



UNIVERSITÀ
DEGLI STUDI
DI PADOVA

Università degli Studi di Padova

DIPARTIMENTO DI MATEMATICA “TULLIO LEVI-CIVITA”

Corso di Laurea Triennale in Matematica

La Legge dei Grandi Numeri:
cinque secoli di storia, risultati e applicazioni

Relatrice:
Prof.ssa Alessandra Bianchi

Laureando: Chiara Fiorentino
Matricola: 1221937

Correlatore:
Prof. Sergio Zoccante

Anno Accademico 2022/2023

data 21/07/2023

*A mio fratello Riccardo
e ai miei genitori*

Indice

Introduzione	7
1 Sviluppo della Legge dei Grandi Numeri	9
1.1 Cardano (1501-1576)	9
1.2 Bernoulli (1654-1705)	10
1.3 Poisson (1781-1840)	14
1.4 Chebyshev (1821-1894)	14
2 Nozioni fondamentali e Legge dei Grandi Numeri	19
2.1 Definizioni e risultati principali	19
2.2 Legge dei Grandi Numeri	24
3 Una legge per variabili aleatorie correlate	31
3.1 Catene di Markov	31
3.2 Teorema ergodico	34
3.3 Metodo Monte Carlo (MCMC)	36
4 Applicazioni	39
4.1 Calcolo di integrali	39
4.2 Algoritmo per l'approssimazione di π	40
4.3 Il problema dell'ago di Buffon	41
4.4 Applicazioni Teoriche	44

Introduzione

L'evoluzione della *Legge dei Grandi Numeri* parte da un'intuizione del matematico italiano Gerolamo Cardano (1505-1575). Egli osservò che le previsioni riguardo ad un evento diventano sempre più precise mano a mano che si ripete l'esperimento. Questa idea rimase però una congettura per circa duecento anni. Il pensiero del tempo impediva infatti lo studio della probabilità: il caso, altro non era che la manifestazione della volontà di Dio e fino ad allora i pochi studi matematici erano limitati ad alcuni giochi d'azzardo.

La prova formale delle osservazioni di Cardano arrivò nel 1713 con la pubblicazione del libro postumo di Jakob Bernoulli (1687-1705) *Ars Conjectandi* e con la dimostrazione del teorema conosciuto come *Teorema di Bernoulli*, che fissò un punto di riferimento per le generazioni successive di matematici. Tra questi si ricordano il francese Simeon Denis Poisson (1781-1840) che coniò il nome di questa legge, Pafnuty Chebyshev (1821-1894) che dimostrò una versione più generale del teorema e Andrej Markov (1856-1922) che generalizzò ulteriormente.

La versatilità di questo teorema rende la *Legge dei Grandi Numeri* applicabile in diversi ambiti quali statistica, economia, teoria della probabilità ed assicurazioni.

L'elaborato si struttura come segue. Nel primo capitolo verrà inquadrato il tema da un punto di vista storico, permettendo di analizzare le cause e gli sviluppi di questa legge. Seguiranno due capitoli in cui vengono fornite le definizioni e i concetti base di probabilità per poi concentrarsi sull'enunciato e sulla dimostrazione della *Legge dei Grandi Numeri* in diverse formulazioni. Infine nell'ultimo capitolo si esploreranno alcune importanti applicazioni tra cui il *metodo Monte Carlo* per il calcolo di integrali e per il calcolo di π , l'*ago di Buffon* e alcuni risultati teorici come il *Teorema di approssimazione di Weierstrass* e il *Teorema di equipartizione asintotica*.

Capitolo 1

Sviluppo della Legge dei Grandi Numeri

In antichità si pensava fosse impossibile trattare con precisione eventi casuali, molto spesso infatti il caso era il frutto di una scelta divina su cui l'uomo non era in grado di fare alcuna previsione. Per tale motivo nessun matematico si era mai azzardato a intraprendere degli studi sul *calcolo delle probabilità*. Solo a partire dal XVI secolo vennero raccolte le prime osservazioni matematiche riguardo al gioco d'azzardo.

