_i(t) \\ dx_i^{(1)}(t) &= \sum_k \frac{\partial b_i(x)}{\partial x_k} x_k^{(1)} dt + \delta_{ij} dt \end{aligned}$$

Per l' Eq.(68) si potrà allora calcolare la correlazione $\langle x_i x_j \rangle_c$ semplicemente dal valore medio di $x_i^{(1)}$, il che risulta non solo molto più agevole del calcolo diretto ma anche meno affetto da fluttuazioni statistiche, come mostrato nel lavoro di Parisi già citato.

Il metodo si presta anche a valutare attraverso la simulazione di Langevin il valore del secondo stato eccitato introducendo la sorgente x_j^2 anziché x_j . Infatti la correlazione $\langle x_j^2 x_i^2 \rangle$ nel caso di un'energia potenziale simmetrica ($V(x) = V(-x)$) non avrà contributi dal primo stato eccitato per questioni di simmetria di parità e il contributo dello stato fondamentale $\langle 0|x_i^2|0 \rangle$ si cancella nella correlazione connessa $\langle x_i x_j \rangle - \langle x_i \rangle \langle x_j \rangle$. Nei sistemi più complessi dotati di simmetrie geometriche più ricche della semplice parità si individueranno osservabili dotate di specifiche proprietà di trasformazione sotto simmetria in modo da filtrare stati di simmetria definita. Il caso della molecola di ammoniaca NH_3 avente la simmetria S_3 si presterebbe a questo tipo di sviluppi, che però esulerebbero dagli obiettivi di questo corso. Nel diagramma di Fig.B.1 si riporta il risultato di un run del programma `LangParFor12.m` consistente in 16 esperimenti condotti in parallelo, 1024 punti di reticolo temporale euclideo e valutazione della slope per le correlazioni di $x(t)$ e $x(t)^2$, confrontate con il calcolo diretto con l'algoritmo di Arnoldi (`eigs`). Ogni esperimento consiste di 2^{21} steps di Langevin per un tempo complessivo $T = 256$, a $g = 0.25$ e la slope è stimata nell'intervallo $[T/128, T/64]$.

ESERCIZIO 78. Scrivere il codice Langevin/Parisi per l'Hamiltoniano

$$H = \frac{1}{2} (p_1^2 + p_2^2) + \frac{1}{2} (x_1^2 + x_2^2) + g (x_1^2 + x_2^2)^2$$

e calcolare il gap di energia confrontandolo con il risultato ottenuto utilizzando il codice `schr2D.m` ottenibile via `dropbox`.

SOLUZIONE. Il codice `matlab` `LangParFor2D.m` si trova su `dropbox` e fa uso del `parallel toolbox` per eseguire in parallelo `Nexp` esperimenti, con `Nexp` multiplo delle macchine virtuali disponibili sul processore. Per ridurre errori statistici e sistematici si devono scegliere con cura i parametri “tecnici” del programma, l’intervallo di tempo euclideo T , suddiviso in N intervalli, scegliendo T il più grande possibile compatibilmente con mantenere piccolo il passo temporale T/N ; il tempo di evoluzione di Langevin suddiviso in “ $Nsteps$ ”, tenendo conto che il processo deve sostenersi fino a raggiungere la distribuzione di equilibrio ma lo step temporale di Langevin produce un errore sistematico che va tenuto piccolo aumentando il numero di steps. Tutti gli altri parametri sono scelti attraverso un menu gestito con `inputdlg`. In Fig. 4.6 il risultato di 16 esperimenti condotti in parallelo con $g = 0.1$. ◀

ESERCIZIO 79. Come nell’esercizio precedente ma nel caso di tre dimensioni. Il potenziale è $V(x, y, z) = \frac{1}{2} r^2 + g r^4/4$ con $r^2 = x^2 + y^2 + z^2$.

ESERCIZIO 80. Scrivere il codice Langevin/Parisi per l’Hamiltoniano ([Sim83])

$$H = \frac{1}{2} (p_x^2 + p_y^2) + x^2 y^2$$

e calcolare il gap di energia confrontandolo con il risultato ottenuto utilizzando il codice `schr2D.m` ottenibile via `dropbox`

Osservazione. Un controllo sulla correttezza del calcolo del gap di energia basato sul Langevin-Parisi è sempre opportuno farlo tante sono le possibilità di errore nella molteplicità di parametri da fissare nell’esperimento. Uno strumento quasi sempre a disposizione è quello che offre la teoria delle perturbazioni. In tutti i libri di testo di Meccanica Quantistica si trovano le istruzioni su come calcolare lo spettro di energia in via approssimata sfruttando la presenza di un parametro, detto “costante di accoppiamento”. Per il valore zero del parametro lo spettro deve essere noto esattamente e le formule di teoria delle perturbazioni esprimono lo spettro attraverso una serie di potenze nella costante di accoppiamento. Nell’esercizio precedente il parametro è indicato con g e per $g = 0$ il potenziale si riduce all’oscillatore armonico. Le formule standard di teoria delle perturbazioni sono facilmente applicabili fino al contributo $O(g^2)$ e non sempre costituiscono un calcolo agevole. L’utilizzo di un linguaggio simbolico può alleviare la fatica di portare in fondo il calcolo e in taluni casi può anche permettere di arrivare a ordini perturbativi ben al di là del secondo ordine. Per fare questo si utilizza una formulazione ricorsiva come descritta in [OD96]. L’algoritmo ricorsivo si presta a una implementazione efficiente nel linguaggio `form` di *J. Vermaseren* [Vea22]. Nel caso di un semplice oscillatore armonico perturbato con un termine quartico $g r^4/4$ il codice è dato da `pertth.frm` rintracciabile su `dropbox` e riportato in appendice. Per il caso dell’esempio precedente (oscillatore in tre gradi di libertà) si può procedere analogamente e i primi termini dello sviluppo perturbativo, ad es. fino a $O(g^{10})$, si ottengono in una frazione di secondo. Per valori piccoli di g questo permette di validare il programma di simulazione stocastica ed escludere errori banali di codifica. Volendo sfruttare appieno la serie perturbativa ad alti ordini, che come è ben noto non converge, si può utilizzare la tecnica di Padé che consiste nel trasformare la serie di potenze in una funzione razionale $P(g)/Q(g)$ che meglio si presta a rappresentare una funzione analitica singolare in $g = 0$. A questo fine si può utilizzare il programma `Mathematica` che offre una funzione “*PadéApproximant*” di facile utilizzo. Per i dettagli chiedere al vostro docente... Nel caso dell’oscillatore anarmonico attraverso questa tecnica si ottengono ottimi risultati per $g < 1$. Il codice `form` per generare lo sviluppo perturbativo si estende a sistemi con più gradi di libertà, la cosa essendo verificata fino a 6 d.o.f.

4.4.5. Galleria Montecarlo. Riportiamo alcuni diagrammi che si riferiscono alla simulazione Montecarlo con il metodo di Metropolis di modelli di spin elementari (Ising2D, Potts, etc.). Nei rapporti finali dei partecipanti al corso negli ultimi due anni si potranno trovare utili informazioni (<https://eoinfnpr.it>).

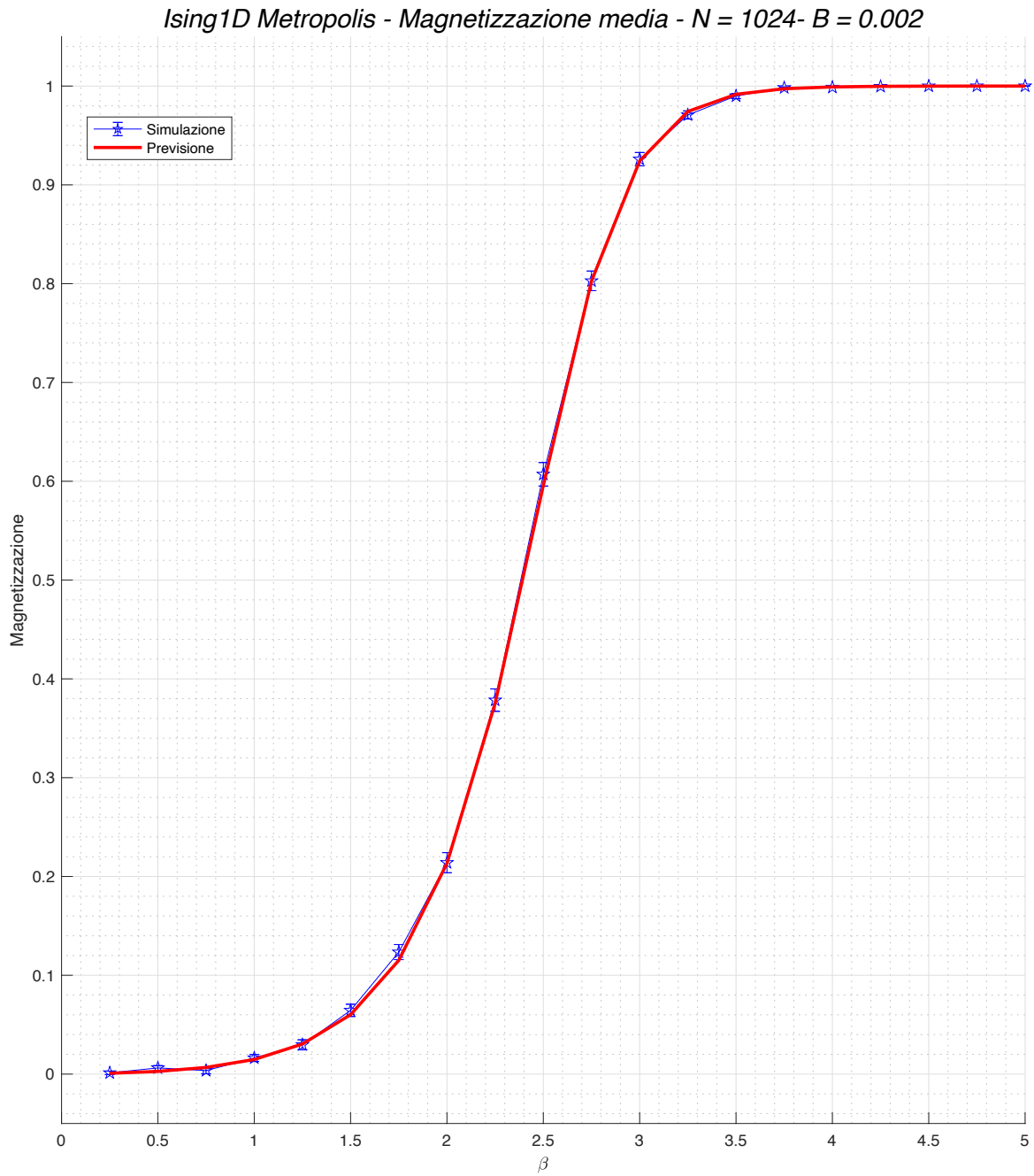


FIGURA 4.4. Un tipico esperimento di Metropolis per l'Ising1D.

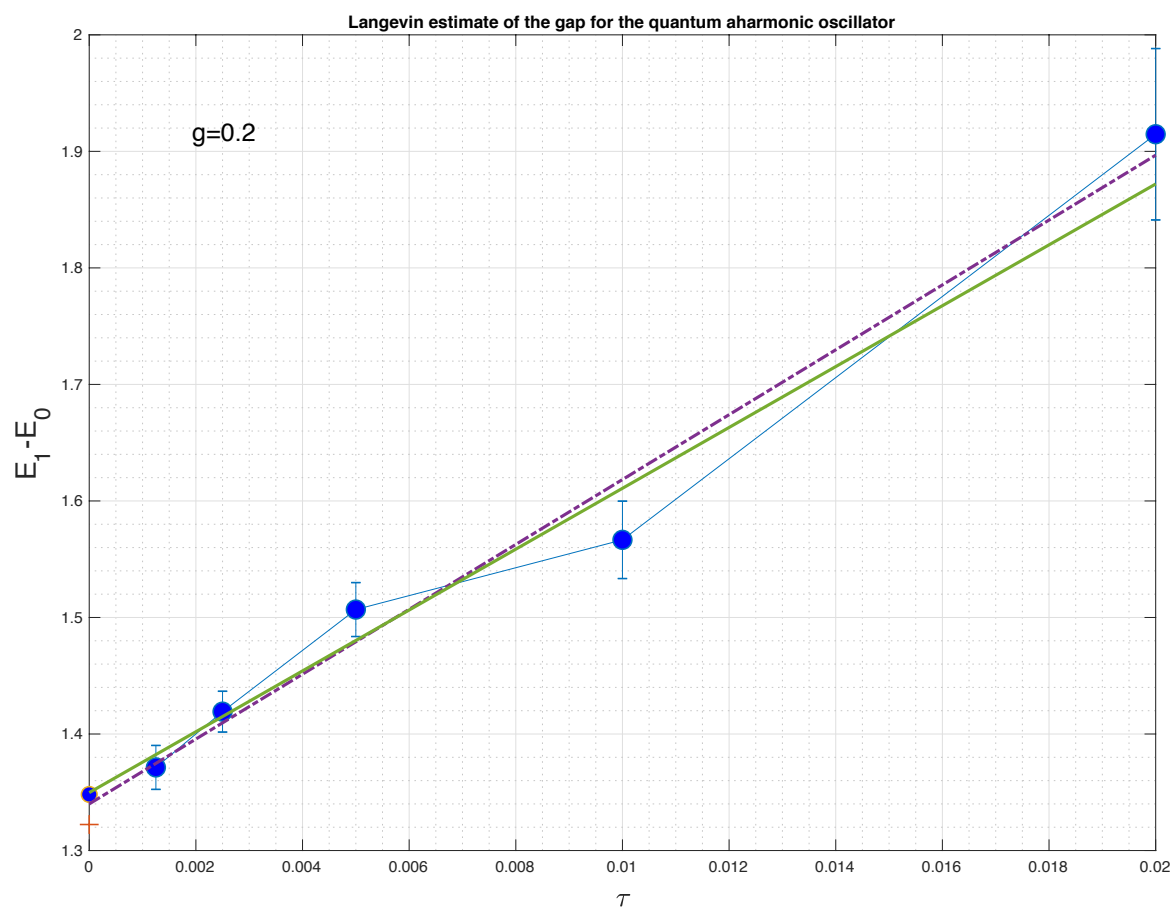


FIGURA 4.5. Un tipico esperimento di quantizzazione stocastica con estrapolazione a $\tau = 0$. La tecnica di fitting deve essere scelta con attenzione.

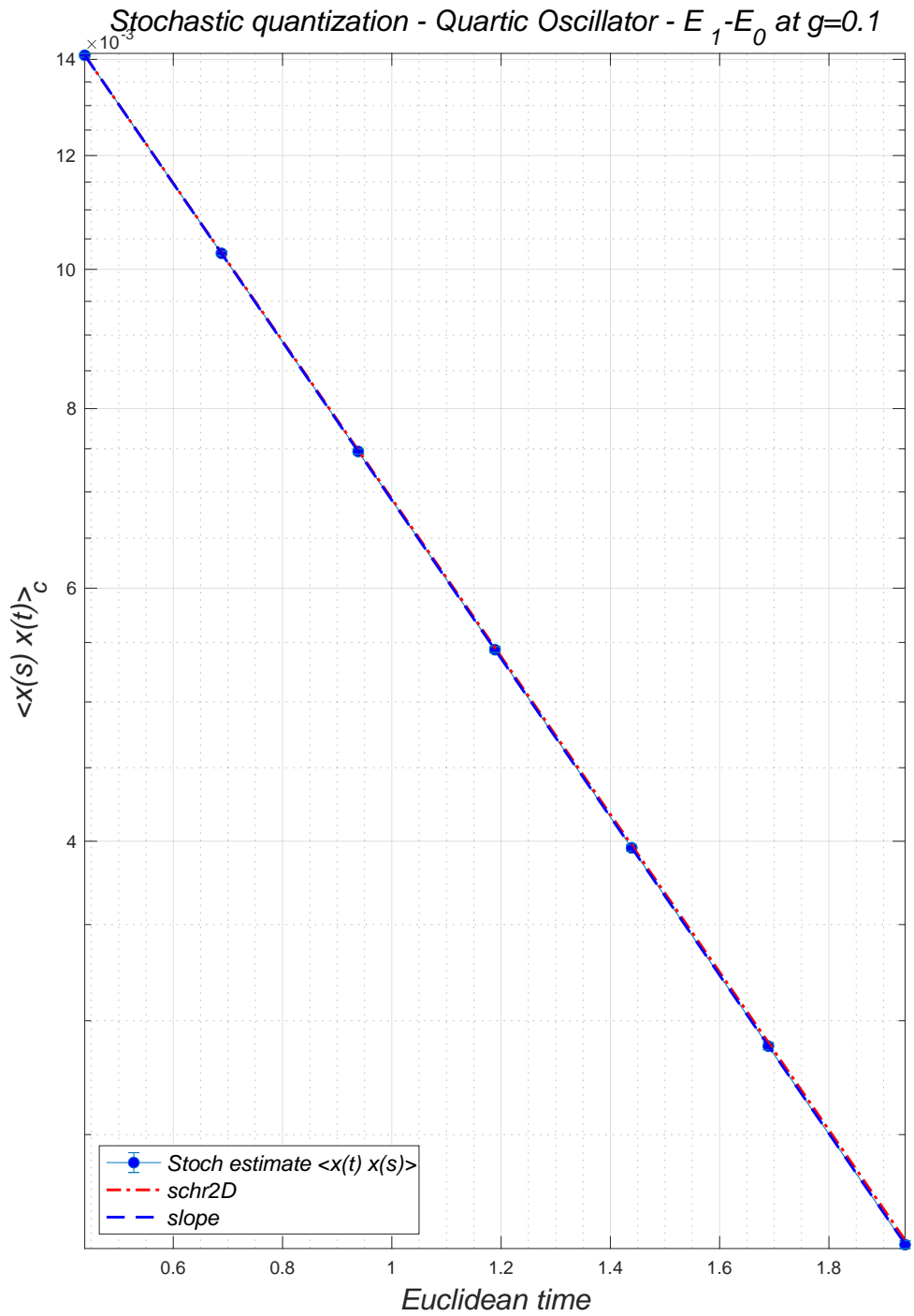


FIGURA 4.6. Applicazione della tecnica di Parisi al potenziale quartico in due gradi di libertà (testo a pag.118).

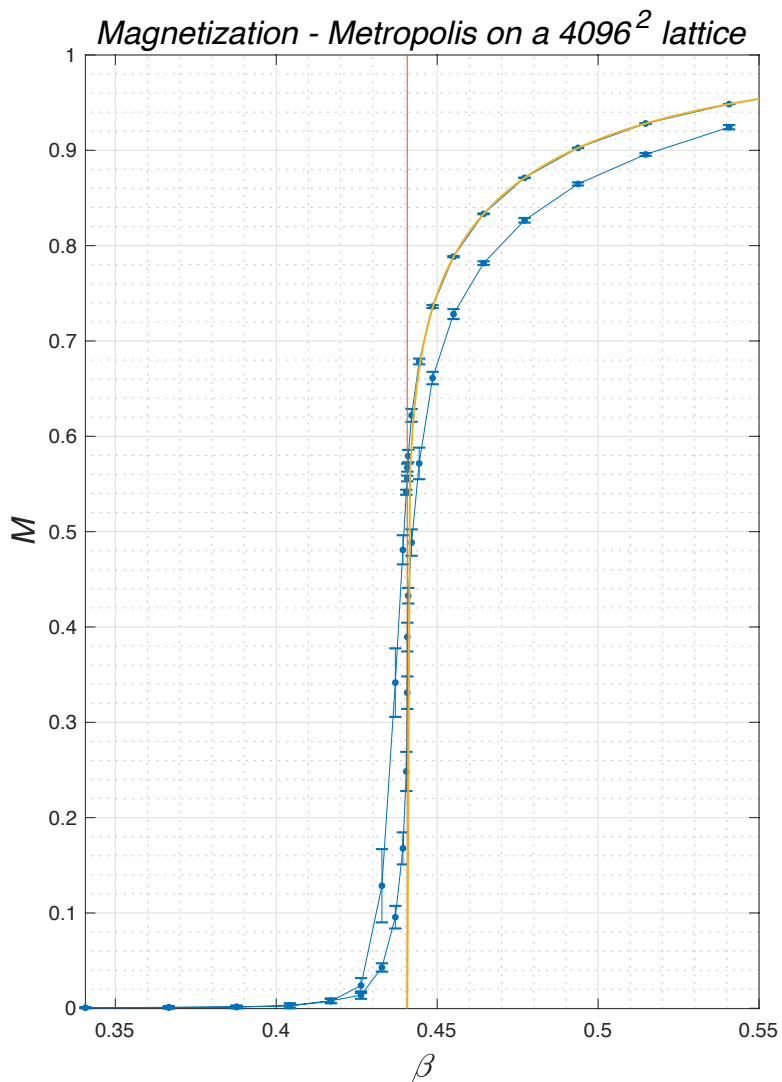


FIGURA 4.7. Misura della magnetizzazione per il modello di Ising in due dimensioni con campo esterno $B = 0$. Esperimento condotto partendo da spin tutti allineati a $\beta = .55$ e scendendo a valori più bassi di β per poi risalire al valore iniziale. Notare che nel primo ramo la magnetizzazione segue con grande precisione la formula di Onsager (curva rossa); si nota invece che nel ramo con β crescente i dati del MC tendono a discostarsi dalla curva esatta, la ragione essendo da cercare nella stabilità dei domini con spin negativo, un effetto dovuto all'algoritmo con mosse locali.

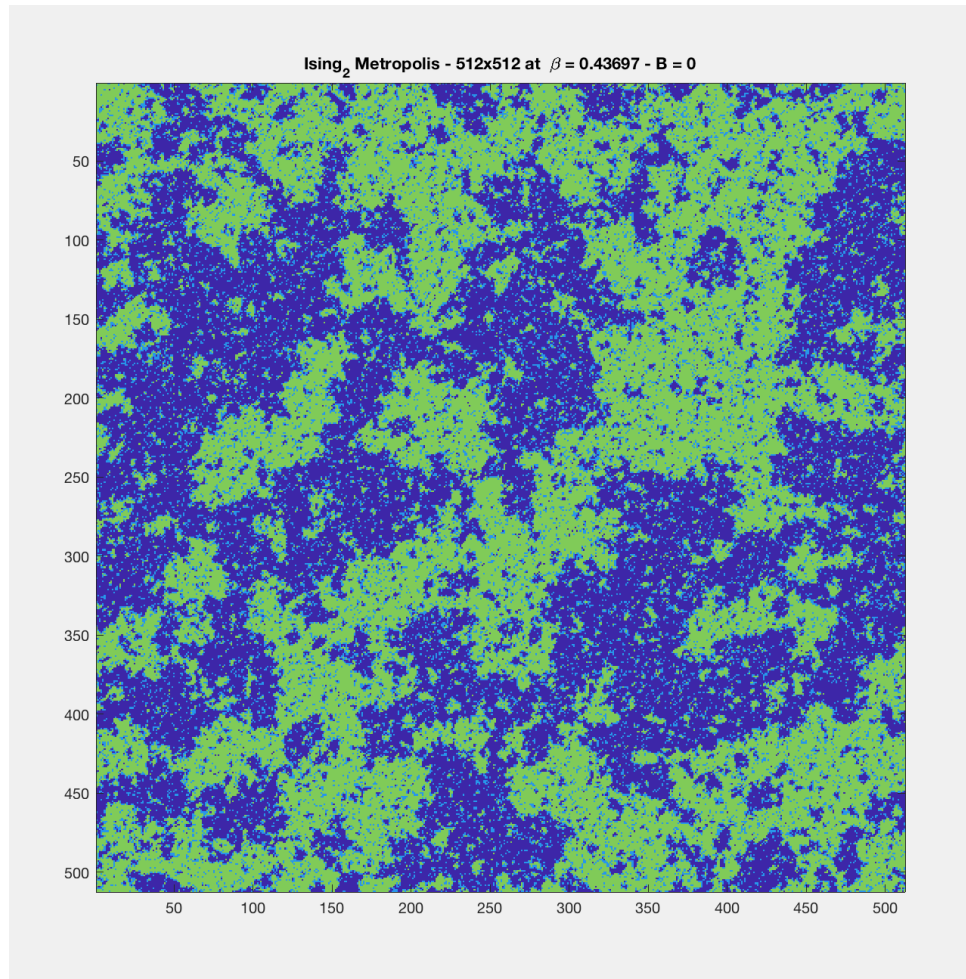


FIGURA 4.8. La formazione di domini di spin allineati per $\beta_c \sim .44$.

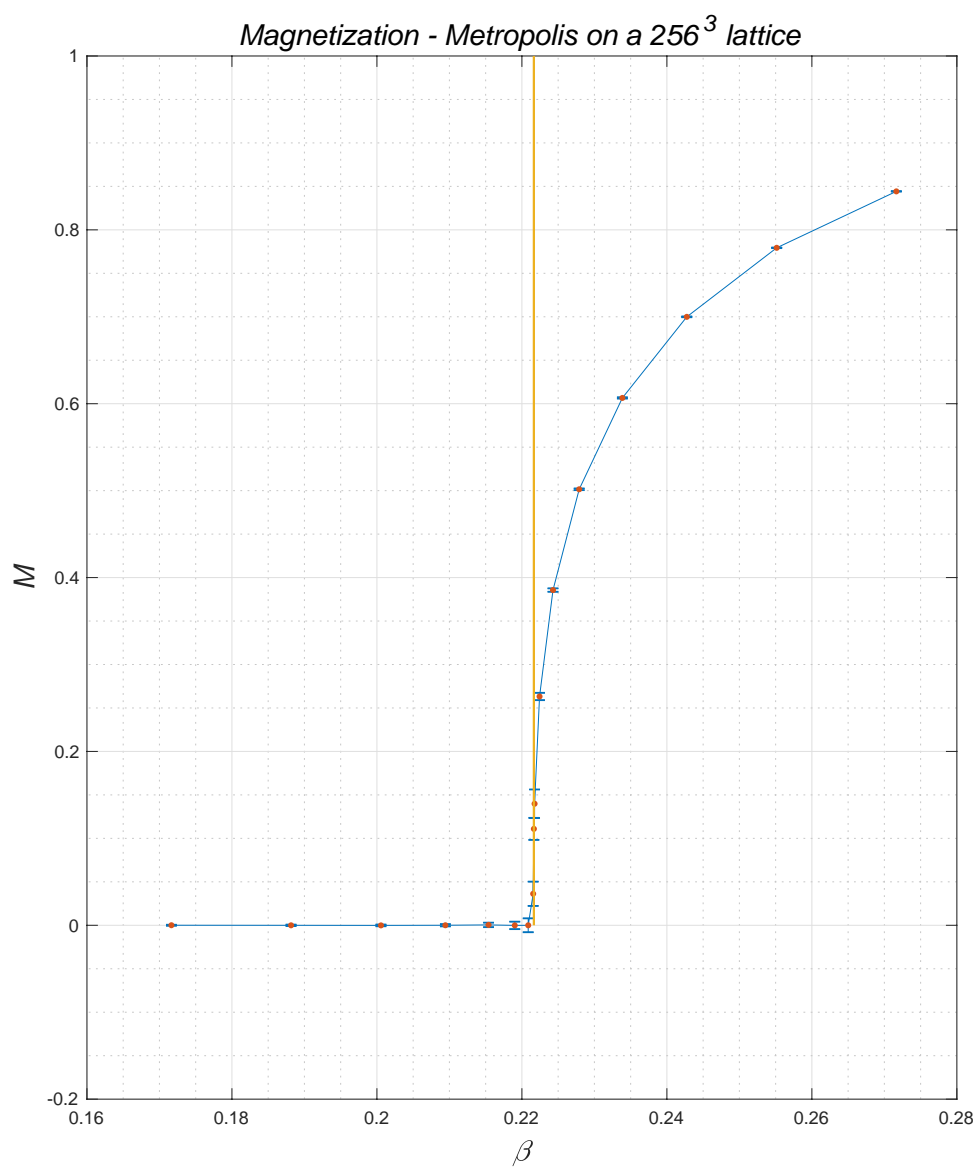


FIGURA 4.9. Misura della magnetizzazione per il modello di Ising in tre dimensioni a campo esterno nullo. La linea verticale indica la stima più precisa analitica del valore di β critico nota in letteratura [BP12] $\beta_c \approx 0.221652$.

om

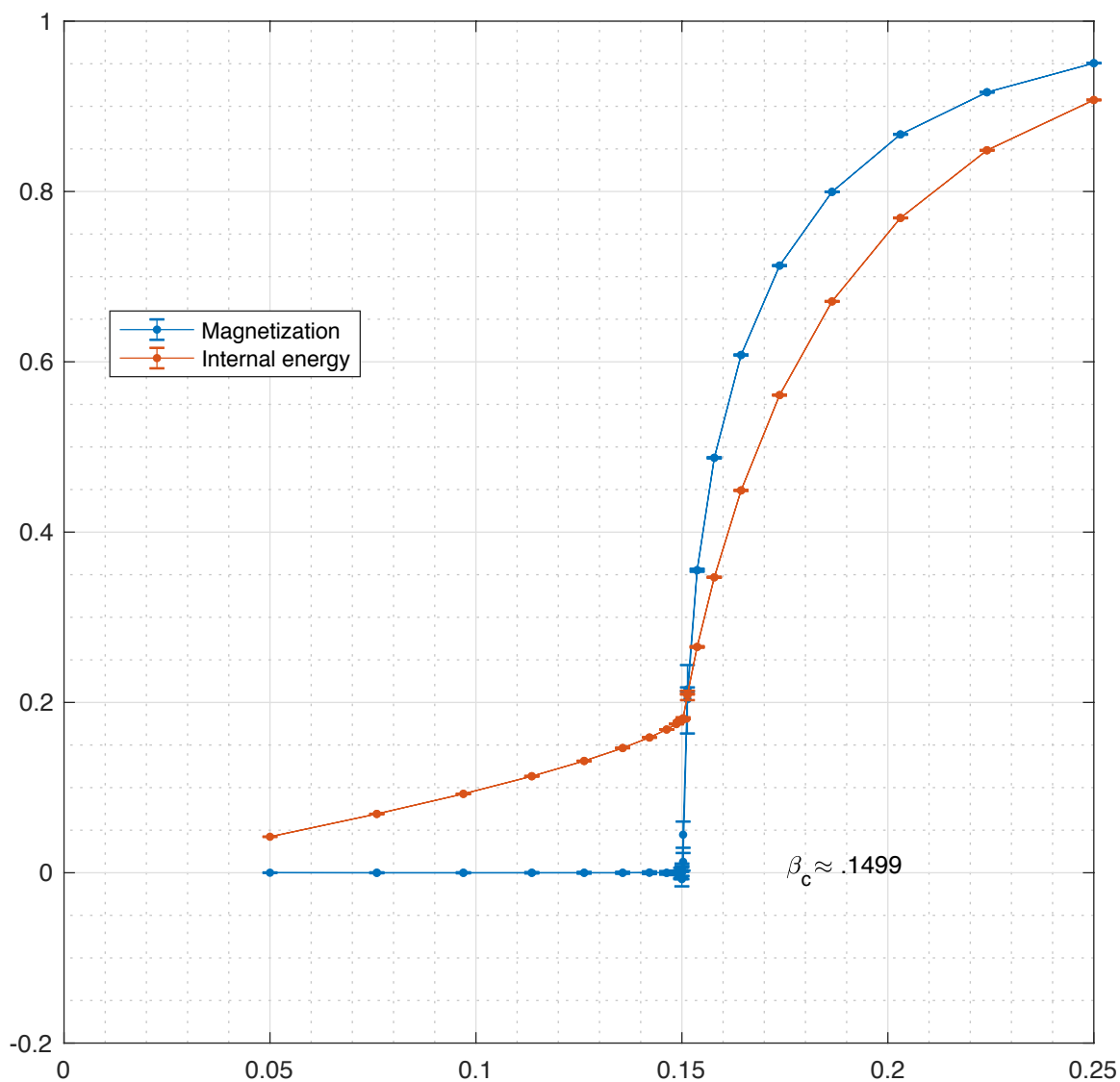


FIGURA 4.10. Misura della Magnetizzazione e dell'Energia interna per il modello di Ising in quattro dimensioni a campo esterno nullo. La linea verticale indica la stima più precisa analitica del valore di β critico nota in letteratura $\beta_c \approx 0.1499$.

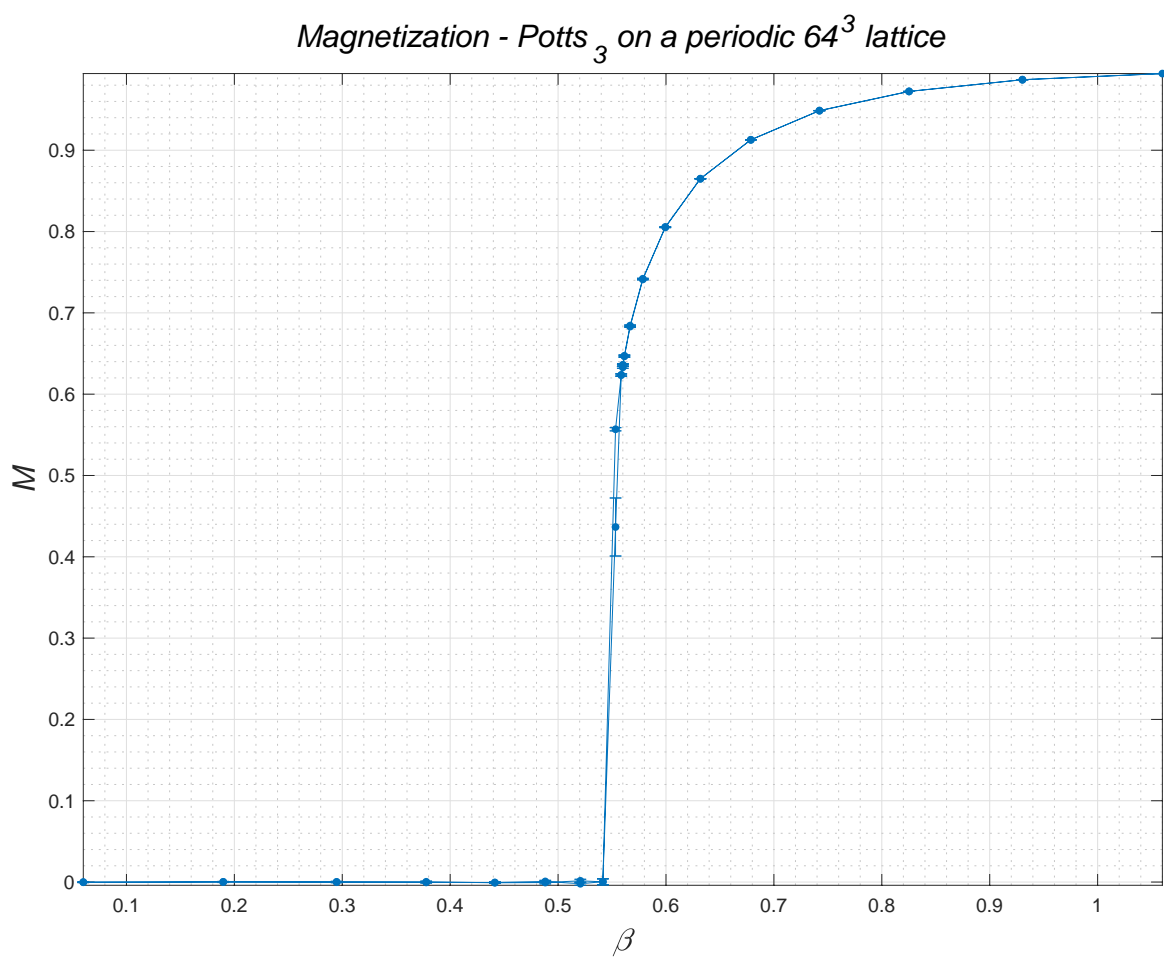


FIGURA 4.11. Magnetizzazione nel modello di Potts con $q = 3$ in 3 dimensioni. Il diagramma presenta un ciclo completo da β_{\max} a β_{\min} e ritorno.

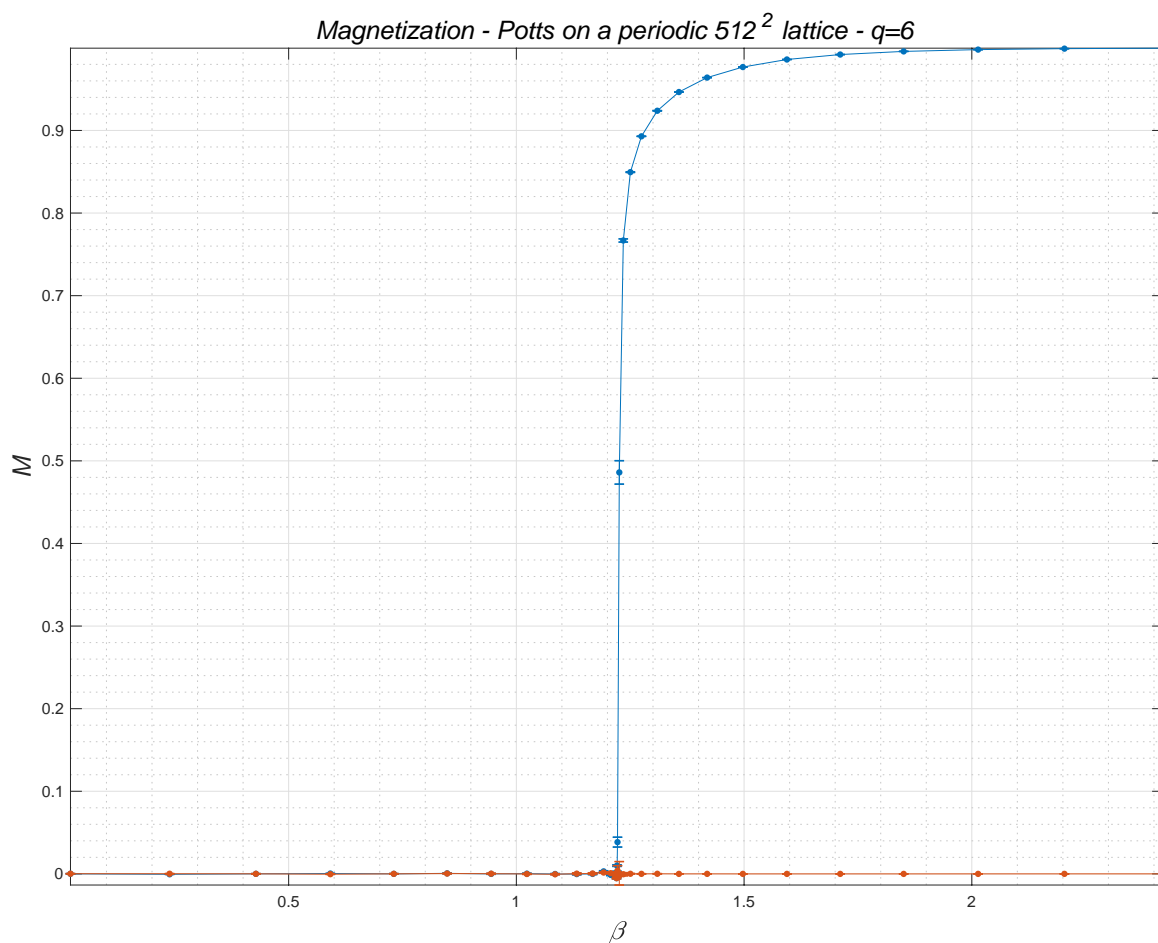


FIGURA 4.12. Magnetizzazione nel modello di Potts con $q = 6$ in 2 dimensioni studiato con un algoritmo Montecarlo di Metropolis. La curva in blu riporta i dati sulla componente degli spin lungo il campo esterno e la curva rossa la componente in direzione ortogonale. Il valore esatto di β_c è 1.23823....

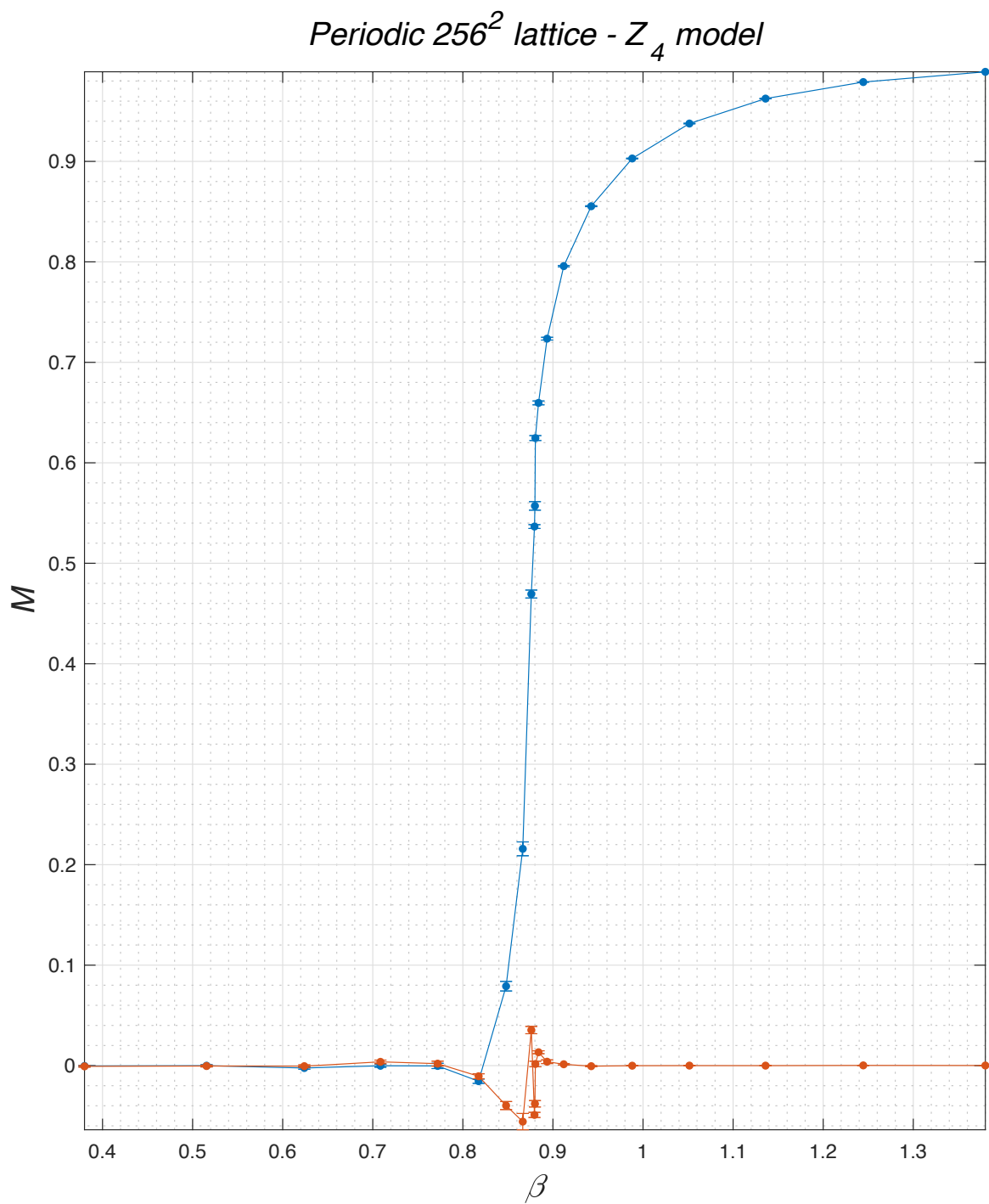


FIGURA 4.13. Magnetizzazione nel $Z(N)$ con $N = 4$ in 2 dimensioni studiato con un algoritmo Montecarlo di Metropolis. La curva in blu riporta i dati sulla componente degli spin lungo il campo esterno e la curva rossa la componente in direzione ortogonale. Il valore esatto di β_c è 0.88137....