A tal proposito questa prima sezione non può non essere dedicata a colui che da molti, è considerato il precursore della probabilità: Gerolamo Cardano.

1.1 Cardano (1501-1576)

Pare a me veramente che non mai si agitasse in uno stesso cervello maggior contrasto fra la ragione e la follia come nel Cardano [...] [7]

Cardano nasce a Pavia nel 1501 ed è il figlio di una giovane donna e di un grande giurista appassionato di Matematica. È un personaggio eclettico e dal carattere tanto brusco quanto geniale: sono diversi i suoi scontri con la Chiesa e con altre figure illustri del tempo, ma altrettanti sono i suoi contributi in Alchimia, Astronomia, Fisica, Matematica e Medicina.

Nonostante la sua laurea in Medicina presso l'Università di Padova è conosciuto per essere il più famoso algebrista del Cinquecento.

Il desiderio di aumentare le vittorie nel gioco dei dadi porta Cardano alla stesura del *Liber de ludo aleae* che riporta alcuni ragionamenti matematici e consigli sul tale gioco. Sebbene sia stato pubblicato un secolo dopo nel 1663 quando queste osservazioni erano ormai state superate, l'opuscolo rappresenta il primo tentativo di sviluppo di una teoria delle probabilità e anticipa seppur senza darne una dimostrazione, la *Legge dei Grandi Numeri*.

Secondo Cardano il gioco d'azzardo si deve basare sull'*equità*: i dadi devono essere onesti, la posta in gioco adeguata e i giocatori rispettosi. Solo se questa situazione è garantita, è possibile trarre osservazioni di indole matematica. Inoltre afferma che assegnare ad un

qualsiasi evento il rapporto $\frac{m}{n}$ tra casi favorevoli (m) e casi possibili (n) non assicura che su n tentativi l'evento si verifica esattamente m volte, ma:

Succede però che in molti circuiti i fatti seguono la congettura da vicino. [3]

Ma quanto vicino?

Nell'ultimo capitolo del fascicolo Cardano suggerisce che aumentando le prove di un qualsiasi esperimento, il valore $\frac{m}{n}$ si stabilizza e diventa sempre più preciso.

1.2 Bernoulli (1654-1705)

Il calcolo delle probabilità diventa un'arte nel 1713 con la pubblicazione postuma di *Ars Conjectandi* di Jakob Bernoulli.

L'opera, che purtroppo risulta incompleta, si divide in quattro capitoli e si conclude con l'intenzione di diffondere l'uso della probabilità in discipline diverse dal gioco d'azzardo.

Artis Conjectandi
Pars Quarta
tradens,
Usum et applicationem praecedentis doctrinae
in civilibus, moralibus et oeconomicis. [1]

Dopo aver dato la definizione di *evento possibile* e di *evento moralmente certo*, Bernoulli introduce il concetto di *probabilità* come il grado di certezza calcolabile tramite il rapporto tra casi favorevoli e numero di casi possibili.

È nel quarto capitolo che il matematico fornisce una dimostrazione rigorosa di quanto accennava Cardano quasi duecento anni prima e che rappresenta la prima formulazione vera e propria della *Legge debole dei Grandi Numeri*.

Il *theorem aureum*, oggi conosciuto come il *Teorema di Bernoulli* è preceduto da alcuni lemmi.

Anche se le notazioni del coefficiente binomiale e del fattoriale non erano ancora state introdotte, in seguito ne verrà fatto uso per poter procedere con maggior compattezza nelle dimostrazioni, che manengono l'idea originale.

Lemma 1. *Siano r, s e $n \in \mathbb{N}$ e si prendano in considerazione le due successioni di numeri*

$$0, 1, 2, \dots, r-1, r, r+1, \dots, r+s;$$

$$0, 1, 2, \dots, n(r-1), \dots, nr, \dots, n(r+1), \dots, n(r+s).$$

Allora tra $n(r+1)$ e $n(r+s)$ c'è un numero di termini non superiore a $s-1$ volte il numero dei termini tra nr e $n(r-1)$ o tra nr e $n(r+1)$. Allo stesso modo tra 0 e $n(r-1)$ c'è un numero di termini non superiore a $r-1$ volte il numero dei termini tra $n(r-1)$ e nr o tra nr e $n(r+1)$.