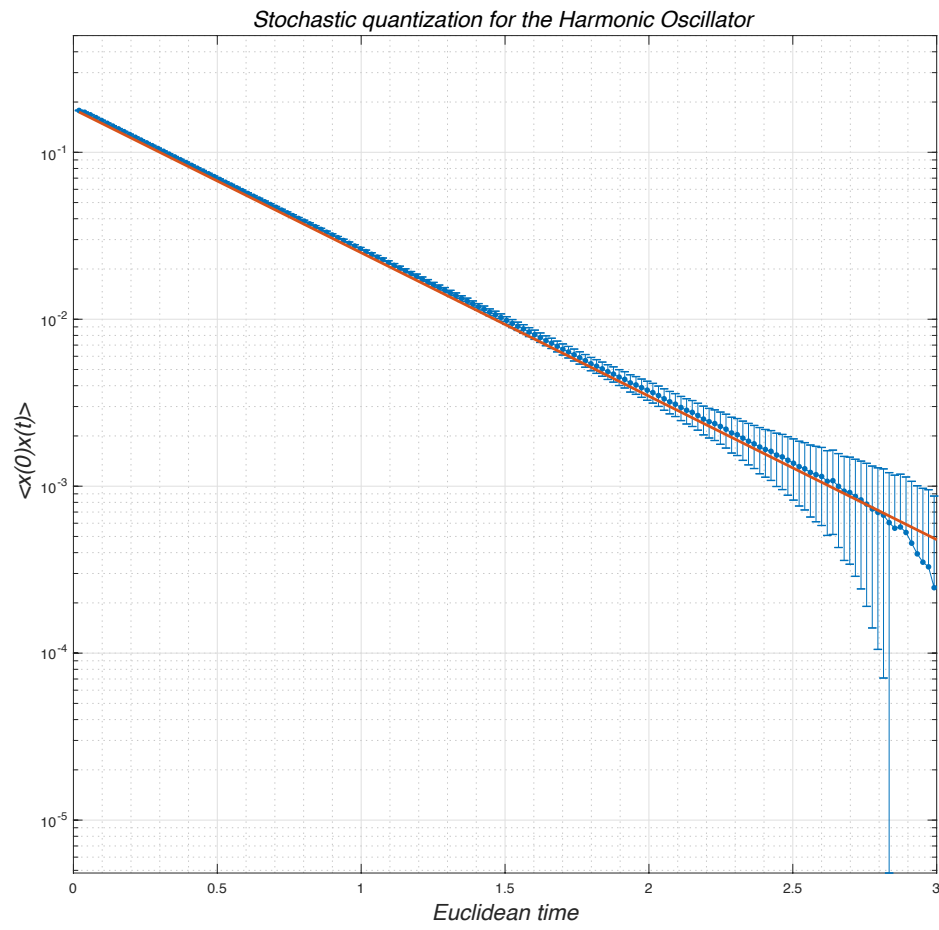


FIGURA 4.14. La correlazione $\langle x(0)x(t) \rangle$ permette di stimare il gap di energia.

APPENDICE A

Complementi

A.1. Generazione di successioni di numeri pseudo-casuali

Il problema di generare una successione di numeri casuali, ingrediente essenziale per ogni tecnica Monte Carlo, è stato affrontato fin dagli anni '40 e si sono trovate varie soluzioni via via più efficaci. Inizialmente si è pensato a sfruttare fenomeni fisici intrinsecamente casuali quali il decadimento radioattivo¹. Risulta evidente che non sarebbe molto pratico dover disporre come accessorio del proprio computer di una sorgente radioattiva e di un contatore Geiger. Si è presto riconosciuto che era tuttavia possibile realizzare un surrogato in termini di processi aritmetici deterministici il cui output presenta tutte le caratteristiche che si richiedono a una successione random - di qui il termine “pseudo-random” che richiama al fatto che le successioni generate dai vari processi aritmetici non hanno nulla di intrinsecamente casuale, al contrario sono esattamente riproducibili, ma presentano caratteristiche statistiche del tutto accettabili. Si consiglia di consultare le lezioni di Pelissetto alla scuola di Parma [Pel94] per una introduzione al problema. Nel trattato di D. Knuth [Knu98] l'intero capitolo III è dedicato al problema di definire e generare successioni di numeri (pseudo-)random. In queste lezioni assumiamo un approccio molto pragmatico: utilizzeremo a scatola chiusa gli strumenti offerti da `matlab` (`rand`, `randi`, `randn`) confidenti nel presupposto che questi si basino sui migliori algoritmi disponibili². Al momento attuale l'algoritmo utilizzato da `Matlab` è il “Mersenne twister”³ che permette di generare successioni di numeri reali nell'intervallo $[0, 1]$ per le quali si può contare su 2^{19937} numeri indipendenti, cioè con autocorrelazione trascurabile. Il principale vantaggio di un algoritmo come il Mersenne twister è costituito dal periodo estremamente lungo (una sottosuccessione di numeri non si ripete prima di circa 10^{6000} estrazioni), la sua portabilità da un linguaggio all'altro, la velocità di esecuzione, quest'ultima qualità è particolarmente apprezzata quando si eseguono processi Monte Carlo su sistemi con un numero elevato di gradi di libertà, ad es. nel caso della simulazione di una porzione di magnete tridimensionale con anche solo 128 spin in ogni direzione ogni “aggiornamento” della configurazione di spin richiede l'estrazione di $\mathcal{O}(2^{21})$ numeri random. L'algoritmo denominato Mersenne Twister ha soppiantato i tradizionali “linear congruential methods” che non danno garanzie per simulazioni a larga scala. Questi sono del tipo

$$X_n = aX_{n-1} + b \pmod{m}$$

dove tutti i numeri in gioco sono interi; m deve essere molto grande perchè determina il periodo di ricorrenza. Del Mersenne twister non riproduciamo lo schema in quanto si tratta di un algoritmo piuttosto complesso. Si possono trovare nei `/MPF2023/Documents` copia degli articoli rilevanti su Wikipedia.

A.2. Distribuzioni di probabilità elementari

Indicheremo con Ω lo spazio degli eventi elementari costituito da un insieme numerico e con X la variabile aleatoria che assume valori in Ω . Ricordiamo che per le distribuzioni discrete valgono le formule $F(z) \equiv \langle z^X \rangle$,

¹Un generatore perfetto di numeri interi casuali è realizzabile utilizzando una moneta perfettamente bilanciata: si lancia la moneta N volte e si registra l'uscita, ad es. $T, T, C, T, C, \text{etc}, C$, associando 1 alla testa e 0 alla croce, in questo modo si ottiene un numero in base 2 compreso tra 0 e $2^N - 1$ e ripetendo il processo si ha una successione $X_1, X_2, \dots, X_n \dots$ del tutto priva di autocorrelazione. Si definirà poi $x_k = X_k/2^N$ per ottenere numeri nell'intervallo $[0, 1)$. Purtroppo dal punto di vista pratico il processo è drammaticamente lento e del tutto inutilizzabile.

²Un esperimento quale la simulazione del modello di Ising in due dimensioni, di cui si conoscono tutte le caratteristiche statistiche in modo esatto, costituisce un convincente test di efficienza per la routine `rand()` utilizzata da `Matlab` per generare numeri casuali

³il termine inglese “twister” significa “tornado”, e vuole forse convogliare l'idea di grande velocità di esecuzione.

$\langle X \rangle = F'(0)$, $\langle X^2 \rangle = F''(1) + F'(1)$, mentre per le distribuzioni continue si preferisce definire $F(t) \equiv \langle e^{tx} \rangle$, e valgono le formule $\langle X \rangle = F'(0)$, $\langle X^2 \rangle = F''(0)$.

A.2.1. Distribuzione di Bernoulli.

Bernoulli

$$\Omega = \{0, 1\}, \quad P(0) = q, \quad P(1) = p, \quad (p + q = 1)$$

$$\text{Funzione generatrice: } F(z) = q + pz$$

$$\text{Media: } \mathbb{E}[X] = F'(1) = p$$

$$\text{Dev.St.: } \mathbb{D}[X] = F''(1) + F'(1) - F'(1)^2 = pq$$

A.2.2. Distribuzione binomiale.

Binomiale

$$\Omega = \{0, 1, 2, \dots, n\}, \quad P(k) = \binom{n}{k} q^{n-k} p^k, \quad (p + q = 1)$$

$$\text{Funzione generatrice: } F(z) = (q + pz)^n$$

$$\text{Media: } \mathbb{E}[X] = np$$

$$\text{Dev.St.: } \mathbb{D}[X] = npq$$

A.2.3. Distribuzione di Poisson. (Si ottiene dalla binomiale nel limite $np \rightarrow \lambda$, $n \rightarrow \infty$)

Poisson

$$\Omega = \mathbb{Z}, \quad P(k) = e^{-\lambda} \lambda^k / k!, \quad \lambda > 0$$

$$\text{Funzione generatrice: } F(z) = e^{\lambda(z-1)}$$

$$\text{Media: } \mathbb{E}[X] = \lambda$$

$$\text{Dev.St.: } \mathbb{D}[X] = \lambda$$

Gaussiana

A.2.4. Distribuzione di Gauss.

$$\Omega = \mathbb{R}, \quad \rho(x) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left\{-\frac{x^2}{2\sigma}\right\}, \quad \sigma > 0.$$

$$\text{Funzione generatrice: } F(t) = \int_{-\infty}^{\infty} dx e^{tx} \rho(x) = e^{\frac{1}{2}\sigma t^2}$$

$$\text{Media: } \mathbb{E}[X] = 0$$

$$\text{Dev.St.: } \mathbb{D}[X] = \sigma$$

Gaussiana n-dim

A.2.5. Distribuzione di Gauss multidimensionale.

$$\Omega = \mathbb{R}^d, \quad \rho(\mathbf{x}) = \frac{e^{-\frac{1}{2}\langle \mathbf{x} | \mathbf{A}^{-1} | \mathbf{x} \rangle}}{(2\pi)^{d/2} \sqrt{\det \mathbf{A}}}, \quad (\mathbf{A} \gg 0).$$

$$\text{Funzione generatrice: } F(\mathbf{t}) = \int_{-\infty}^{\infty} d\mathbf{x} e^{\langle \mathbf{y} | \mathbf{x} \rangle} \rho(\mathbf{x}) = e^{\frac{1}{2}\langle \mathbf{y} | \mathbf{A} | \mathbf{y} \rangle}$$

$$\text{Media: } \mathbb{E}[X] = 0$$

$$\text{Correlazione: } \langle x_i x_j \rangle = A_{ij}$$

La matrice A di dimensione $d \times d$ è *positiva definita*, ossia $\langle \mathbf{y} | A | \mathbf{y} \rangle > 0$ per ogni vettore \mathbf{t} non nullo. Il calcolo della funzione di distribuzione per una densità di probabilità gaussiana in \mathbb{R}^d si effettua come segue, avendo indicato con R la matrice ortogonale che diagonalizza A e con $\Lambda = \text{diag}\{\lambda_j\}$ la matrice diagonale $RA R^{-1}$:

$$\begin{aligned}
 F(\xi) &= \mathcal{N} \int e^{\langle \xi | \mathbf{x} \rangle - \frac{1}{2} \langle \mathbf{x} | A^{-1} | \mathbf{x} \rangle} d^d \mathbf{x} = \\
 &= \mathcal{N} \int e^{\langle \xi | R^{-1} \mathbf{x}' \rangle - \frac{1}{2} \langle R^{-1} \mathbf{x}' | A^{-1} R^{-1} \mathbf{x}' \rangle} d^d \mathbf{x}' = \\
 &= \mathcal{N} \int e^{\langle R \xi | \mathbf{x}' \rangle - \frac{1}{2} \langle \mathbf{x}' | \Lambda^{-1} \mathbf{x}' \rangle} d^d \mathbf{x}' = \\
 &= \mathcal{N} \prod_{j=1}^d \int e^{(R \xi)_j x_j - \frac{1}{2} \lambda_j^{-1} x_j^2} d x_j = \\
 &= \mathcal{N} \prod_{j=1}^d e^{\frac{1}{2} \lambda_j (R \xi)_j^2} \sqrt{2\pi \lambda_j} = \\
 &= \mathcal{N} (2\pi)^{\frac{d}{2}} \sqrt{\det A} e^{\frac{1}{2} \langle (R \xi) | \Lambda (R \xi) \rangle} \\
 &= e^{\frac{1}{2} \langle \xi | A \xi \rangle}
 \end{aligned}$$

A.3. Il metodo del bootstrap

Immaginiamo di avere una singola serie di dati come quella utilizzata per illustrare l'esperimento di Rutherford e Geiger. L'idea del bootstrap [Efr79] consiste nel generare un numero arbitrariamente grande di altre serie di dati estratte da una probabilità discreta costruita direttamente sull'istogramma della prima serie: detti $(0, 1, 2, \dots, N)$ i valori della variabile aleatoria X in questione e $P = (p_0, p_1, p_2, \dots, p_N)$ la frequenza di questi valori nella prima serie di misure, si generano altre serie di dati considerando P come la distribuzione di probabilità di X . A questo scopo si può utilizzare ad esempio il metodo degli alias per generare una serie di 10000 samples distribuiti secondo $P/\text{sum}(P)$. Si provi a immettere le istruzioni che seguono al prompt di matlab, avendo precedentemente definito P , ad es. nel caso di Rutherford-Geiger

```

P = [57, 203, 383, 525, 532, 408, 273, 139, 45, 27, 10, 4, 0, 1, 1];
P = P/sum(P);

```

e si prosegue utilizzando il metodo degli alias:

```

K = Kalias(P);
X = Krand(10000, K)
histogram(X, 'normalization', 'pdf')

```

Attraverso la generazione di un numero qualunque di serie di dati come X tra loro indipendenti si può arrivare a stimare le varie proprietà statistiche del fenomeno aleatorio anche senza avere a priori una definita distribuzione di probabilità dipendente da parametri da adattare l'esperimento e si parla allora di *analisi non-parametrica*. L'approccio appare a prima vista troppo semplice per essere di qualche utilità ma sulla base dell'esperienza e dell'analisi statistica del metodo (si vedano [Efr82, DH97]) siamo in presenza di uno strumento molto utile di indagine statistica.

A.4. La distribuzione di Poisson nel decadimento α

Bateman argomenta come segue: sia λdt la probabilità che avvenga un decadimento α nell'intervallo di tempo dt . Per tempi brevi rispetto al tempo di dimezzamento possiamo considerare λ costante. Sia $W_n(t)$ la probabilità che n particelle α colpiscano lo schermo sensibile nell'intervallo di tempo $[0, t)$. Allora la probabilità che ne arrivino $n+1$ entro $t+dt$ sarà calcolabile dal contributo di due possibilità: a) ne arrivi una nell'intervallo dt e ne arrivino n prima di t oppure b) non ne arrivi nessuna nell'intervallo dt e ne siano arrivate $n+1$ in precedenza. Si trova allora, trattandosi di eventi indipendenti,

$$W_{n+1}(t+dt) = P(a) + P(b) = \lambda dt W_n(t) + (1 - \lambda dt) W_{n+1}(t).$$

Nel limite $dt \rightarrow 0$ si trova allora

$$\frac{dW_n}{dt} = \lambda (W_n - W_{n+1}).$$

Si tratta di un sistema di equazioni accoppiate che si risolvono moltiplicando tutte le incognite W_n per il fattore $e^{\lambda t}$: $U_n = W_n e^{\lambda t}$. Si trova

$$\frac{dU_{n+1}}{dt} = \lambda U_n$$

Dato che $W_0(0) = 1$ e $W_n(0) = 0$ per $n > 1$ si trova $W_0(t) = e^{-\lambda t}$ e per integrazioni successive $W_n(t) = \frac{(\lambda t)^n}{n!} e^{-\lambda t}$ che è proprio la distribuzione di Poisson con valor medio λt . In realtà il calcolo può essere semplificato introducendo la funzione generatrice $F(z, t) = \sum_{n \geq 0} W_n(t) z^n$. Derivando rispetto a t troviamo

$$\begin{aligned} \frac{\partial F}{\partial t} &= \sum_{n \geq 0} \frac{dU_n}{dt} z^n = \lambda \sum_n (W_{n-1} - W_n) z^n \\ &= \lambda (z - 1) F \end{aligned}$$

da cui $F(z, t) = \exp\{\lambda t(z - 1)\}$ che è giusto la funzione generatrice della distribuzione di Poisson. ◀

A.5. I teoremi di Gershgorin e Perron-Frobenius

Per i due teoremi che seguono si veda [Var62].

TEOREMA A.1 (Gershgorin). Sia M una matrice qualunque $N \times N$. Lo spettro della matrice è contenuto nell'unione dei cerchi

$$\{z \in \mathbb{C} \mid |z - M_{ii}| \leq \sum_{j \neq i} |M_{ji}|\}.$$

Corollario: se la matrice M è una matrice stocastica, si trova $|z - M_{ii}| \leq 1 - M_{ii}$. Tutti i “cerchi di Gershgorin” sono allora contenuti nel cerchio unitario e passano tutti per il punto $z = 1$. Il teorema seguente, fondamentale nell'analisi delle catene di Markov finite, si trova sul testo di Varga oppure sul trattato di Gantmacher ([Gan59], Vol.II, Ch. XIII).

TEOREMA A.2 (Perron-Frobenius). Si assuma che la matrice A di dimensione N a elementi non-negativi sia irriducibile, cioè non esiste alcuna permutazione degli indici che la trasformi nella forma ridotta

$$\begin{pmatrix} A_{r \times r} & A_{r \times s} \\ 0 & A_{s \times s} \end{pmatrix}$$

con $r + s = N$. Allora valgono le seguenti proprietà:

- lo spettro di A contiene un autovalore reale positivo uguale al raggio spettrale $\rho(A)$;
- in corrispondenza di questo autovalore positivo esiste un unico autovettore avente tutte componenti positive;
- il raggio spettrale $\rho(A)$ è una funzione strettamente crescente degli elementi di matrice;
- se k è il numero di autovalori con modulo pari al raggio spettrale la matrice è detta *ciclica di indice* k e può essere ridotta attraverso una permutazione degli indici alla forma

$$\begin{pmatrix} 0 & B_{12} & 0 & \dots & 0 \\ 0 & 0 & B_{23} & \dots & 0 \\ \vdots & \vdots & \ddots & \dots & \vdots \\ 0 & 0 & 0 & \dots & B_{k-1,k} \\ B_{k,1} & 0 & 0 & \dots & 0 \end{pmatrix}.$$

- Gli autovalori di modulo $\rho(A)$, per una matrice ciclica di indice k , sono *semplici* e coincidono con le radici k -esime dell'unità $\exp\{2\pi i j/k\}$, e in generale per ogni autovalore λ anche $\lambda \exp\{2\pi i j/k\}$ appartiene allo spettro.

Il teorema determina le proprietà asintotiche del processo di Markov. Infatti per tempi lunghi la matrice di transizione sarà data da

$$P^t \underset{t \rightarrow \infty}{\sim} \sum_{j=0}^{k-1} e^{2\pi i j t/k} \Pi_j + \mathcal{O}(\lambda_{\text{sub}}^t),$$

dove i termini trascurabili sono associati a tutti gli autovalori in modulo inferiori a uno. Si afferma di solito che il limite della distribuzione di probabilità esiste solo per il caso *primitivo*, cioè per $k = 1$. Questo è a rigore

vero, ma irrilevante per le applicazioni dove quello che conta è il comportamento del processo *mediato su un tempo finito*. Infatti è chiaro che il contributo alla probabilità a tempi lunghi degli autovettori con autovalore $\exp\{2\pi i j/k\}$, $j > 0$ risulta fortemente attenuato se procediamo ad una media su un tempo finito (ricordiamo che per una qualunque radice k -esima dell'unità $\omega \neq 1$ si ha $\sum_{r=1}^k \omega^r = 0$). Si veda anche il seguente § A.7.

A.6. Soluzione esatta del modello di Ehrenfest

Indichiamo con $p(n, t)$ la probabilità che la prima urna contenga n particelle al tempo t ; conviene introdurre la funzione generatrice definita come

$$G(x, t) = \sum_{n=0}^N p(n, t) x^n.$$

G soddisfa ad una relazione di ricorrenza definita dalla probabilità di transizione (Eq.(34)) ⁴:

$$p(n, t+1) = \frac{n+1}{N} p(n+1, t) + \frac{N-n+1}{N} p(n-1, t).$$

Sostituiamo questa espressione per p nella definizione di $G(x, t+1)$ ed eliminiamo sistematicamente ogni occorrenza dell'indice n in favore di $x \frac{d}{dx}$, secondo l'ovvia identità $n x^n \equiv x \frac{d}{dx} x^n$:

$$\begin{aligned} G(x, t+1) &= \\ &= \sum_{n=0}^{N-1} \left(\frac{n+1}{N} p(n+1, t) x^n + \frac{N-n+1}{N} p(n-1, t) x^n \right) \\ &= \sum_{n=0}^{N-1} \left(\frac{1}{N} p(n+1, t) \frac{d}{dx} x^{n+1} + p(n-1, t) x^n - \frac{1}{N} x^2 \frac{d}{dx} x^{n-1} p(n-1, t) \right) \\ &= \frac{1-x^2}{N} \frac{\partial}{\partial x} G(x, t) + x G(x, t) \end{aligned}$$

La distribuzione di probabilità stazionaria deve soddisfare l'equazione $G(x, t+1) = G(x, t)$ ossia $(1+x) \frac{\partial}{\partial x} G(x, t) = N G(x, t)$ che ha come unica soluzione $G(x, t) = \mathcal{N} (1+x)^N$. La normalizzazione ($\sum p(n) = G(1, t) = 1$), impone $\mathcal{N} = 2^{-N}$. Ne segue che la distribuzione di equilibrio è data da $p(n) = 2^{-N} \binom{N}{n}$, cioè dalla *distribuzione binomiale simmetrica* ($p = q$) e si trova immediatamente $\mathbb{E}[n] = N/2$ e $\mathbb{D}[n] = N/4$. Notare che se pensiamo al modello fisico di un gas in equilibrio tra due recipienti uguali collegati tra loro, il modello suggerisce che il gas si distribuisce in modo uguale tra i due recipienti e le fluttuazioni di pressione, cioè le fluttuazioni nel numero di molecole, sono dell'ordine di \sqrt{N} . Per $N = 10^{23}$ è chiaro che le fluttuazioni sono percentualmente molto piccole ($O(10^{-12})$). Con un calcolo analogo si trovano tutti gli altri autovalori e autovettori della matrice di transizione (precisamente gli autovalori sono dati da $\{j/N \mid j = -N, -N+2, \dots, N-2, N\}$). La presenza dell'autovalore -1 implica che la matrice di transizione M è *ciclica di indice 2*. Si veda il dettaglio del calcolo nell'esercizio seguente.



ESERCIZIO 81. Considerare il modello di Ehrenfest modificato definito a pag.83. La catena di Markov è definita in questo caso da

$$(70) \quad p_j(t+1) = \left(1 - \frac{j-1}{N}\right) p_{j-1}(t) + 2 \frac{j}{N} \left(1 - \frac{j}{N}\right) p_j(t) + \left(\frac{j+1}{N}\right) p_{j+1}(t)$$

Conviene introdurre la funzione generatrice $F(z, t) \equiv \sum_{j=0}^N p_j(t) z^j$. La funzione $F(z, t)$ evolve nel tempo in accordo con l'(70) e cioè

$$\begin{aligned} (71) \quad F(z, t+1) &= \sum_{j=0}^N p_j(t+1) z^j \\ &= \sum_{j=0}^N \left(\left(1 - \frac{j-1}{N}\right) p_{j-1}(t) + 2 \frac{j}{N} \left(1 - \frac{j}{N}\right) p_j(t) + \left(\frac{j+1}{N}\right) p_{j+1}(t) \right) z^j \end{aligned}$$

⁴il primo coefficiente è da considerare uguale a zero per $n = N$ e il secondo per $n = 0$

Adesso possiamo sistematicamente sostituire ad ogni termine contenente jz^j l'espressione equivalente $z \frac{dz}{dz} z^j$; in questo modo si elimina dappertutto j dalla formula per $F(z, t+1)$:

$$(72) \quad F(z, t+1) = \frac{z(z-1)^2}{N^2} \frac{\partial^2}{\partial z^2} F(z, t) + \frac{1-z}{N^2} (1-z_2 N z) \frac{\partial}{\partial z} F(z, t) + z F$$

Ne possiamo concludere che la distribuzione stazionaria della catena di Markov soddisfa all'equazione

$$(73) \quad F''(z) + \left(\frac{1}{z} + \frac{2N}{1-z} \right) F'(z) + \frac{N^2}{z(z-1)} F(z) = 0$$

L'equazione è del tipo ipergeometrico e si può determinare la soluzione in

$$(74) \quad F(z) = {}_2F_1(-N, -N, 1, z) = C \sum_{j=0}^N (-N)_j^2 z^j / (j!)^2 = C \sum_{j=0}^N \binom{N}{j}^2 z^j$$

La costante di normalizzazione è fissata dalla condizione $F(1) = 1$. Secondo una formula che si trova sul testo di Rozanov si ha infine

$$(75) \quad F(z) = \binom{2N}{N}^{-1} \sum_{j=0}^N \binom{N}{j}^2 z^j.$$

Si tratta di una distribuzione che generalizza quella di Bernoulli ed è un caso particolare della distribuzione di Pólya. Per saperne di più si consulti [Fis63] oppure [Fel68] (Cap.II.6).

ESERCIZIO 82. Calcolare lo spettro della matrice stocastica che caratterizza il modello di Ehrenfest.

SOLUZIONE. *La matrice*

$$M p_n = \frac{n+1}{N} p_{n+1} + \frac{N-n+1}{N} p_{n-1}$$

ha uno spettro che a priori deve essere contenuto nel disco $|\lambda| \leq 1$. Il calcolo numerico dello spettro si fa agevolmente utilizzando **matlab**, con l'avvertenza che per dimensioni grandi della matrice (oltre 200×200) possono presentarsi anomalie causate dalla precisione finita. Si può allora ricorrere a programmi che consentono di utilizzare precisione multipla (**advanpix**) o addirittura infinita (**Mathematica**). Qui troviamo lo spettro per via analitica. Introduciamo il polinomio $p(x) = \sum_{n=0}^N p_n x^n$. L'azione di M si traduce nell'operatore

$$\mathcal{M} p(x) = \frac{1-x^2}{N} \frac{dp(x)}{dx} + x p(x).$$

L'equazione agli autovalori

$$\mathcal{M} p(x) = \lambda p(x)$$

costituisce una semplice equazione differenziale lineare del primo ordine, già considerata nell'esercizio precedente nel caso particolare $\lambda = 1$:

$$\frac{1-x^2}{N} \frac{dp(x)}{dx} + x p(x) = \lambda p(x)$$

che si risolve per separazione delle variabili

$$\begin{aligned} \frac{1}{N} p'(x) &= \frac{\lambda - x}{1-x^2} p(x) \\ \frac{d \ln(p(x))}{dx} &= \frac{1}{2} N \left(\frac{\lambda-1}{1-x} + \frac{\lambda+1}{1+x} \right) \\ p(x) &= (1+x)^{\frac{1}{2} N(\lambda+1)} (1-x)^{\frac{1}{2} N(1-\lambda)} \end{aligned}$$

Dato che $p(x)$ deve essere un polinomio, gli esponenti nella formula precedente saranno due interi non negativi, diciamo k_1 e k_2 :

$$k_1 = \frac{1}{2} N(\lambda+1), \quad k_2 = \frac{1}{2} N(1-\lambda)$$

da cui otteniamo $k_1 + k_2 = N$ e $\lambda = (k_1 - k_2)/N$. Concludiamo che lo spettro consiste nell'insieme di numeri reali $\lambda_n = \{(2k/N - 1) \mid 0 \leq k \leq N\}$. ◀

ESERCIZIO 83. Dimostrare che la matrice di transizione del modello di Ehrenfest soddisfa l'equazione $MT + TM = 0$ essendo T la matrice diagonale con elementi di matrice $T_{n,n} = (-1)^n$. (ciò implica anche che lo spettro di autovalori di M deve essere simmetrico intorno all'origine delle coordinate).

A.7. Calcolo del tempo medio di ritorno

Calcoliamo le funzioni generatrici $\mathcal{F}_j(s)$ e $\mathcal{G}_j(s)$ introdotte nel Probl.(3.2.2) a Pag.85 utilizzando la decomposizione spettrale della matrice P . Per semplicità ammettiamo che P sia diagonalizzabile, anche se questo non sarebbe strettamente necessario. Sia perciò

$$P = \sum_{\lambda \in \sigma(P)} \lambda \Pi_\lambda$$

dove la somma è estesa a tutto lo spettro di P e le matrici Π_λ costituiscono i *proiettori spettrali*. Si ha allora

$$\mathcal{G}_j(s) = \sum_{\lambda} \frac{s\lambda}{1-s\lambda} (\Pi_\lambda)_{jj}$$

Dalla relazione contenuta nell'Eq.(42) si ricava

$$\frac{d\mathcal{F}_j}{ds} = \frac{\mathcal{G}'_j(s)}{(1 + \mathcal{G}_j(s))^2}$$

e inserendo la rappresentazione spettrale si ottiene così il tempo medio di ritorno

$$(76) \quad \tau_j = \frac{d\mathcal{F}_j}{ds} = \frac{\sum_{\lambda} \frac{\lambda}{(1-s\lambda)^2} (\Pi_\lambda)_{jj}}{(\sum_{\lambda} \frac{1}{1-s\lambda} (\Pi_\lambda)_{jj})^2}$$

$$(77) \quad = \frac{\frac{1}{(1-s)^2} (\Pi_1)_{jj} + \sum'_{\lambda} \frac{\lambda}{(1-s\lambda)^2} (\Pi_\lambda)_{jj}}{(\frac{1}{1-s} (\Pi_1)_{jj} + \sum'_{\lambda} \frac{1}{1-s\lambda} (\Pi_\lambda)_{jj})^2}$$

$$(78) \quad = (\Pi_1)_{jj}^{-1} + \mathcal{O}(1-s).$$

(\sum' indica la somma estesa agli autovalori diversi da uno). Ora il proiettore Π_1 è dato da $(\Pi_1)_{nm} = u_n^R u_m^L$ essendo u^R e u^L gli autovettori *destri* e *sinistri* di P con autovalore uno.

$$\Pi_1 u^R = u^R, u^L \Pi_1 = u^L.$$

Una considerazione necessaria riguarda la normalizzazione degli autovettori: u^R rappresenta la *distribuzione di probabilità stazionaria*, ed è quindi da normalizzare secondo $\sum_j u_j^R = 1$. L'autovettore sinistro u^L invece ha tutte componenti uguali $u_j^L = u = \text{costante}$, per via della proprietà di P che ne fa una matrice stocastica (Eq.36). La condizione che Π_1 sia un proiettore impone $u = 1$ (si lascia questo da verificare). Dato che l'autovettore sinistro ha tutte componenti uguali a uno segue che nel limite $s \rightarrow 1$ vale la relazione

$$(79) \quad \tau_j = 1/u_j^R.$$

Si noti che nel caso di matrice di transizione *irriducibile* tutte le componenti della distribuzione stazionaria u_R sono strettamente positive, in altri termini tutti gli stati sono raggiungibili in tempo medio finito.

ESERCIZIO 84. Dalla relazione Eq.76 ricavare i primi due termini dello sviluppo di Taylor per $\mathcal{F}_j(s)$ e dedurre la varianza del tempo medio di ritorno $\mathbb{D}[\tau_j]$.

SOLUZIONE. Il codice Mathematica ⁵ che segue calcola le funzioni generatrici \mathcal{F} e \mathcal{M} per il modello di Ehrenfest.

```

EhrenfestRetTimes.nb
EhrenfestM[n_] :=
  SparseArray[{{i_, j_}/; i-j==1->1-(j-1)/n, {i_, j_}/; i-j== -1-> i/n}, {n+1, n+1}]
M[n_] := M[n] = Normal[EhrenfestM[n]];
R[n_, s_] := R[n, s] = Simplify[s M[n] . Inverse[IdentityMatrix[n+1] - s M[n],
  Method->CofactorExpansion]];
M[n_, s_] := M[n, s] = Simplify[Tr[R[n, s], List]];

```

⁵EhrenfestRetTimes.nb si trova su Dropbox.

```

F[n_, s_] := F[n, s] = Simplify[M[n, s] / (1 + M[n, s])];

K=15;
Simplify[F[K, s]] /. s -> 1
{1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1} (* ritorno c.p.1 per tutti gli stati iniz. *)
Series[D[F[K, s], s], {s, 1, 1}]; (* Output omissso *)
tau = Limit[D[H[K, s], s], s -> 1]
{32768, 32768/15, 32768/105, 32768/455, 32768/1365, 32768/3003,
 32768/5005, 32768/6435, 32768/6435, 32768/5005, 32768/3003,
 32768/1365, 32768/455, 32768/105, 32768/15, 32768}
(* 2^15/Binomial(15, k) *)

```

I valori di τ_n coincidono con i reciproci delle componenti della distribuzione stazionaria (Sec. A.6). Si ha qui un esempio di processo che sebbene tenda all'equilibrio prevede che ogni stato si ripresenti con probabilità uno ma su tempi medi che possono diventare molto grandi (immaginare cosa sia un tempo $2^{10^{23}}$ anche ammettendo che il passo temporale sia 10^{-10} sec!), con il che, su scale temporali limitate, il processo appare irreversibile. Il testo già citato di Mark Kac ([Kac59]) contiene una discussione approfondita del modello di Ehrenfest e della sua interpretazione in termini di meccanica statistica classica. ◀

PROBLEMA A.7.1. Per il modello di Ehrenfest dimostrare le relazioni

$$\begin{aligned} \mathbb{P}(n_A(t-1) = m-1, n_A(t+1) = m-1 | n_A(t) = m) &= (m/N)^2 \\ \mathbb{P}(n_A(t-1) = m+1, n_A(t+1) = m-1 | n_A(t) = m) &= m(N-m)/N^2 \\ \mathbb{P}(n_A(t-1) = m+1, n_A(t+1) = m+1 | n_A(t) = m) &= (1-m/N)^2 \end{aligned}$$

dove $n_A(t)$ è il numero di palle nella prima urna al tempo t e m è un valore intero arbitrario positivo. Se ne deduce che se il processo va fuori dall'equilibrio, ossia m assume valori vicini a N , la configurazione $m-1 \rightarrow m \rightarrow m-1$ è molto più probabile di $m+1 \rightarrow m \rightarrow m-1$ e di $m+1 \rightarrow m \rightarrow m+1$. In pratica la curva $n_A(t)$ lontano dal valore medio $N/2$ tende ad essere convessa (si veda [Kac59], III.7).

PROBLEMA A.7.2. Modificare il modello di Ehrenfest introducendo un elemento che lo avvicina alla fisica dei gas: trovare il modo di assegnare una temperatura al gas in entrambe le urne e determinare la probabilità di transizione in modo che sia il numero di molecole sia la loro temperatura influiscano sulla probabilità che una molecola passi da un'urna all'altra.

SOLUZIONE. Si consideri il programma `matlab` EhrenfestT.m, presente su `dropbox`. In figura un'esperimento da cui si vede come la temperatura va all'equilibrio e nel contempo la differenza nel numero di molecole inizialmente cresce per via della differenza di pressione e poi si stabilizza sulla equidistribuzione. Il principio si basa sull'idea che ogni molecola ha una energia cinetica kT uniforme in ciascuna urna. La temperatura determina anche la pressione quindi bisogna aumentare la probabilità che si abbia una transizione $A \rightarrow B$ se $T_A > T_B$. L'istruzione `pA→B` diventa allora

```

if rand < nA*TA/(nA*TA+nB*TB),
    nA = nA-1; TB=(nB*TB+TA)/(nB+1); nB = nB+1;
else
    nB = nB-1; TA=(nA*TA+TB)/(nA+1); nA = nA+1;
end

```

Nel caso che una molecola passi da A a B essa distribuisce la sua energia nell'urna B

```

TB=(nB*TB+TA)/(nB+1);

```

altrimenti

```

TA=(nA*TA+TB)/(nA+1);

```

All'equilibrio $T_A = T_B$ e il processo prosegue come nel modello originale. ◀

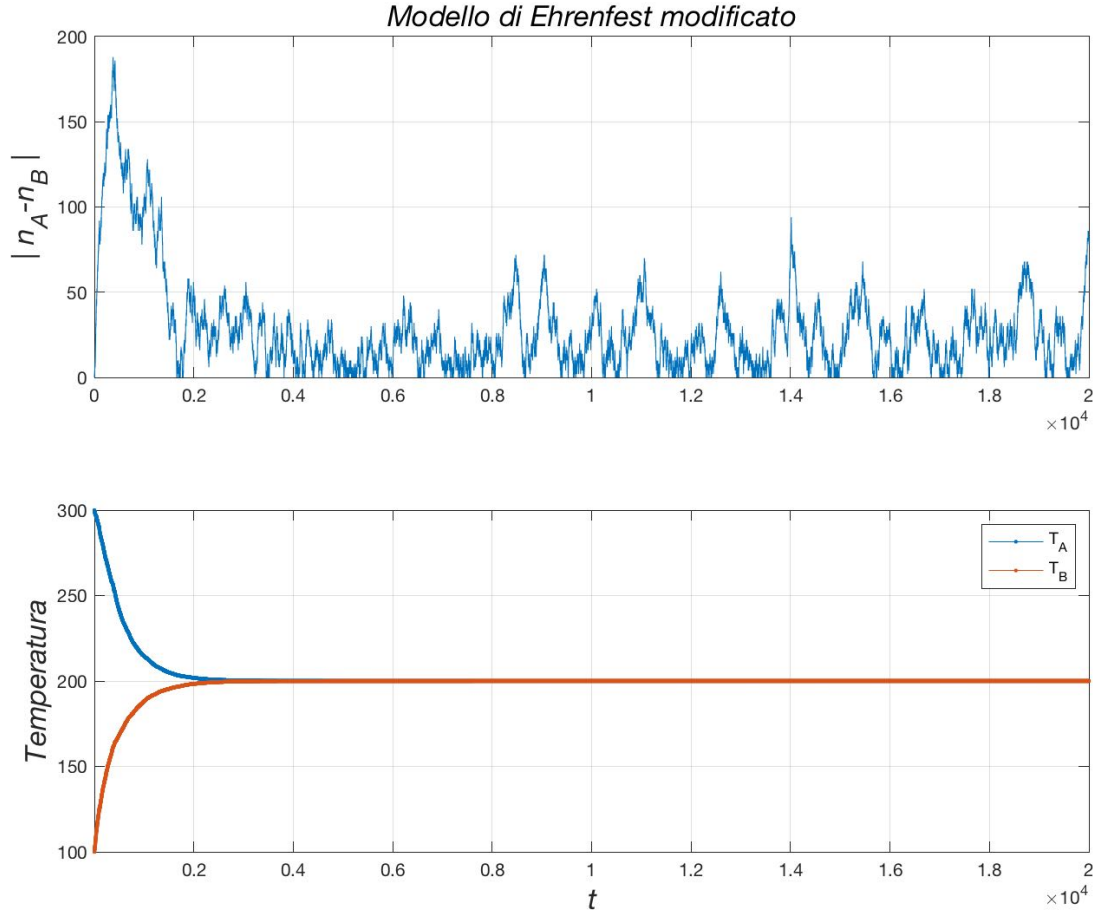


FIGURA A.1. Un esperimento per il modello di Ehrenfest modificato con l'introduzione della temperatura.

A.8. Processi diffusivi

Sia $\mathbf{x}(t) : [0, +\infty) \rightarrow \mathbb{R}^n$ un processo stocastico che evolve nel tempo secondo la seguente equazione:

$$(80) \quad d\mathbf{x}(t) = \mathbf{b}(\mathbf{x}(t), t) dt + \sqrt{\sigma(\mathbf{x}(t), t)} d\mathbf{w}(t)$$

dove $\mathbf{b}(\mathbf{x}(t), t)$ è un campo vettoriale, detto campo di *drift*, che agisce in modo deterministico sul sistema; $d\mathbf{w}(t)$ è una variabile casuale distribuita in modo gaussiano con $\mathbb{E}(d\mathbf{w}) = 0$ e

$$\mathbb{E}(dw_i(t) dw_j(s)) = dt \delta(t - s) \delta_{ij},$$

La funzione σ rappresenta il coefficiente di diffusione, cosicché $\sqrt{\sigma} d\mathbf{w}(t)$ fornisce il contributo aleatorio allo spostamento infinitesimo, come descritto in figura. Si può pensare alla dinamica rappresentata dall'Eq. (80) come dovuta al moto in presenza di un campo di forze $\mathbf{F}(\mathbf{x})$ e di una forza di attrito in regime critico; l'attrito è così forte che la particella raggiunge immediatamente la velocità limite \mathbf{F}/μ . In più la particella è soggetta a urti con le molecole del mezzo viscoso che impartiscono uno spostamento di rinculo casuale distribuito in modo Gaussiano. Dunque in totale $d\mathbf{x} = \mathbf{b}(\mathbf{x}, t) dt + \sqrt{\sigma} d\mathbf{w}(t)$ e le variabili gaussiane $w_i(t)$ soddisfano $\mathbb{E}(w_i(t) w_j(s)) = \delta_{ij} \inf(t, s)$ o semplicemente $\mathbb{E}(dw_i(t) dw_j(t)) = \sqrt{\sigma} \delta_{ij} dt$.

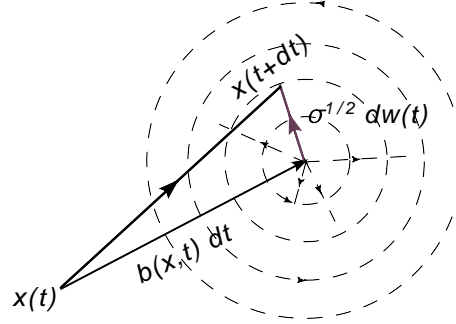


FIGURA A.2. Lo spostamento $\mathbf{x}(t) \rightarrow \mathbf{x}(t + dt)$ decomposto in parte casuale e in trascinamento (*drift*)

Ora, assumiamo che la probabilità che $\mathbf{x}(t)$ si trovi al tempo t in una certa porzione di spazio $\mathcal{I} \subset \mathbb{R}^n$ sia data da:

$$(81) \quad \mathbb{P}_{\mathcal{I}}(t) = \int_{\mathcal{I}} \rho(\mathbf{x}, t) d\mathbf{x}$$

dove $\rho(\mathbf{x}, t)$ è la densità di probabilità. Si trova, sotto opportune condizioni di continuità, che $\rho(\mathbf{x}, t)$ soddisfa l'equazione differenziale

$$(82) \quad \frac{\partial \rho}{\partial t} = \frac{\sigma}{2} \Delta \rho - \nabla \cdot (\rho \mathbf{b})$$

detta *equazione di Kolmogorov* o di *Fokker-Planck*. Il primo addendo rappresenta la parte detta *diffusiva* del processo, mentre il secondo è il classico termine dell'equazione di continuità per il campo di velocità \mathbf{b} di cui la (82) rappresenta una generalizzazione⁶. L'argomento è il seguente: il valor medio di una qualunque funzione delle coordinate è dato al tempo t dall'integrale $\int_{\mathbb{R}^n} \rho(\mathbf{x}, t) f(\mathbf{x}) d^n \mathbf{x}$. Al tempo successivo $t + dt$ la densità sarà $\rho(\mathbf{x}, t + dt)$ ma possiamo anche calcolare il valor medio di $f(\mathbf{x})$ tenendo conto che con probabilità $\rho(\mathbf{x}, t)$ la posizione al tempo $t + dt$ è data dall'Eq. (80) e quindi

$$\int \rho(\mathbf{x}, t + dt) f(\mathbf{x}) d^n \mathbf{x} = \mathbb{E}_w \left(\int \rho(\mathbf{x}, t) f(\mathbf{x} + \mathbf{b}(\mathbf{x}, t)dt + \sqrt{\sigma} d\mathbf{w}(t)) d^n \mathbf{x} \right)$$

Adesso possiamo sviluppare in serie la $f(\cdot)$ tenendo conto che $\mathbb{E}(d\mathbf{w}) = 0$ e $\mathbb{E}(d\mathbf{w}^2) = O(dt)$ il che comporta che dobbiamo tenere in considerazione lo sviluppo di Taylor di $f(\cdot)$ fino al secondo ordine; si trova:

$$\int \rho(\mathbf{x}, t + dt) f(\mathbf{x}) d^n \mathbf{x} = \mathbb{E}_w \left[\int \rho(\mathbf{x}, t) \left(f(\mathbf{x}) + dt \frac{\partial f(\mathbf{x})}{\partial x} \mathbf{b}(\mathbf{x}, t) + \frac{1}{2} dt \sigma \frac{\partial^2 f}{\partial x^2} \right) \right]$$

Ricostruendo le formule in n dimensioni e *integrando per parti* si trova allora l'Eq.(82).

Nel caso in cui il drift \mathbf{b} sia il gradiente di una funzione, allora è immediato determinare la distribuzione invariante, cioè la soluzione stazionaria dell'Eq.(82). Sia infatti

$$\mathbf{b}(\mathbf{x}) = -\frac{\sigma}{2} \nabla S(\mathbf{x}).$$

Segue che il membro di destra dell'equazione si riduce a

$$\frac{1}{2} \sigma \nabla \cdot (\nabla \rho + \rho \nabla S)$$

che quindi si annulla per $\rho = \mathcal{Z}^{-1} \exp\{-S\}$. Nell'applicazione alla meccanica quantistica, la funzione S rappresenta l'azione (euclidea, cioè a tempo immaginario) e per consistenza è presente un fattore $1/\hbar$

$$\rho(\mathbf{x}) = \exp \left\{ -\frac{1}{\hbar} S(\mathbf{x}) \right\}$$

⁶Si ricorda ad es. l'equazione di continuità per la densità di un fluido o per la densità di carica, identica all'Eq.(82) con $\sigma = 0$, essendo \mathbf{b} la *corrente*.

e l'equazione stocastica assume la forma

$$d\mathbf{x}(t) = -\frac{\sigma}{2\mathbf{h}} \nabla S(\mathbf{x}) dt + \sqrt{\sigma} d\mathbf{w}(t)$$

La costante di diffusione σ regola l'evoluzione temporale ma non influenza la soluzione stazionaria, quindi si può scegliere arbitrariamente, ad es. $\sigma = 2\mathbf{h}$, il che semplifica formalmente l'equazione; sono però a disposizione altre scelte che possono rendere più veloce la convergenza alla soluzione stazionaria. Ad esempio è anche possibile scegliere σ uguale ad una matrice (positiva definita); nel caso in cui S sia quadratica ($\frac{1}{2} < \mathbf{x}, A\mathbf{x} >$), è evidente il vantaggio, come ha suggerito Parisi, di porre $\sigma = A^{-1}$; infatti si può dare il caso in cui gli autovalori di A siano molto diversi come ordine di grandezza, cioè l'indice di condizionamento dato dal rapporto tra l'autovalore più grande e quello più piccolo sia molto grande. In questa situazione il calcolo numerico della soluzione dell'equazione stocastica risulta molto penalizzata (per raggiungere l'equilibrio è necessario attendere i modi più lenti). La scelta di Parisi mette tutti i modi al passo con un notevole vantaggio di efficienza dell'algoritmo⁷.

Come esempio affrontiamo la risoluzione di un problema particolare

$$d\mathbf{x}(t) = -A\mathbf{x}dt + \sqrt{2} d\mathbf{w}(t),$$

(dove abbiamo posto per semplicità $\mathbf{h} = 1$), la cui soluzione è

$$(83) \quad \mathbf{x}(t) = e^{-At}\mathbf{x}(0) + \int_0^t e^{A(t'-t)} d\mathbf{w}(t')$$

dove $\mathbf{x}(0)$ è la condizione iniziale. Dato che \mathbf{x} risulta sovrapposizione lineare di variabili gaussiane (sempre che lo sia $\mathbf{x}(0)$), la soluzione è data ad ogni istante da una variabile aleatoria gaussiana; per caratterizzarla è allora sufficiente conoscerne la media e la (co-)varianza. Dalla (83) si trova $\mathbb{E}(\mathbf{x}(t)) = e^{-At}\mathbb{E}(\mathbf{x}(0))$, essendo $\mathbb{E}(d\mathbf{w}(t)) = 0$. Nel caso in cui A sia definita positiva si ha $\lim_{t \rightarrow \infty} \mathbb{E}(\mathbf{x}(t)) = 0$. Per semplicità poniamo $\mathbf{x}(0) = 0$, il che non influenza il risultato a tempi grandi, e troviamo

$$\begin{aligned} \mathbb{E}(x_i(t) x_j(t)) &= \mathbb{E}\left(\int_0^t dt' (e^{A(t'-t)})_{ik} d\mathbf{w}_k(t') \int_0^t dt'' (e^{A(t''-t)})_{jh} d\mathbf{w}_h(t'')\right) \\ &= \int_0^t \int_0^t (e^{A(t'-t)})_{ik} (e^{A(t''-t)})_{jh} \mathbb{E}(d\mathbf{w}_k(t') d\mathbf{w}_h(t'')) \\ &= \int_0^t \int_0^t (e^{A(t'-t)})_{ik} (e^{A(t''-t)})_{jh} 2dt' \delta_{kh} \delta(t' - t'') \\ &= 2 \int_0^t (e^{2A(t'-t)})_{ij} dt' \\ &= \left(\frac{1 - e^{-2At}}{A}\right)_{ij} \end{aligned}$$

Si è sfruttato il fatto che la matrice A è simmetrica. Nel limite per $t \rightarrow \infty$ si trova

$$\mathbb{E}(x_i(t) x_j(t)) \rightarrow (A^{-1})_{ij}$$

Se adottiamo la scelta di Parisi, $\sigma = A^{-1}$, la soluzione diviene

$$\mathbf{x}(t) = \mathbf{x}(0)e^{-t} + \int_0^t e^{t'-t} A^{-1/2} d\mathbf{w}(t')$$

da cui segue

$$\mathbb{E}[x_i(t) x_j(t)] = (A^{-1})_{ij} (1 - e^{-2t})$$

Quindi la soluzione dipende dalla scelta di σ , ma non la distribuzione stazionaria, che è data da

$$\rho(\mathbf{x}, t) \rightarrow \mathcal{N} \exp\left\{-\frac{1}{2} \langle \mathbf{x} |, A | \mathbf{x} \rangle\right\}$$

Un problema interessante è quello di realizzare algoritmi efficienti per studiare numericamente la soluzione di equazioni differenziali stocastiche; teniamo conto infatti che i casi in cui si sa risolvere esattamente l'equazione

⁷Il calcolo della radice quadrata $A^{-1/2}$ non pone eccessivi problemi; inoltre è effettuata una volta sola nel corso del calcolo.

si riducono a poca cosa (caso lineare o riconducibile a). Un metodo molto semplice è basato sull'osservazione seguente: per integrazione si ha dall'Eq. (56) (avendo assunto $\sigma = 1$)

$$x(t + \tau) = x(t) + \int_t^{t+\tau} b(x(t')) dt' + w(t + \tau) - w(t),$$

e dunque la parte casuale è data *esattamente* da un contributo gaussiano con media 0 e varianza $\sqrt{\tau}$. L'integrale invece bisogna approssimarlo in modo accurato. La formula del trapezio è la prima possibilità⁸

$$\int_t^{t+\tau} b(x(t')) dt' \approx \frac{1}{2} \tau (b(x(t)) + b(x(t + \tau))) .$$

Questo tuttavia obbliga a risolvere un'equazione implicita per ottenere $x(t + \tau)$. In alternativa si può approssimare ulteriormente la formula inserendo $x(t + \tau) \approx x(t) + b(x(t))\tau + w(t + \tau) - w(t)$ nella formula del trapezio.

Per altri algoritmi, più accurati, si può consultare [KP99]. Per l'applicazione delle equazioni stocastiche al Montecarlo si veda invece [DH87].

A.9. Dimostrazione della Eq.(66), correlazione quantistica

Sia $q(t)$ l'operatore di posizione in descrizione di Heisenberg, cioè $q(t) = \exp(iHt) q \exp(-iHt)$. Passiamo a tempo immaginario e calcoliamo

$$\begin{aligned} \langle E_0 | q(it) q(is) | E_0 \rangle &= \langle E_0 | q e^{-H(s-t)} q | E_0 \rangle e^{E_0(s-t)} \\ &= \sum_n \langle E_0 | q | E_n \rangle \langle E_n | q | E_0 \rangle e^{-(E_n - E_0)(s-t)} \\ &= \sum_n |\langle E_0 | q | E_n \rangle|^2 e^{-(E_n - E_0)(s-t)} \end{aligned} \quad (84)$$

D'altra parte, scelto un vettore arbitrario $|\psi\rangle$, utilizzando la relazione

$$|E_0\rangle = \lim_{T \rightarrow \infty} e^{-(H - E_0)T} |\psi\rangle / \langle E_0 | \psi \rangle$$

si ha anche

$$\begin{aligned} \langle E_0 | q(it) q(is) | E_0 \rangle &= \lim_{T \rightarrow \infty} \frac{\langle \psi | e^{-HT} q e^{-H(s-t)} q e^{-HT} | \psi \rangle}{\langle \psi | e^{-HT} e^{-H(s-t)} e^{-HT} | \psi \rangle} \\ &\sim \frac{\int \mathcal{D}x(.) \exp \left\{ - \int_{-\infty}^{\infty} d\tau \left(\frac{1}{2} \dot{x}^2 + V(x(\tau)) \right) \right\} x(t) x(s)}{\int \mathcal{D}x(.) \exp \left\{ - \int_{-\infty}^{\infty} d\tau \left(\frac{1}{2} \dot{x}^2 + V(x(\tau)) \right) \right\}} \end{aligned} \quad (85)$$

L'ultimo passaggio presume familiarità con il formalismo dei *path integrals* di Feynman come sviluppati ad es. su [Fey72]. Si veda [Pol77] per l'applicazione al calcolo del gap di energia e per l'estensione a sistemi più interessanti.

⁸Risalente addirittura a Isaac Newton!

APPENDICE B

Codici matlab

Nota. Le versioni aggiornate dei codici si possono trovare nel folder `~dropbox/MPF2023/Programmi/Matlab`

B.1. Probabilità applicata al poker

Il programma `matlab` che segue calcola il numero di poker serviti in una serie di N smazzate con $4 \times nC$ carte attraverso un'estrazione a caso (con quattro giocatori $nC = 8$). L'ingrediente principale è la funzione `randperm` che genera una qualunque permutazione di n oggetti con uguale probabilità a priori; ciò equivale a mescolare le carte, meglio di un croupier. Si mettono in ordine crescente i valori delle carte della mano assegnata al primo giocatore e si cercano le combinazioni per cui la prima carta è uguale alla quarta oppure la seconda è uguale alla quinta. Il programma è stato utilizzato per produrre l'istogramma a pagina 16. La versione più aggiornata di `poker.m` si trova su `dropbox/MPF2023/Programmi/Matlab/Cap.2/Poker`. L'altro quesito proposto è: qual'è la probabilità che ci siano due o più giocatori con un poker servito? Questo è ovviamente di interesse per chi si trovi un poker e debba valutare quanto sia sicura la vincita! A questo si può provvedere con il programma `pokermult4G.m`. Da una simulazione con 50 esperimenti da 10^7 smazzate si può stimare questa probabilità in $1.138 \times 10^{-5} \pm 1.6 \times 10^{-7}$.

```
function [p, Prb, pexact] = poker(nC,N)
% Random draws from a deck of nC different cards in 4 suits
% p returns the number of pokers served to the first player
% in N shuffles
% Usage:
%   p=poker(nC,N)
%   default: nC=8, N=1000; nC should not exceed 13.

if nargin<2, N=1000; end
if nargin<1, nC=8;      % eight cards 7-10,J,Q,K,A
end

C = 1:nC;
M = [C,C,C,C];          % 4 suits

p=0;
for j=1:N
    G = M(randperm(4*nC));
    G = sort(G(1:5));
    if (G(1)==G(4) || G(2)==G(5))
        p=p+1;
    end
end

pexact = 30 / ((4*nC-1)*(4*nC-2)*(4*nC-3))
Prb = binomial(p,q,N,10);
end
%-----
function prob = binomial(p,q,N,bt) % DISTRIBUZIONE BINOMIALE
prob=[];
for k=0:bt
    prob=[prob,nchoosek(N,k)*q^(N-k)*p^k];
end
```

B.2. Il problema del botteghino

Per la soluzione si parte da un vettore con $2n$ componenti $(1, 1, 1, \dots, 1)$; si sceglie a caso un sottoinsieme di $n-m$ elementi e in corrispondenza a questo si pone il valore a -1 . (1 corrisponde a una persona in possesso di una banconota da 5 €, -1 altrimenti. Se la somma parziale delle componenti del vettore si mantiene non-negativa siamo nel caso in cui nessuno deve aspettare per avere il resto. Il codice si spiega da solo...

Il sottoinsieme si sceglie a caso con la routine “randsample”, che è realizzata secondo lo schema che si trova nel testo di Knuth, Vol.2, page 142.

```

                                Botteghino.m
function S = Botteghino(n, m, c, sample, nsw, exps)
% Il problema del botteghino: vedere al Cap. 2, esercizi.
% Usage:
% S = botteghino(n, m, c, sample, nsw, exps)
% numero di spettatori=2n, numero di spettatori con banconote da
% 5Eur=n+m; sample = numero di code simulate
% ©2005-2023 E. Onofri

N = 2*n;
q = n-m;                                % # of people without 5Eur bill
S = [];

j = 0:c; prob = 1-prod((n-m-j)./(n+m+j+1)); % conjecture...

for ex = 1:exps,                        % start experiments
    Scurr=zeros(nsw,1);

    for k = 1:nsw
        hit = 0;
        for j = 1:sample
            x = ones(1,N);
            x(randsample(N,q)) = -1;

            if(any(cumsum(x) < -c))      % notice the use of "any"
                hit = hit+1;            % no 5Eur bills left in the cashier
            end
        end
        Scurr(k) = 1-hit/sample;        % prob of no-hitting
    end

    S = [S, Scurr];
end
figure('windowstyle','docked')
histogram(S(:),'EdgeAlpha',.5,'LineWidth',2, 'DisplayStyle','bar',...
    'BinMethod','fd','EdgeColor','b')
grid
grid minor

title('Overall statistic','FontSize',14,'FontWeight','normal')

hold on
plot(prob*[1,1],[0,.25],'pk-','MarkerSize',8)
legend('simulation','exact')

kruskalwallis(S); % see Matlab documentation - produces the boxplot

title('Statistics of individual experiments','FontSize',12, 'FontWeight','normal')
xlabel('Exp.no.')
set(gcf,'windowstyle','docked')
hold on
M = median(S);
dS = std(M)/sqrt(exps);
M = mean(M);

```

```

plot([0,exps+1], prob*[1,1], 'g-', 'LineWidth', 1)
plot([0,exps+1], M*[1,1], 'r-', 'LineWidth', 1)
plot([0,exps+1], (M-dS)*[1,1], 'r-')
plot([0,exps+1], (M+dS)*[1,1], 'r-')

disp(['Check the general formula:'])
result=[sprintf('%1.4g', M), ' ± ', sprintf('%0.2g', dS)];
disp(['exp = ', result, ' - theory = ', num2str( prob)]);

```

B.3. Il problema di De Mère

```

function P = DeMereplot(n1,n2,L)
% Analisi del "caso De Mere": probabilita` di un 6 su quattro tiri di dado
% a sei facce contro probabilita` di una coppia di 6 in 24 tiri di due dadi
%
% Usage:
%   P = demereplot(n1,n2,L)
% Si eseguono "L" diversi esperimenti con un numero
% di tiri compreso tra 10^n1 e 10^n2 (n2 <= 6 e n1 >= 1).
% Default: n1 = 1, n2 = 5, L = 10.
% P contiene medie e deviazione standard delle medie
% La deviazione standard delle giocate risulta praticamente
% costante ~1/2; il diagramma riporta la deviazione standard delle
% medie.
%
% ©E. Onofri 2017-2019 (da David, Gillies, et many others)

% user interface
if nargin<3, L = 8,end;
if nargin<2, n2 = 5,end;
if nargin<1, n1 = 1,end;

if n1>n2, disp('n1 deve essere inferiore a n2!'); n1 = n2-1;end
% end user interface

n2 = min(n2,6); n1 = max(n1,1);
n = floor(logspace(n1, n2, L));

P = [];
for j = n,
    [P1,D1,P2,D2] = demere(j);
    P = [P; [P1,D1,P2,D2,j]];
end

fig = figure;
set(fig,'Windowstyle','docked');

errorbar(log(n)/log(10), P(:,1), P(:,2), '-o', 'MarkerSize', 6, ...
    'MarkerFaceColor', 'blue', 'LineWidth', 1); hold on

errorbar(log(n)/log(10), P(:,3), P(:,4), '-o', 'MarkerSize', 6, ...
    'MarkerFaceColor', 'red', 'LineWidth', 1);

grid on
grid minor

plot([n1,n2], 1-(5/6)^4*[1,1], '-.b')
plot([n1,n2], 1-(35/36)^24*[1,1], '-.r')

legend('P1', 'P2', '1-(5/6)^4', '1-(35/36)^{24}')
legend('boxoff')
xlabel('log_{10}(samples)')

```

```
%-----
function [P1, D1, P2, D2] = demere(Nsamples)
disp(['throwing dice ', num2str(Nsamples), ' times'])
if nargin<1, Nsamples = 1000; end

S1 = zeros(Nsamples,1);
S2 = zeros(Nsamples,1);

for j = 1:Nsamples
    x = randi(6,1,4); % un sei su 4 tiri
    if any(x==6), S1(j)=1;
    end

    y = randi(6,2,24); % 2 sei su 24 tiri
    if any(sum(y)==12), S2(j) = 1;
    end
end

P1 = mean(S1); D1 = std(S1)/sqrt(Nsamples-1);
P2 = mean(S2); D2 = std(S2)/sqrt(Nsamples-1);
```



B.4. Il problema dei “derangements”

(vedi Esercizio 35 a pag.50).

```
function [P,dP] = derangements(N, smpl, sweeps)
% Calcolo Montecarlo della probabilita' che una permutazione a caso non
% lasci alcun oggetto invariante - stima asintotica a N grande p=1/e

if nargin<3, sweeps=100;end
if nargin<2, smpl=10000;end
if nargin<1, N=20;end

T=[];
for sw=1:sweeps,
    n=0;
    for j=1:smpl,
        if(any(randperm(N)==1:N)), n=n+1; end;
    end;
    p= 1 - n/smpl;
    T=[T,p];
end
P=mean(T);
dP=std(T);
```

B.4.1. Subfactorial.

```
function [D,P] = subfactorial(n,r)
% Ritorna il numero di permutazioni senza punti fissi
% (derangements) di 1:n (r=0), oppure piu' in generale con $r$ punti
% fissi.
if nargin<2, r=0; end
if nargin<1, n=8; end
j=0:n;
T = gamma(j+1); % n!=gamma(n+1) %
D = -sum(1./T.*(-1).^j)*gamma(n+1);
P = D/gamma(n+1)/gamma(r+1);
```

B.5. Il problema di Buffon

(vedi *Esercizio 4* a pag.18).

Buffon needles

Buffon experiment

```
function [A,dA] = BuffonExp(Sw,d,N)
% Buffon needle experiment
% The code Buffon.m is called 'Sw' times and the result is
% analysed in statistical terms

if nargin < 3, N = 1000; end
if nargin < 2, d = .5; end
if nargin < 1, Sw = 50; end

Atot = []; dAtot = [];

for j = 1:Sw,
    [A,dA] = Buffon(d,N);
    Atot = [Atot,A];
    dAtot = [dAtot,dA];
end

errorbar(Atot,dAtot);
hold on
A = mean(Atot);
dA = std(Atot)/sqrt(Sw);
plot([1,Sw],[1,1]*(A+dA),...
      [1,Sw],[1,1]*(A-dA),...
      [1,Sw],[1,1]*A,'LineWidth',2)

grid on;
title('Buffon experiment')
xlabel('Sweeps')
ylabel('Estimate of \int \sin(x)')

% ----- subroutine Buffon.m
function [A, dA] = Buffon(d, N)
% Buffon needle experiment
if nargin < 2, N = 100, end;
if nargin < 1, d = .25, end;

L=1;

h=5*(2*rand(1,N)-1)*L;          % il foglio e` rettangolare 10 x 20
x=10*(2*rand(1,N)-1)*L;
t=rand(1,N)*pi;
nohit=(floor(h) == floor(h+d*sin(t)));

A = (1-mean(nohit))*L*pi/d;

dA = std(nohit)*L*pi/d/sqrt(N);
```



B.6. Algoritmi per le equazioni differenziali stocastiche

La tipica equazione differenziale stocastica

$$dx(t) = b(x(t))dt + \sqrt{\sigma}dw(t)$$

si può affrontare utilizzando algoritmi numerici. Il più semplice è quello di *Eulero*

$$x_{n+1} = x_n + b(x_n)\Delta t + \sqrt{\sigma\Delta t}N(0,1)$$

dove, avendo fissato una griglia temporale $\{t_0, t_1, \dots, t_N \dots\}$ si definisce $x_n = x(t_n)$ e $N(0, 1)$ è estratta a caso dalla distribuzione normale. Come conseguenza del fatto che il termine di “rumore Browniano” varia con la radice quadrata dell’intervallo temporale, l’algoritmo risulta piuttosto rozzo nel senso che la sua accuratezza scala come la radice quadrata del numero di esemplari generati, essenzialmente come in un tipico Montecarlo. Esistono vari algoritmi “migliorati” che permettono di raggiungere accuratezze che scalano con una potenza superiore $N^{-\gamma}$ nel qual caso si parla di algoritmo “di ordine γ ”. L’esempio più semplice è quello “dei trapezi”

$$x^* = x_n + b(x_n) \Delta t + \sqrt{\sigma \Delta t} N(0, 1)$$

$$x_{n+1} = x_n + \frac{1}{2}(b(x_n) + b(x^*)) \Delta t + \sqrt{\sigma \Delta t} N(0, 1)$$

(si veda il rattato di Kloeden e Platen [KP99]) per algoritmi di ordine più elevato). Il codice `linearSDE.m` nel folder `~/MPF2023/Programmi/Matlab/Montecarlo/Langevin` realizza l’algoritmo nel caso lineare (Ornstein-Uhlenbeck). ◀

B.6.1. Il codice Langevin/Parisi per le correlazioni. Il codice matlab `LangParFor12.m` calcola il gap di energia per il primo e il secondo stato eccitato di un’Hamiltoniana a un grado di libertà utilizzando il calcolo della correlazione $C_1(s, t) = \langle x(t)x(s) \rangle$ e $C_2(s, t) = \langle x(t)^2 x(s)^2 \rangle$. Applicando l’Eq.68 la correlazione C_1 si calcola dalla media lungo la traiettoria di Langevin $\frac{1}{T} \int_0^T x_1(t) dt$, mentre la C_2 è data dalla media $\frac{1}{T} \int_0^T 2x_0(t)x_1(t) dt$. Nella Fig.B.1 si riporta il risultato per una serie di 16 esperimenti condotti in parallelo.

Correlazioni

```
function [setup, results] = LangParFor12(setup)
% Parisi-Langevin algorithm for the anharmonic oscillator
% Source: lambda*x(1)^2 + mu*x(1)
% -- doing the calculation for <x x> and <x^2 x^2> in the same run
% to estimate the gap E1-E0 and E2-E0
%
% Using "parfor" if there are enough cores to speed up the calculation
% Hamiltonian = p^2/2m + x^2/2 + g x^4
% results is a structure with all data
% Calling schrlD.m to compute the spectrum through eigs
%
% Usage:
% [setup, results] = LangParFor12(setup)
% © E.Onofri 2023 Corso di 'Metodi Probabilistici della Fisica'

% Qui omettiamo i dettagli dell'input dei dati iniziali che utilizza 'inputdlg'

.....

t = linspace(0, T, N);
a = T/N; % Eucl.time lattice size

d = ones(N,1);
D2 = spdiags([-d, d],0:1,N,N); % diff.seconda discreta
D2(1,N) = 1; % condizioni periodiche
D2 = D2+D2';
D2 = D2/a^2;
clear d

tau = Tlang/steps; % Langevin time step
D = sqrt(2*tau*hbar/a); % diffusion coefficient

Xcum1 = [];
Xcum2 = [];

noise = D*randn(N,1);

src = 1; % location of source
```

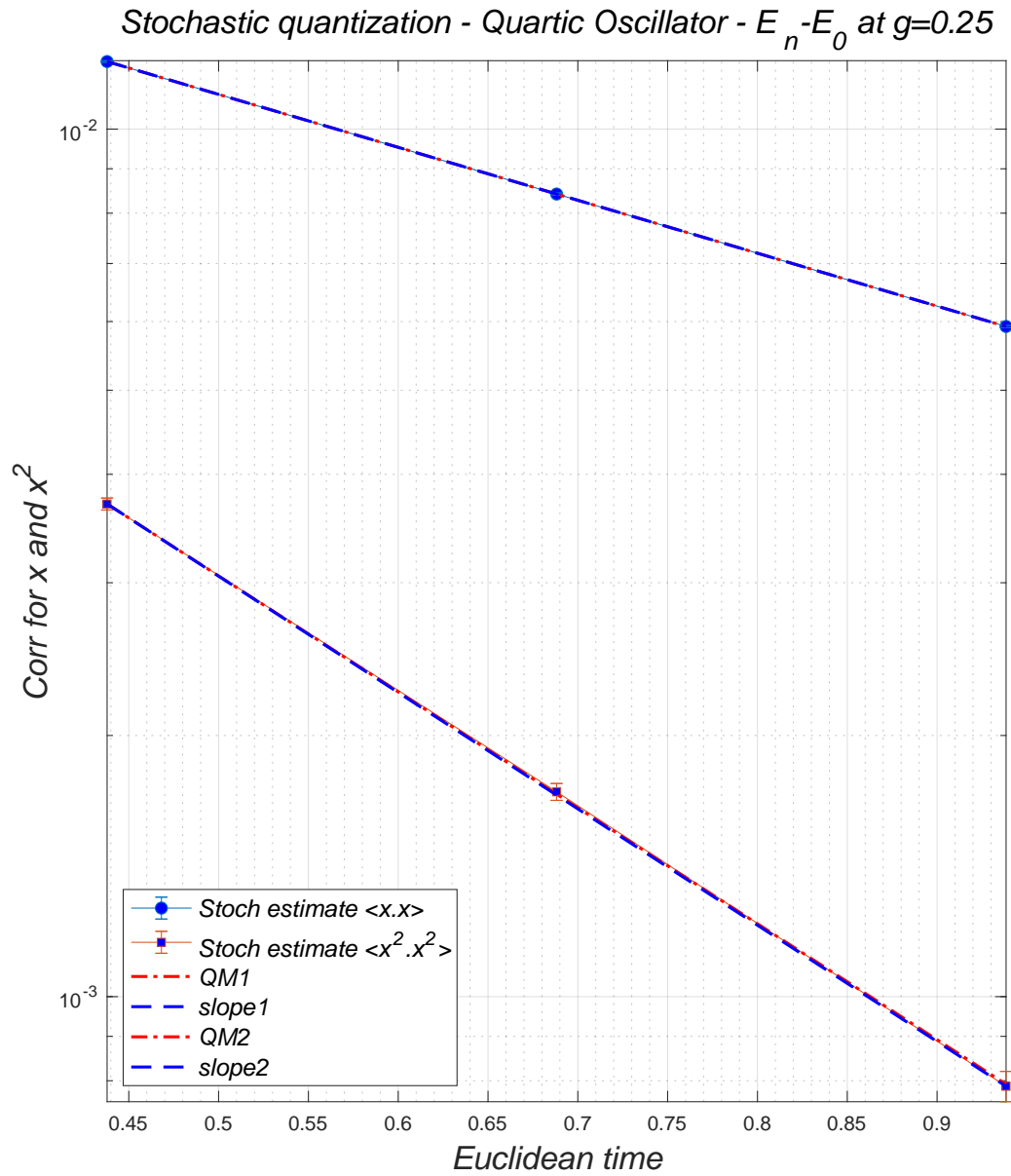


FIGURA B.1. La tecnica Langevin/Parisi per l'oscillatore anarmonico. Le correlazioni $\langle x(t)x(s) \rangle$ e $\langle x(t)^2 x(s)^2 \rangle$ vengono analizzate per valutare i primi due livelli energetici. Testo alla sezione B.6.1

```

parfor r = 1:exps                                % run "exps" histories
    x0 = zeros(N,1);
    x1 = zeros(N,1);
    x2 = zeros(N,1);
    Xc1 = [];
    Xc2 = [];

```

```

for j = 0:therm                                % start thermalization phase
    noise = D*randn(N,1);
    x0 = x0 + (D2*x0 - x0 - 4*g*x0.*x0.*x0) * tau + noise;
    x1 = x1 + (D2*x1 - x1 - 12*g*x0.*x0.*x1) * tau;
    x1(src) = x1(src) + tau;
    x2 = x2 + (D2*x2 - x2 - 12*g*x0.*x0.*x2) * tau;
    x2(src) = x2(src) + 2*tau*x0(src);
end

for j = 0:skip:steps                            % start Langevin run

    for m = 1:skip
        noise = D*randn(N,1);
        x0src = x0(src);
        % Euler step
        x0eu = x0 + (D2*x0 - x0 - 4*g*x0.*x0.*x0) * tau + noise;
        x1eu = x1 + (D2*x1 - x1 - 12*g*x0.*x0.*x1) * tau;
        x1eu(src) = x1eu(src) + tau;
        x2eu = x2 + (D2*x2 - x2 - 12*g*x0.*x0.*x2) * tau;
        x2eu(src) = x2eu(src) + 2*tau*x0src;

        % method of trapezes
        x0 = x0 + 0.5*tau*((D2*x0 - x0 - 4*g*x0.*x0.*x0)+...
            (D2*x0eu - x0eu - 4*g*x0eu.*x0eu.*x0eu)) + noise;

        x1 = x1 + 0.5*tau*((D2*x1eu - x1eu - 12*g*x0eu.*x0eu.*x1eu) +...
            (D2*x1 - x1 - 12*g*x0.*x0.*x1));
        x1(src) = x1(src) + tau;

        x2 = x2 + 0.5*tau*((D2*x2eu - x2eu - 12*g*x0eu.*x0eu.*x2eu) +...
            (D2*x2 - x2 - 12*g*x0.*x0.*x2));
        x2(src) = x2(src) + tau*(x0src + x0(src));
    end

    Xc1=[Xc1, x1];                                % correlazione < x(s) . x(t) >
    Xc2=[Xc2, 2*x2.*x0];                          % correlazione < x(s)^2 . x(t)^2 >

end;

Xcorr1 = mean(Xc1')';
Xcum1 = [Xcum1, Xcorr1];                          % average on steps*exps/skip...
Xcorr2 = mean(Xc2')';
Xcum2 = [Xcum2, Xcorr2];                          % average on steps*exps/skip...
end

interv = round(N/trim(1)):4:round(N/trim(2));
Xcum1 = Xcum1(interv,:);
Xcum2 = Xcum2(interv,:);
trim = t(interv);

E1=[];
E2=[];

for n=1:exps,
    s1 = fitcorr(Xcum1(:,n), setup, interv);
    s2 = fitcorr(Xcum2(:,n), setup, interv);        % fit the exponential decay
                                                    % get the energy gap and
                                                    % amplitude

    E1 = [E1; hbar*s1];
    E2 = [E2; hbar*s2];
end

X1 = mean(Xcum1')'; dX1 = std(Xcum1')'/sqrt(exps-1);

```



```

X2 = mean(Xcum2')'; dX2 = std(Xcum2')'/sqrt(exps-1);

[s1,f,Xf1] = fitcorr(X1, setup, interv); % fit the exponential decay
[s2,f,Xf2] = fitcorr(X2, setup, interv);

slop1 =log(X1(1)/X1(3))/(trim(3)-trim(1))
slop2 =log(X2(1)/X2(3))/(trim(3)-trim(1))

dE1 = std(E1)/sqrt(exps-1); E1 = mean(E1);
dE2 = std(E2)/sqrt(exps-1); E2 = mean(E2);
disp(['<E1-E0>=',num2str(E1),' +/- ',num2str(dE1)])
disp(['<E2-E0>=',num2str(E2),' +/- ',num2str(dE2)])

Eqm = spectrum(1, g, 8192, hbar)

dev1 = ceil(100*(Eqm(1)-E1)/dE1);
dev2 = ceil(100*(Eqm(2)-E2)/dE2);

disp(['Dev1 is ',num2str(abs(dev1)), '% of std(E1)'])
disp(['Dev2 is ',num2str(abs(dev2)), '% of std(E2)'])

figure('windowstyle','docked')
errorbar(trim, X1, dX1,'o-','MarkerSize',6,'MarkerFacecolor','b');
hold on
errorbar(trim, X2, dX2,'s-','MarkerSize',6,'MarkerFacecolor','b');

Xeigs1 = exp(-trim*Eqm(1)/hbar);
Xslop1 = exp(-trim*slop1/hbar);
Xeigs2 = exp(-trim*Eqm(2)/hbar);
Xslop2 = exp(-trim*slop2/hbar);

plot(trim,Xeigs1*X1(1)/Xeigs1(1),'r-.','LineWidth',1.5); % QM correlation
plot(trim,Xslop1*X1(1)/Xslop1(1),'b--','LineWidth',1.5); % from slope
plot(trim,Xeigs2*X2(1)/Xeigs2(1),'r-.','LineWidth',1.5); % QM correlation
plot(trim,Xslop2*X2(1)/Xslop2(1),'b--','LineWidth',1.5); % from slope

set(gca,'YScale','log')
legend('Stoch estimate <x.x>',...
       'Stoch estimate <x^2.x^2>',...% 'exp fit1','exp fit2',
       'QM1','slope1','QM2','slope2',...
       'FontSize',12,'FontAngle','it',...
       'location','southwest')
axis tight

xlabel('Euclidean time','FontSize',16,...
       'FontAngle','it')
ylabel('Corr for x and x^2','FontSize',16,'FontAngle','it')
title(['Stochastic quantization - Quartic Oscillator - E_n-E_0 at g=',num2str(g)],...
       'FontSize',16,'FontAngle','it','FontWeight','normal')

end

.....

% End main program %%

% -----

function [s, f, Xf] = fitcorr(Xc, setup, interv)
% Extract the slope of the correlation produces in LangCluster
% E(x(t)x(s)) ~f*exp(-s|t-s|)
% using fminsearch; Xf returns the fit

N = str2num(setup{1}); % extract parameters
T = str2num(setup{2});

```

```

hbar = 1. ;
t = linspace(0,T,N)';

opt = optimset('TolFun', 1e-4,...
               'TolX', 1e-4,...
               'MaxIter', 1e6,...
               'MaxFunEval', 1e6,...
               'Display','on');
ti = t(interv);

z = fminsearch(@corr , [1., 1.], opt, ti, Xc, T);

s = z(1);
f = z(2);

Xf = f*(exp(-s*t)+exp(-s*(T-t)));
end
%-----

function err = corr(z, t, Xc, T)
s=z(1);
f=z(2);

err = norm( Xc - f*(exp(-s*t) + exp(-s*(T-t))) );
end

```

B.6.2. Il codice form per lo sviluppo perturbativo.

```

                                perttth
* Perturbation series for the quartic oscillator - 1 deg.o.f.
* Based On Eq.s(66-67) in "Appunti dalle lezioni di Meccanica
Quantistica" (www.eoinfnp.it)
* On statistics;
#define MAX "12"                * perturbative series up to (g^12)
#define E0 "0"                  * ground state perturbation
S n, m;                        * commuting variables * S denotes "Symbols"
F a, ad, bra, ket, R;          * non commuting operators or "Functions"
format 250;
* "L" denotes expressions to be evaluated
L V = (a+ad)*(a+ad)*(a+ad)*(a+ad)/4; * define the perturbation operator q^4

repeat; Id a*ad = ad*a+1; * put V in normal order
endrepeat;
.sort

L VKet = V*ket('E0');          * V|0> is invariant through the iteration
L VBra = bra('E0')*V;          * and so is <0|V
L eta{0} = ket('E0'); * the perturbed eigenstate will be eta{0}+g eta{1}+g^2 eta{2} + ...

repeat;                        * basic rules For creation/annihilation op.s
Id ad*ket(n?) = ket(n+1);
Id a*ket(n?) = n*ket(n-1);
Id bra(n?)*ad = n*bra(n-1);
Id bra(n?)*a = bra(n+1);
endrepeat;
* Notice that this is a peculiar representation of creation/annihilation
* operators which avoids using square roots! The inner product is given by
*
*                               bra(n).ket(m) = n! \delta_{n,m}
.sort
#do i=1,'MAX'
L dlt{'i'} = VBra*eta{'i'-1}; * Eq.66
L eta{'i'} = -R*V*eta{'i'-1} + * Eq.67

```

```

#do p = 1,'i'-1          * statements with "#" in front are executed by the preprocessor
+ dlt{'p'}*R*eta{'i'-'p'}
#enddo
;
repeat;
Id ad*ket(n?) = ket(n+1);
Id a*ket(n?) = n*ket(n-1);
Id R*ket('E0') = 0;      * Resolvent operator
Id R*ket(n?) = ket(n)/(n-'E0');
Id bra(n?)*ket(m?) = delta_(n,m)*fac_(n)/fac_('E0');
endrepeat;
* Print dlt{'i'};
* Print eta{'i'};
.sort
# $a=dlt{'i'};
# write "%", $a
#enddo
.end

```

Il programma si generalizza al caso di sistemi con più gradi di libertà, ad es. per generare lo sviluppo perturbativo per l'oscillatore isotropo con perturbazione quartica. Può sorgere la difficoltà di incontrare livelli degeneri nel qual caso il codice senza modifiche può fallire per "divisione per zero". In un caso il codice si realizza facilmente, quello in cui la perturbazione preserva la parità e quindi non accoppia stati degeneri al primo stato eccitato. Questo accade per esempio nel caso dell'oscillatore anarmonico isotropo $V = r^2/2 + g r^4/4$. Il potenziale r^4 non accoppia gli stati degeneri $|1,0\rangle$ e $|0,1\rangle$ e il codice può tranquillamente generare la serie perturbativa per stato fondamentale e primo stato eccitato, in questo modo permettendo di valutare il gap di energia attraverso la serie perturbativa. Si può trovare il codice su [dropbox](#) nel folder `~/dropbox/MPF2023/Programmi/form/` dove sono ritrovabili anche un recente manuale e un tutorial di **Form**.

B.7. Codici per matrici random

Codici matlab corretti per lo studio delle matrici *random simmetriche reali*, *simmetriche a valori interi e infine Hermitiane* si trovano su

`~dropbox/MPF2023/Programmi/Matlab/WignerMatrices.`

Il primo riferimento è quello di Wigner [Wig67]¹ cui ha fatto seguito una mole impressionante di lavori scientifici su riviste di Matematica e di Fisica teorica. Si veda "random matrices" utilizzando *Google Scholar*: 3.700.000 indirizzi o "Cicuta random matrices" (più di 600).

Riportiamo qui soltanto le componenti essenziali dei vari codici.

B.7.1. Matrici random reali simmetriche.

```

function W = SRM(N)          % integer random matrices
    a = randn(N);
    d = diag(a);
    u = triu(a,1);
    W = diag(d) + u + u';

```

Si utilizza la routine `triu.m` per estrarre "u", la parte triangolare superiore della matrice creata con `randn`, e la si somma con la trasposta e con la diagonale.

B.7.2. Matrici random Hermitiane.

```

function W = HRM(N)          % Hermitian random matrices
    a = (randn(N)+i*randn(N))/sqrt(2);
    d = real(diag(a));
    u = triu(a,1);
    W = diag(d) + u + u';
end

```

¹Su `~/dropbox/MPF2023/Documenti/Montecarlo` si trova copia dell'articolo.

L'algoritmo è identico a quello per le matrici reali simmetriche, tranne che gli elementi di matrice sono generati come numeri complessi random (gaussiani). Lo spettro è ancora dato dalla legge del semicerchio di Wigner $\rho(x) \propto \sqrt{4N - x^2}$.

B.7.3. Matrici random unitarie.

```
function U=ue(N)
% Unitary Ensemble
% return a random unitary matrix NxN
% uniformly distributed according to Haar measure on U(N)
% Algorithm from F. Mezzadri, Bristol U. 2006
if nargin<1, N=100;end

Z = randn(N)+i*randn(N);      % Gaussian complex
[Q,R] = qr(Z);                % QR-decomposition
D = diag(R);                  %
D = diag(D./abs(D));           % (1,-1) diagonal
U = Q*D;                      % return the unitary samples
end
```

L'algoritmo fa uso della decomposizione "qr" (vedi il testo di Golub e Van Loan [GL90]). $[Q, R] = \text{qr}(M)$ costruisce una matrice unitaria Q e una triangolare superiore R tali che $M = Q \times R$.

B.7.4. Matrici random a valori interi.

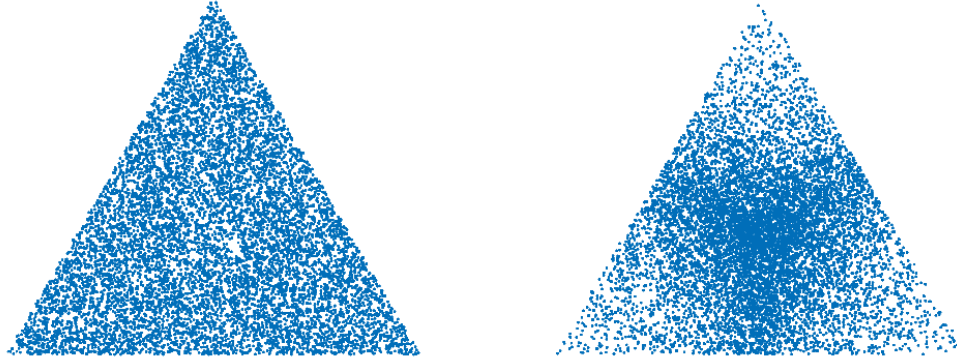
```
function W = IRM(n,N)
% Returns a random matrix with integer matrix elements in the interval [-n,n]
a = randi(2*n+1,N)-n-1;
d = diag(a);
u = triu(a,1);
W = diag(d)+u+u';
end
```

B.7.5. Matrici random stocastiche. Dato un intero N il codice seguente costruisce una matrice stocastica (elementi non negativi, somma degli elementi su ogni colonna pari a uno) estratta uniformemente nel simpleso $\Sigma = \sum_i S_{ij} = 1$.

```
xi = -log(rand(N));
S = xi ./ (ones(N,1)*sum(xi));
```

Lo spettro, ad eccezione dell'autovalore uno, presenta una distribuzione quasi uniforme nel disco $|z| \leq 1/\sqrt{N}$, con una concentrazione sull'asse reale. Questo implica che per una matrice stocastica a elementi random l'equilibrio nell'evoluzione della catena di Markov si raggiunge a un ritmo $N^{-k/2}$; cioè ad esempio per una matrice 100×100 le componenti del vettore di probabilità diverse dalla distribuzione di equilibrio vengono ridotte di un fattore 10 ad ogni passo.

Il codice richiede qualche spiegazione. Definire ξ come logaritmo di una matrice "rand" significa prendere gli elementi di matrice distribuiti secondo una densità $\rho(x) \propto \exp\{-x\}$. Ne segue che gli elementi su una colonna, essendo indipendenti, hanno una distribuzione $\rho_j(\xi_{n,j}) \propto \exp\{-\sum_n \xi_{n,j}\}$. Fissata la somma a uno, la distribuzione risulta uniforme nelle altre direzioni. Si potrebbe pensare di procedere in modo più semplice con $\xi = \text{rand}(N)$; $S = \xi ./ (\text{ones}(N,1) * \text{sum}(\xi))$. Il risultato tuttavia *non* sarebbe uniforme sul simpleso Σ . Per convincersene si può fare l'esperimento con $N = 3$ e visualizzare la distribuzione nel simpleso, ora triangolo equilatero di vertici $[1, 0, 0]$, $[0, 1, 0]$, $[0, 0, 1]$. Nelle figure seguenti, a sinistra si trova il caso con la distribuzione esponenziale a destra quella con la distribuzione uniforme nel cubo $|\xi| < 1$.



Modulo le fluttuazioni per il numero limitato di esemplari (10000), si vede bene come nel secondo caso la densità di punti diminuisce nei pressi dei tre vertici. Tutto sta naturalmente nel decidere qual'è la “vera” definizione di matrici stocastiche random. Come nel caso del paradosso di Bertrand a pag 36, non è detto che la scelta sia univoca. Nel caso unitario o reale simmetrico c'è la simmetria che suggerisce la misura di probabilità. Nel presente caso di matrici stocastiche la scelta che sembra più naturale è quella di chiedere una distribuzione uniforme degli elementi di matrice nel simpleso $\sum_i S_{ij} = 1$. Si suggerisce di effettuare esperimenti con entrambe le definizioni e studiare la statistica dello spettro. Su [~dropbox/MPF2023/Programmi/Matlab/WignerMatrices](https://www.dropbox.com/sh/MPF2023/Programmi/Matlab/WignerMatrices) si trovano implementate entrambe le scelte nel file `rand_stoch.m`.

B.8. Esercizi

AVVERTENZA. Il simboli * e ** indicano gli esercizi più impegnativi. Il simbolo # segnala che la soluzione, almeno parziale, è riportata in fondo al capitolo.

PROB B.8.1.# Si mostri in modo formale che per N esperimenti indipendenti $\{x_i\}$ lo scarto quadratico medio della media si può stimare come $\frac{\sigma}{\sqrt{N}}$, essendo σ la deviazione standard $\sqrt{\langle x^2 \rangle - \langle x \rangle^2}$.

PROB B.8.2.** Si consideri una matrice M $n \times n$ Hermitiana i cui elementi di matrice M_{ij} con $i \geq j$ sono estratti a caso in modo indipendente da una distribuzione Gaussiana normale. Determinare la densità di probabilità $P(\lambda \in [x, x + dx])$ per gli autovalori, nel limite di n molto grande. La distribuzione è nota come “legge del semicerchio di Wigner” ([Wig67, Meh89]).

PROB B.8.3.** Costruire un algoritmo che permetta di estrarre matrici unitarie a caso, uniformemente rispetto alla misura invariante $[dU] = [d(gU)] = [d(Ug)]$.

PROB B.8.4.** Si consideri l'insieme delle matrici **simmetriche** a elementi interi $M_{ij} \in \mathbb{Z}$ costruite secondo la regola: ogni elemento di matrice M_{ij} con $i \geq j$ è scelto in modo indipendente in modo casuale con probabilità uniforme nell'intervallo $(-n, -n + 1, \dots, n - 1, n)$. Ogni valore nell'intervallo è estratto perciò con probabilità $1/(2n + 1)$.

Costruire un algoritmo Matlab per costruire in modo efficiente un insieme di matrici di questo tipo e studiarne lo spettro. Per matrici di grandi dimensioni si ritrova una distribuzione degli autovalori identica a quella già incontrata (legge del semicerchio). Individuare il raggio del semicerchio in funzione di n e della dimensione N delle matrici.

PROB B.8.5. Una pulce esegue un salto ogni minuto primo da uno all'altro di quattro cani C_1, \dots, C_4 . La decisione avviene (apparentemente) a caso con probabilità di salto $C_j \rightarrow C_k$ data dalla matrice stocastica

$$P(C_j \rightarrow C_k) = \begin{pmatrix} 0 & 1/4 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 \\ 0 & 1/2 & 0 & 1/2 \\ 1/2 & 1/4 & 1/2 & 1/2 \end{pmatrix}$$

Si calcoli la probabilità che la pulce si trovi sul cane C_1 dopo un'ora dalla prima rilevazione che la trovava su C_2 . Dimostrare che il risultato è praticamente indipendente dal cane di partenza.

PROB B.8.6.[#] Si consideri il gruppo S_n delle permutazioni su n oggetti. Si chiede qual'è la probabilità che una permutazione scelta a caso non lasci fisso alcun elemento? (Una variante del problema dice: 40 signori lasciano il proprio cappello all'ingresso del teatro. All'uscita viene a mancare l'energia elettrica e i cappelli vengono riconsegnati al buio. Qual'è la probabilità che nessuno ritorni in possesso del proprio cappello?).

B.9. Soluzioni

SOLUZIONE. **B.8.1** Sia $\langle X \rangle = \langle \frac{1}{N} \sum_{i=1}^N x_i \rangle = \mu$ la media degli esperimenti tutti ugualmente distribuiti in probabilità e statisticamente indipendenti. La deviazione standard di X è

$$\begin{aligned} \Delta^2 X &= \langle X^2 \rangle - \mu^2 = \frac{1}{N^2} \left\langle \sum_i \sum_j x_i x_j \right\rangle - \mu^2 \\ &= \frac{1}{N} \langle x^2 \rangle + \frac{N(N-1)}{N^2} \mu^2 - \mu^2 = \frac{1}{N} (\langle x^2 \rangle - \mu^2) \end{aligned}$$

In pratica, mentre ogni singolo esperimento è regolato dalla stessa distribuzione di probabilità, la media di N esperimenti tenderà a concentrarsi sul valore medio a priori con deviazione che diminuisce con l'accumularsi di risultati. È essenziale che i successivi esperimenti siano statisticamente indipendenti tra loro in modo da potere applicare la formula $\langle x_i x_j \rangle = \langle x_i \rangle \langle x_j \rangle$. ◀

SOLUZIONE. **B.8.2** Si vedano [Wig67, Meh89] per l'approccio analitico che risale alla metà del secolo scorso. Il problema si può esplorare "sperimentalmente" con ² *matlab*:

```

function E=wigner(N, samples)
E=[];
for n=1:samples
    M = randn(N); % distrib. normale per gli elementi di matrice
    M=(M+M')/2; % simmetrizzazione
    E=[E; eig(M)]; % accumula gli autovalori
end
hist(E, samples*N/1000); % disegna l'istogramma

```

◀

Si noterà che gli autovalori si distribuiscono secondo una densità concentrata nell'intervallo $[-\sqrt{2N}, \sqrt{2N}]$. Ciò si ripete nel caso di matrici hermitiane

```

function E=wigner(N, samples)
E=[];
for n=1:samples
    M = randn(N)+i*randn(N);
    M=(M+M')/2/sqrt(2);
    E=[E; eig(M)];
end
hist(E, samples*N/1000);

```

²Il codice che riportiamo qui contiene in realtà un'imprecisione, che però diventa irrilevante per grandi valori di N in quanto la diagonale della matrice influisce poco sul risultato. Un algoritmo esatto si trova nelle routines *HermRandMat.m*, *SymRandMat.m* e *IntRandMat.m*.

Un piccolo numero di esperimenti può essere sufficiente per stabilire la legge generale. Per la teoria si vedano i riferimenti consigliati. ◀

SOLUZIONE. **B.8.3** Il caso unitario è incluso negli studi di Mehta et al, quindi i riferimenti sono gli stessi. Per eseguire esperimenti numerici si può utilizzare un algoritmo non banale ³

```

function U=ue(N)
% Return a random unitary matrix NxN
% uniformly distributed according to Haar measure on U(N)
% Algorithm by F. Mezzadri, Bristol U. 2006
Z = randn(N)+i*randn(N);      % Gaussian complex
[Q,R] = qr(Z);                % QR-decomposition
D = diag(R);
D = diag(D./abs(D));          % (1,-1) diagonal
U = Q*D;                      % return the unitary sample

```

◀

SOLUZIONE. **B.8.4** Si utilizza la routine `randi` che estrae numeri interi in modo uniforme:

```
>> M = randi(2*n+1,N)-n-1 ; matrice NxN di interi random
```

Ogni elemento di matrice sarà così un intero compreso tra $-n$ e n . Per rendere la matrice simmetrica si procede così : si estrae la diagonale

```
>> D=diag(M);
```

quindi la parte triangolare superiore

```
>> U=triu(M,1);
```

e infine si ricomponi una matrice simmetrica:

```
>> M = diag(D)+U+U';
```

Il codice è `irm.m` mentre un esperimento per studiare la statistica degli autovalori di queste matrici si può condurre utilizzando il codice `IrmRandMat.m`. Si trova che lo spettro coincide con quello del semicerchio per dimensioni sufficientemente grandi. ◀

SOLUZIONE. **B.8.6** Per fissare le idee: S_3 è costituito dalle sei permutazioni

$$\begin{pmatrix} 123 \\ 123 \end{pmatrix} \begin{pmatrix} 123 \\ 312 \end{pmatrix} \begin{pmatrix} 123 \\ 231 \end{pmatrix} \begin{pmatrix} 123 \\ 132 \end{pmatrix} \begin{pmatrix} 123 \\ 213 \end{pmatrix} \begin{pmatrix} 123 \\ 321 \end{pmatrix}$$

dunque $D(3) = 1/3$. Ma il numero di permutazioni è $n!$ e non è quindi il caso di procedere nello stesso modo per $n = 16$ che comporta circa 1.3×10^{12} permutazioni. Utilizzando la funzione `randperm` è molto facile farsi un'idea di quanto può valere $D(16)$. Secondo la simulazione (20 pacchetti di 10000 permutazioni scelte a caso, molto inferiore al numero totale di permutazioni possibili) si ha che $D(16) \approx .369 \pm 0.004$ (Si veda la Fig. 2.17). Il codice sta, volendo, in una riga:

```

function p = derangements(N, smpl)
n=0; for j=1:smpl, if(any(randperm(N)==1:N)), n=n+1; end; end;p=n/N;

```

Vediamo cosa ci può dire la formula di inclusione/esclusione. Partiamo dal fatto che per ogni i ($1 \leq i \leq n$) l'insieme delle permutazioni che lascia fisso i è un gruppo di permutazioni su $n-1$ elementi, indichiamolo con A_i . Inoltre $A_i \cap A_j$ è l'insieme di permutazioni che lascia due elementi fissi, e così via. Avremo perciò (indicando con $|A|$ il numero di elementi di A)

$$\begin{aligned}
 n! D(n) &= n! - \sum_i |A_i| + \sum_{i>j} |A_i \cap A_j| - \sum_{i>j>k} |A_i \cap A_j \cap A_k| + \dots \\
 &= n! - \binom{n}{1} (n-1)! + \binom{n}{2} (n-2)! - \dots + (-1)^k \binom{n}{k} (n-k)! + \dots + (-1)^n \\
 &= \sum_{k=0}^n (-1)^k \frac{n!}{k!(n-k)!} (n-k)!
 \end{aligned}$$

³Ringrazio vivamente F. Mezzadri per avermi fornito il codice.

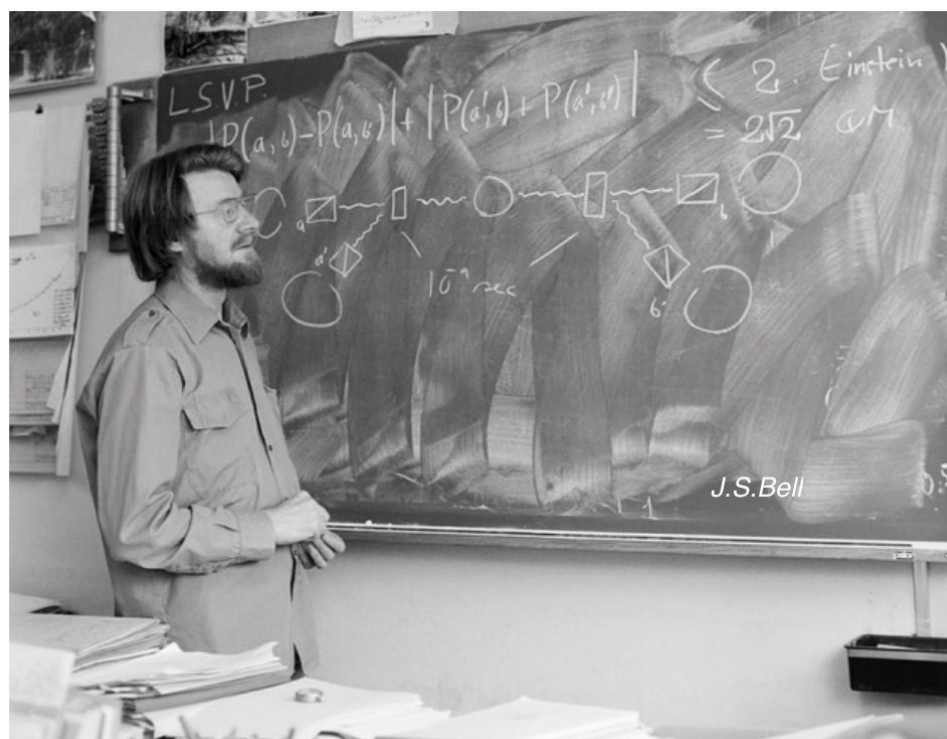
per cui

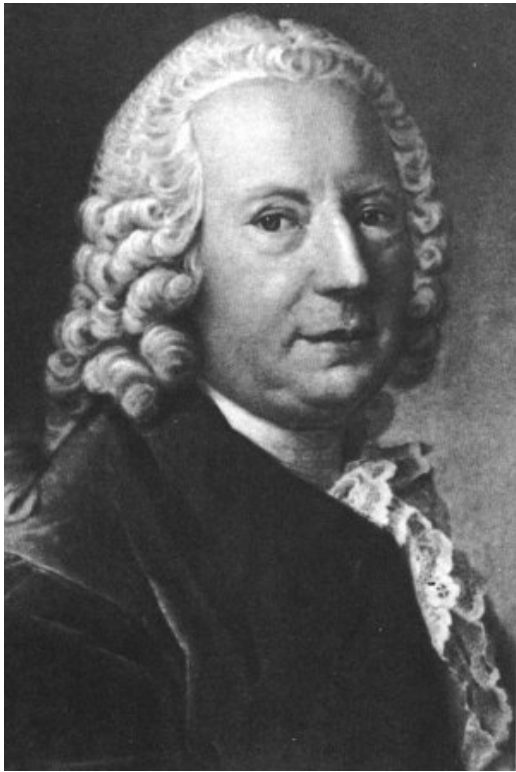
$$D(n) = \sum_{k=0}^n (-1)^k \frac{1}{k!}.$$

Per n grande si ha perciò $D(n) \approx \exp(-1) \approx .3679$. Il problema era stato risolto già da Poincaré [**Poi12**] dove si trova anche la formula di inclusione/esclusione (vedi Pag. 49). ◀

APPENDICE C

Galleria di ritratti

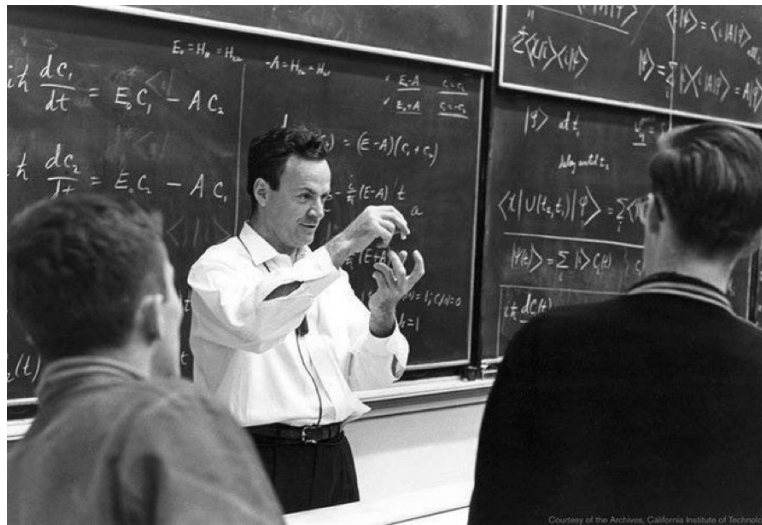




D. Bernoulli



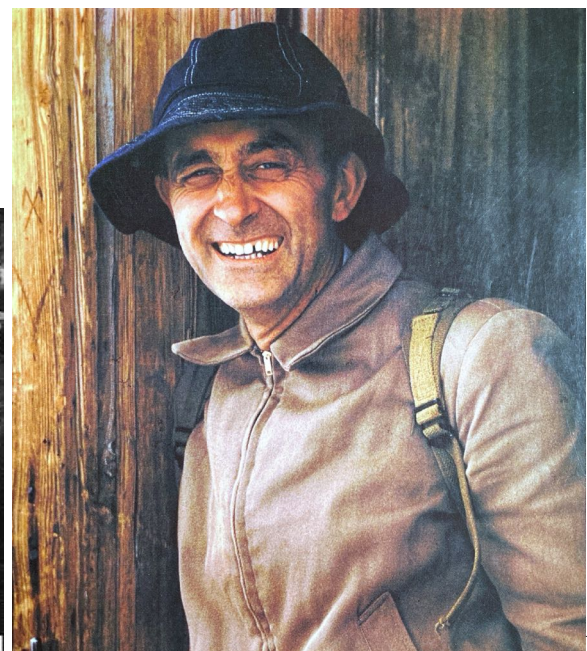
B. Riemann



R. Feynman



Fermi



E.



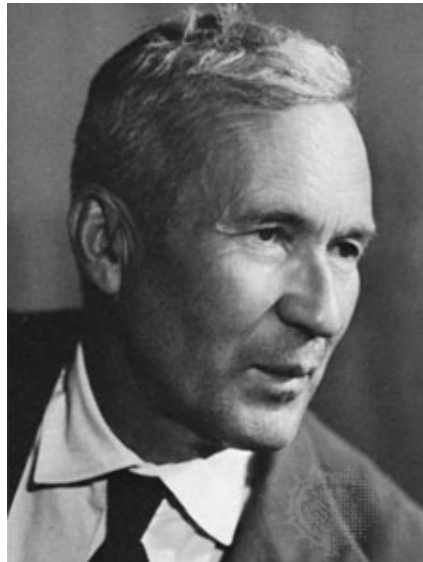
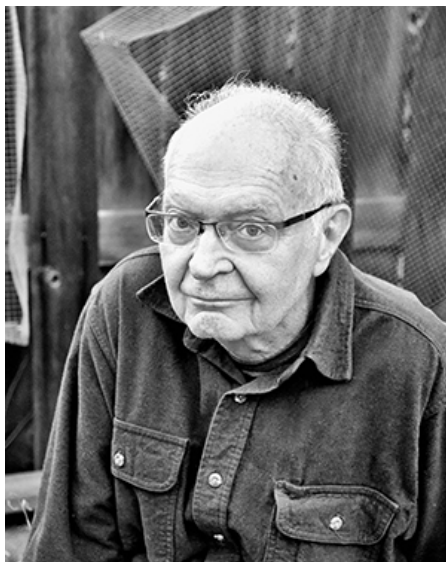
A portrait of a man with a white cravat and a dark coat, looking slightly to the right. The background is dark and textured.



C.F. Gauss and D. Hilbert



M. Kac



D. Knuth and A.N. Kolmogorov



A. Markov



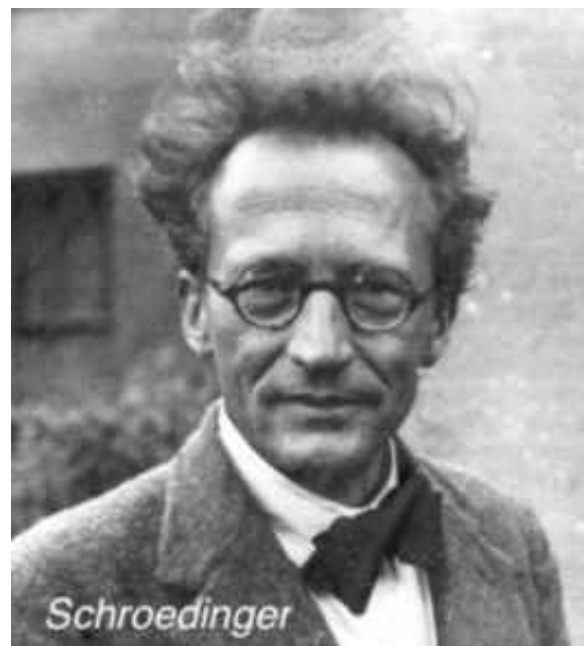
M.L. Mehta



B. Pascal e H. Poincaré



G. Polya





J. Von Neumann e E. Wigner



Stan Ulam

Indice analitico

- \mathbb{D} , 24, 54
- \mathbb{E} , 24, 54
- \mathbb{P} , 9
- $\mathbb{P}(A|B)$, 44
- $\delta(x)$, 55
- $\zeta(s)$, 26
- ehrenplot.m, 81
- polyfitweighted.m, 31
- BuffonExp.m, 19
- ballot.m, 35
- demere.m, 27
- demoivre.m, 64
- gambler.m, 36
- lyapounov.m, 69

- alias
 - metodo degli, 101
- Amir, A., 42

- Baldi, 94
- Bayes
 - formula di, 44
- Bell, ritratto 161
 - disuguaglianza di, 73
- Bernoulli, 54, 61, 63, 134, ritratto 162
- Bertrand
 - paradosso di, 36
- Biglie indistinguibili, 41
- Bilancio dettagliato
 - principio del, 107
- binomiale
 - distribuzione, 17, 61, 62
- Bohm D., 73
- Bookpile, 89
- Borel-Cantelli
 - Lemma, 70
- botteghino
 - codice per il problema del, 146
 - il problema del, 30
- boxplot, 34
- Buffon, 18
 - codice per il problema di, 149
 - esperimento di, 18
- Calore specifico, 112
- Catene di Markov, 81
- CERN/CMS, 62
- Chebyshev
 - disuguaglianza di, 62
 - teorema di , 62
- Cicuta, G., 2, 155, 158
- coda alla biglietteria, 33
- Codice di Langevin/Parisi, 150
- Codici matlab , 149
- Codici matlab , 145
- coefficienti binomiali, 38
- correlazione quantistica, 144

- De Mère, 26, 27
 - codice per il problema di, 147
- De Moivre
 - teorema di, 63
- Decadimento α , 67
- decomposizione spettrale, 83
- delta di Dirac, 55
- Derangements, 50, 56
 - codice Montecarlo, 148
- Destri, 2, 96
- Deviazione standard, 54
- deviazione standard della media, 158
- distribuzione
 - binomiale, 17, 28, 61, 62, 65, 134
 - di Bernoulli, 54, 134
 - di Gauss, 58, 64, 134
 - di Gauss multidimensionale, 58
 - di Poisson, 65, 66, 134
 - di Polya, 68
 - funzione di, 54
 - ipergeometrica, 84, 138
 - normale, 58, 134
 - stazionaria, 83
- drift, 113

- Ehrenfest, ritratto 164
 - modello continuo di, 94
 - modello di, 81
 - modello modificato, 83
 - soluzione del modello di, 137

- ensemble
 - unitario, 159
 - unitario, 156
- entangled
 - stato, 73n
- errore della media, 31
- Esempi, 15, 18, 56, 87, 89, 101
- Esercizi, 16, 17, 24--26, 28, 30, 33, 35,
 - 36, 44, 46, 50, 61, 67, 68, 83--85,
 - 88, 93, 94, 139
- Esercizi su catene di Markov, 86--101
- eventi
 - indipendenti, 43
 - spazio degli, 13
- f.d.d., 54
- f.g., 56
- Faris: un semplice caso di disuguaglianza di Bell, 73
- Fermi, ritratto 163
- Fermi, E., 99, ritratto 163
- Feynman, ritratto 163
- Fisz, 11, 68
- fitdist, 65, 72
- fitgmdist, 72
- Fokker-Planck, 116, 141
- Form
 - language, 2, 154
- Funzione $\zeta(s)$ di Riemann, 26
- Funzione di distribuzione, 54
- Funzione generatrice, 56
 - della somma di v.a., 61
 - e legge dei grandi numeri, 63
 - per i derangements, 56
 - per il problema del botteghino, 59
 - per il tempo di ritorno, 85
 - per la distribuzione Gaussiana, 58
- Galileo, 26n
- gambler's ruin, 35
- Gauss, ritratto 164
 - distribuzione di, 58, 64
- Generatori di numeri casuali, 133
- Gershgorin
 - teorema di, 136
- Gillies, 10n
- Gnedenko
 - metodo della riflessione, 33
- GSL, library, 112
- Hausdorff
 - esempio di distribuzione non normale, 70
- Heisenberg, ritratto 163
- Higgs
 - particella di, 62
- Hilbert, ritratto 164
- histfit, 65, 72
- inclusione/esclusione, 49
- indipendenza statistica, 43
- Ising
 - modello di, 106, 112, 120
- Ising 2D
 - magnetizzazione spontanea, 111n
- Kac, 11, 81, 99, ritratto 165
- Knuth, ritratto 165
 - metodo degli alias, 102
- Kolmogorov, 113, 141, ritratto 165
- Langevin, ritratto 164
 - equazione di, 112
- Langevin-Parisi
 - codice di, 150
- Laplace, ritratto 162
- Legge dei grandi numeri, 63
- legge del semicerchio, 157
- Lemma di Borel-Cantelli, 70
- Lepage, G.P., 112
- Limite centrale
 - teorema del, 68
- linear congruential methods, 133
- Lyapounov, 69
- Magnetizzazione, 109, 111n
 - modello di Ising, 124
 - modello di Potts, 129
- Markov, ritratto 166
 - catene di, 81
 - processi di, 92
- Markov chains, 81
- Markov processes, 93
- Mathematica, 19, 37, 52, 96, 139
- matlab
 - boxplot, 34
 - funzione anonima, 24
 - nchoosek, 15
 - polyfitweighted, 31
 - polynomial fit, 31
- matrice
 - irriducibile, 85
 - stocastica, 82
- matrici random, 155
 - a elementi interi, 156, 157
 - hermitiane, 155, 157
 - reali simmetriche, 155
 - stocastiche, 156
 - unitarie, 156, 157, 159

- Mehta, 158, ritratto 166
- Mersenne, 11n
- Mersenne twister, 133
- metodo
 - dei trapezi, 150
- metodo Montecarlo, 99
- Metodo degli alias, 101
- metodo dei trapezi, 144
- Metropolis
 - algoritmo di, 107
- Metropolis, N., 10, 99
- Mezzadri, 159n
- Micciola
 - tesi, 10n
- misura di probabilità, 13
- modello di Ehrenfest, 81
- modello di Ising
 - in due dimensioni, 111
 - in tre dimensioni, 126
 - in una dimensione, 110
- modello di Potts, 109
- Montecarlo
 - metodo, 99
- Monty Hall problem, 45
- Mosteller, 21, 28
- Moto Browniano, 85, 112, 141
 - correlazione, 114
- Moto Browniano
 - in campo gravitazionale, 115
- nchoosek(), 15
- Nelson, 116
- non-universalità della distribuzione di Gauss, 70
- Normale
 - distribuzione, 58
- oeis.org (online encyclopedia), 37
- Onsager, 106n
 - formula di, 111n
- Ornstein-Uhlenbeck, 114
- Padé
 - approssimanti di, 119
- Paradosso di Bertrand, 36
- parametro di rilassamento, 94
- Parisi, 143
 - calcolo della correlazione, 118
- Parisi-Langevin, 118
- Parisi-Wu, 112, 116, 117
- Pascal, 26n, ritratto 167
- Pauli, ritratto 163
- Pelissetto A., 133
- Pelissetto A., 11
- permutazioni, 50, 158
 - algoritmo di Knuth-Lehmer, 89
- Perron-Frobenius
 - teorema di, 136
- Pila di libri
 - problema della, 89
- pione, 73n
- Poincaré, 160, ritratto 167
- Poisson
 - distribuzione di, 66
- poker
 - codice, 15
 - probabilità nel gioco del, 15
 - simulazione, 17, 145
- Polya, ritratto 167
 - distribuzione di, 68
- Polyakov, 144
- Potts
 - modello di, 109, 112
- Probabilità
 - condizionata, 44
 - formula della probabilità totale, 44
 - spazio di, 13
- Probabilità
 - misura, 14
- problema
 - dei portalettere, 101
 - del botteghino, 30
 - codice matlab, 34
 - del giocatore, 35
- Processi di Markov, 81
- Processi diffusivi, 141
- Processi stocastici, 81
- Quantizzazione stocastica, 116
- Ramsey-De Finetti, 14
- Random walk, 84
- random variables, 54
- Random walk, 87
 - asimmetrico, 114, 115
- Random walk sul cerchio, 88
- randperm, 52, 159
- Riemann, ritratto 162
 - funzione ζ , 26
- Ritorno
 - tempo medio di, 85
- Ritorno al punto di partenza, 85
 - per il random walk, 87
- Rosenbluth, A., 107
- Roulette, 28, 29
- Rutherford, 67
- SageMath, 37

Schroedinger, ritratto 167
Sokal A., 19
Somma di variabili aleatorie, 60
 gaussiane, 64
Stieltjes
 integrale di, 55
Subfactorial[n], 52
successioni casuali, 100

tempo medio di ritorno, 85, 139
Teorema
 del limite centrale, 68
 di Gershgorin, 136
 di Perron-Frobenius, 136
trapezi
 metodo dei, 150

Ulam S., 10, 107, ritratto 168

variabili
 aleatorie, 54
varianza, 54
Vegas, 112
Vermaseren, J., 2
Von Neumann, J., 107, ritratto 168

Walker
 algoritmo di, 102
Whittaker-Robinson
 esempio di Hausdorff, 70
Wick, 'The infamous boundary', 73
Wigner
 legge del semicerchio, 157
wigner.m, 158
Wigner, E., 155, 158, ritratto 168

Bibliografia

- [AAR82] J. Dalibard A. Aspect and G. Roger. Experimental tests of bell's inequalities using time-varying analyzers. *P.R.L.*, 25:1804, 1982.
- [AER35] B. Podolsky A. Einstein and N. Rosen. Can Quantum-Mechanical Description of Physical Reality be Considered Complete? *Phys. Rev.*, 47:777–780, 1935.
- [AJ82] F. C. Alcaraz and L. Jacobs. Z(N) generalization of the Baxter-Wu model. *J. Phys. A: Math Gen.*, 15:L357–L361, 1982.
- [Ami21] A. Amir. *Thinking probabilistically*. Cambridge U.P., Cambridge and New York, 2021.
- [BCR84] K. Baclawski, M. Cerasoli, and G.C. Rota. *Introduzione alla Probabilità*. UMI, 1984.
- [Bel64] J. S. Bell. On the Einstein-Podolsky-Rosen paradox. *Physics*, 1(3):195, 1964.
- [Bel87] J. S. Bell. *Speakable and unspeakable in quantum mechanics*. Cambridge UP, Cambridge, 1987.
- [Ber04] B. Berg. *Markov Chain Monte Carlo simulations and their statistica analysis*. World Scientific, 2004.
- [BP12] P. Butera and M. Pernici. High-temperature expansions of the higher susceptibilities for the Ising model in general dimension d. *Phys.Rev. E*, 86:011139, 2012.
- [BS99] I. Beichi and F. Sullivan. Pay me now or pay me later. *Computing in Science and Engineering*, page 59, July/August 1999.
- [BV12] G. Boffetta and A. Vulpiani. *Probabilità in Fisica*. Springer, 2012.
- [BW14] B.M.McCoy and T.T. Wu. *The Two Dimensional Ising model*. Dover, 2014.
- [C.98] Bruce C. *The Strange Case of Mrs Hudson Cat*. Vintage, Random House, London, 1998.
- [Com74] L. Comtet. *Advanced Combinatorics*. Reidel, 1974.
- [Cre83] M. Creutz. *Quarks, gluons and lattices*. Cambridge University Press, 1983.
- [D'A03] G. D'Agostini. *Bayesian Reasoning in Data Analysis*. World Sci., 2003.
- [D'A12] G. D'Agostini. Scoperte scientifiche annunciate a colpi di “sigma”. Incontri di Fisica, Università di Roma La Sapienza, 2012.
- [Dav62] F. N. David. *Games, Gods and Gambling*. Hafner Pub. Co., New York, 1962.
- [DH87] P.H. Damgaard and H.Hüffel. Stochastic Quantization. *Phys.Rep.*, 152((5-6)):227, 1987.
- [DH97] A.C. Davison and D.V. Hinkley. *Bootstrap methods and their application*. Cambridge U.P., 1997.
- [DSC98] P. Diaconis and L. Saloff-Coste. What do we know about the Metropolis Algorithm? *J. Computer ans System Sciences*, 57:30–36, 1998.
- [EE07] T. Ehrenfest and P. Ehrenfest. Über zwei bekannte Einwände gegen das Boltzmannsche H-Theorem. *Physikali-sche Zeitschrift*, 8:311–314, 1907.
- [Efr79] B. Efron. ”bootstrap methods: Another look at the jackknife”. *The Annals of Statistics*, 7((1)):1–26, 1979.
- [Efr82] B. Efron. *The Jackknife, the Bootstrap and other resampling plans*. SIAM, 1982.
- [Fel68] W. Feller. *An Introduction to Probability Theory and its Applications*. Wiley series in probability and mathematical statistics. J. Wiley & Sons, New York, 1968.
- [Fey72] R. P. Feynman. *Statistical Mechanics*. W. A. Benjamin, Inc., Reading, Mass., 1972.
- [Fis63] M. Fisz. *Probability Theory and Mathematical Statistics*. J. Wiley & Sons, New York, 3rd edition, 1963.
- [FLB75] Guerra F., Rosen L., and Simon B. The $P(\phi)_2$ euclidean quantum field theory as classical statistical mechanics. *Ann.of Math.*, 101(2):191–259, 1975.
- [Gan59] F. R. Gantmacher. *The Theory of Matrices*. Chelsea Pub. Co., New York, 1959.
- [GGAT78] G.C. Ghirardi, G.Omero, A.Rimini, and T.Weber. The stochastic interpretation of Quantum Mechanics. *Rivista del Nuovo Cimento*, 1(3):1–34, 1978.
- [Gil00] D. Gillies. *Philosophical Theories of Probability*. Routledge, London and New York, 2000.
- [Gin00] M. Gintis. *Game Theory Evolving*. Princeton UP, 2000.
- [GL90] G.H.Golub and C.F.Van Loan. *Matrix Computations*. Johns Hopkins U.P., 1990.
- [Gne79] B. Gnedenko. *Teoria della Probabilità*. MIR, 1979.
- [GR65] L. S. Gradshteyn and I. M. Ryzhik. *Table of Integrals, Series and Products*. Academic Press, New York, IV edition, 1965.

- [Hit03] D. B. Hitchcock. A History of the Metropolis-Hastings Algorithm. *The American Statistician*, 53(4):254–257, 2003.
- [HKW78] A. Hintermann, H. Kunz, and F.Y. Wu. Exact results for the Potts Model in two dimensions. *J. Stat. Phys.*, 19(6):623–632, 1978.
- [HW41] H.A.Kramers and G.H. Wannier. Statistics of the two-dimensional ferromagnet. part i. *Phys.Rev.*, 60:252, 1941.
- [ID89] C. Itzykson and J. M. Drouffe. *Théorie Statistique des Champs*. InterEditions, Paris, 1989.
- [I.J11] I.J.Zucker. 70+ years of the watson integral. *J.Stat.Phys.*, 145:591–612, 2011.
- [J.82] Baxter R. J. *Exactly Solved Models in Statistical Mechanics*. Acad. Press, London, 1982.
- [Jay58] E.T. Jaynes. Probability Theory in Science and Engineering. Colloquium Lectures in Pure and Applied Science, Dallas, TX, 1958. Field Research Lab., Socony Mobil Oil Co.
- [Jef38] H. Jeffreys. *Theory of Probability*. Oxford at the Clarendon Press, 1938.
- [Kac46] M. Kac. Random walks and the theory of brownian motion. In N. Wax, editor, *Selected papers on Noise and Stochastic Processes*. Dover, 1946.
- [Kac59] M. Kac. *Probability and Related Problems in Physical Sciences*. Interscience, New York, 1959.
- [Knu98] D. Knuth. *The art of computer programming*. Addison–Wesley, 1998.
- [KP99] P. E. Kloeden and E. Platen. *Numerical Solution of Stochastic Differential Equations*, volume 23 of *Applications of Mathematics*. Sringer, Berlin, 1999.
- [LCM06] M. Lax, Wei Cai, and MinXu. *Random Processes in Physics and Finance*. Oxford U.P., 2006.
- [Lep78] G.P. Lepage. A New Algorithm for Adaptive Multidimensional Integration. *J. Comput. Phys.*, 27:192–203, 1978.
- [Mal03] G. Maltese. *Enrico Fermi in America*. Le Ellissi. Zanichelli, Bologna, 2003.
- [M.D91] Vose M.D. A Linear Algorithm For Generating Random Numbers With a Given Distribution. *IEEE Trans. Soft. Eng.*, 17(9):972–974, 1991.
- [Meh89] M.L. Mehta. *Matrix Theory, Selected Topics and Useful Results*. Les Editions de Physique, Les Ulis (F), 1989.
- [MI77] M.L.Glasser and I.J.Zucker. Extended Watson integrals for the cubic lattices. *Proc.Nat.Acad.Sci.*, 74(5):1800–1801, 1977.
- [MN98] M. Matsumoto and T. Nishimura. Mersenne twister: a 623-dimensionally equidistributed uniform pseudo-random number generator. *ACM Transactions on Modeling and Computer Simulation*, 8((1)):3–30, 1998.
- [Mos65] F. Mosteller. *Fifty challenging problems in probability with solutions*. Dover, New York, 1965.
- [MU49] Nicholas Metropolis and S. Ulam. The monte carlo method. *Journal of the American Statistical Association*, 44(247):335–341, 1949. PMID: 18139350.
- [Nel67] E. Nelson. *Dynamical Theories of Brownian Motion*. Princeton University Press, Princeton, N.J., 1967.
- [Nym72] J. E. Nymann. On the probability that k positive integers are relatively prime. *J. Number Theory*, 4:469–473, 1972.
- [OD96] E. Onofri and C. Destri. *Istituzioni di Fisica Teorica*. La Nuova Italia Scientifica, 1996.
- [Ono84] E. Onofri. *Lezioni sulla Teoria degli Operatori Lineari*. Ediz. Zara, Parma, 1984.
- [Ons44] L. Onsager. Crystal Statistics. I. A two-dimensional model with an order-disorder transition. *Phys.Rev.*, 65:117–149, 1944.
- [Par81] G. Parisi. Correlation function and computer simulations. *Nuclear Physics*, B180 [FS2]:378–384, 1981.
- [P.B98] P.Baldi. *Calcolo delle probabilità e statistica*. Collana di istruzione scientifica, serie di Matematica. McGraw-Hill, Milano, 1998.
- [Pel94] A. Pelissetto. *Introduction to the Monte Carlo method*, in “Elementary particles, Quantum Fields and Statistical Mechanics”, M. Bonini, G. Marchesini and E. Onofri Ed.s, chapter III. Università di Parma, 1994.
- [Pen89] R. Penrose. *The Emperor New Mind*. Oxford U. Press, 1989.
- [Pen97] R. Penrose. *La mente nuova dell'imperatore*. BUR. Rizzoli, Milano, 1997.
- [Poi12] H. Poincaré. *Calcul des Probabilités*. (disponibile su gallica.fr), Paris, gauthiers–villars edition, 1912.
- [Pol77] A. M. Polyakov. Quark confinement and topology of gauge theories. *Nucl.Phys.*, 120B:429–458, 1977.
- [Pot52] R. P. Potts. Some generalized order-disorder transformations. *Proc. Camb.Philos. Soc.*, 48:106, 1952.
- [PR69] Yu. V. Prohorov and Yu. A. Rozanov. *Probability Theory*. Springer Verlag, Berlin, 1969.
- [PW81] G. Parisi and Y.-S. Wu. ”perturbation theory without gauge fixing”. *Sci. Sinica*, pages 483–496, 1981.
- [QRF00] A. Quarteroni, R.Sacco, and F.Saleri. *Matematica Numerica*. Springer–Verlag Italia, Milano, 2000.
- [Reb83] C. Rebbi. *Lattice Gauge Theory and Montecarlo simulations*. World Sci., Singapore, 1983.
- [RG10] E. Rutherford and H. Geiger. The probability variations in the distribution of alfa particles. *Phil. Mag.*, pages 698–707, 1910.
- [Roz69] Yu. A. Rozanov. *Probability Theory: a concise course*. Dover, 1969.
- [RPP12] A. Rotondi, P. Pedroni, and A. Pievatolo. *Probabilità, Statistica e Simulazione*. Springer, iii edition, 2012.
- [Sch17] David N. Schwartz. *The last Man Who Knew Everything*. Basic books, 2017.
- [Sim83] B. Simon. Some quantum operators with discrete spectrum but classically continuous spectrum. *Annals of Physics*, 146:209–220, 1983.

- [Sok91] A. D. Sokal. How to beat critical slowing down: 1990 update. In U.M.Heller et al, editor, *Lattice 90*. North-Holland, 1991.
- [Sok97] A. D. Sokal. *Functional integration (Cargèse, 1996)*, chapter Monte Carlo methods in Statistical Mechanics: Foundations and New Algorithms, pages 131–192. Plenum, New York, 1997.
- [Sta99] R. P. Stanley. *Enumerative Combinatorics*, volume 2. Cambridge U.P., 1999.
- [Var62] R. S. Varga. *Matrix iterative analysis*. Prentice–Hall, Englewood Cliffs, N.J., 1962.
- [Vea22] J.A.M. Vermaseren et al. *FORM reference manual*, 2022.
- [Wal77] A. J. Walker. An efficient method for generating discrete random variables with general distributions. *ACM Trans. Math. Softw.*, 1977.
- [Wax54] N. Wax, editor. *Selected papers on Noise and Stochastic Processes*. Dover, New York, 1954.
- [Wic95] D. Wick. *The Infamous Boundary*. Birkhäuser, Boston, 1995.
- [Wig67] E. P. Wigner. Random Matrices in Physics. *SIAM Review*, 9(1):1–23, 1967.
- [WR44] E. Whittaker and G. Robinson. *The Calculus of Observations*. Blackie & Son Limited, London and Glasgow, fourth ed. edition, 1944.
- [Wu82] F. Y. Wu. The Potts Model. *Rev. Mod. Phys.*, 54(1):235–268, 1982.
- [WW69] A. T. Whittaker and G. N. Watson. *A Course in Mathematical Analysis*. Cambridge University Press, Cambridge, 1969.
- [WY76] F.Y. Wu and Y.K.Yang. Duality transformations in a many-components spin model. *J. Math. Phys.*, 17(3):439–440, 1976.
- [Xia11] Cheng Xiaxi. *Bell’s theorem and experimental tests*. PhD thesis, Universität Wien, 2011.