Dimostrazione. $n(r+s) - n(r+1) = n(s-1)$ mentre $n(r+1) - nr = n = nr - n(r-1)$. Il secondo caso è del tutto analogo. \square

Lemma 2. Nello sviluppo di $(r + s)^n$ compaiono $n + 1$ termini.

Dimostrazione. I termini che compaiono nello sviluppo di $(r + s)^n$ sono del tipo $r^k s^{n-k}$ con $k = 0, \dots, n$ e sono quindi $n + 1$. \square

Lemma 3. Si definisca t come $t = r + s$ e si consideri $(r + s)^{nt}$. Sia poi M il coefficiente corrispondente alla potenza $r^{nr} s^{ns}$. Allora M è il termine massimo dello sviluppo e vale

$$\frac{M}{M_-} < \frac{M_-}{M_{--}}$$

dove M_- è il termine che precede M nello sviluppo, mentre M_{--} è il termine che precede M_- .

Dimostrazione. Ai tempi di Bernoulli era già noto che: $(a + b)^n = \sum_{k=0}^n \binom{n}{k} a^{n-k} b^k$. Di conseguenza:

$$M = \frac{(nt)!}{(ns)!(nt - ns)!} r^{nr} s^{ns} = \frac{(nt)!}{(ns)!(nr)!} r^{nr} s^{ns}$$

A seconda di quale fattoriale semplificare al denominatore, M può essere scritto in due modi equivalenti:

$$M = \begin{cases} \frac{nt(nt-1)\dots(nr+1)}{(ns)!} r^{nr} s^{ns} \\ \frac{nt(nt-1)\dots(ns+1)}{(nr)!} r^{nr} s^{ns} \end{cases}$$

Allora il termine M_- ed il termine M_+ che precedono e seguono M sono:

$$M_- = \frac{nt(nt-1)\dots(nr+2)}{(ns-1)!} r^{nr+1} s^{ns-1}$$

$$M_+ = \frac{nt(nt-1)\dots(ns+2)}{(nr-1)!} r^{nr-1} s^{ns+1}$$

mentre i termini che precedono M_- e seguono M_+ sono:

$$M_{--} = \frac{nt(nt-1)\dots(nr+3)}{(ns-2)!} r^{nr+2} s^{ns-2}$$

$$M_{++} = \frac{nt(nt-1)\dots(ns+3)}{(nr-2)!} r^{nr-2} s^{ns+2}.$$

Quindi con opportune semplificazioni:

$$\frac{M}{M_-} = \frac{(nr+1)s}{nsr} \quad \frac{M_-}{M_{--}} = \frac{(nr+2)s}{(ns-1)r}$$

$$\frac{M}{M_+} = \frac{(ns+1)r}{nrs} \quad \frac{M_+}{M_{++}} = \frac{(ns+2)r}{(nr-1)s}.$$

Segue immediatamente che $\frac{M}{M_-} < \frac{M_-}{M_{--}}$ e $\frac{M}{M_+} < \frac{M_+}{M_{++}}$. Infine riflettendo analogamente sugli altri termini dello sviluppo si deduce che M è il massimo. \square

Lemma 4. Siano ora L e Λ i termini dello sviluppo $(r+s)^{nt}$ posti n posizioni a sinistra e a destra di M rispettivamente. Per n opportunamente grande $\frac{M}{L}$ e $\frac{M}{\Lambda}$ assumono un valore elevato.

Dimostrazione. Sviluppando i binomiali si ottiene che:

$$L = \frac{nt(nt-1)\dots(nr+n+1)}{(ns-n)!} r^{nr+n} s^{ns-n};$$

$$\Lambda = \frac{nt(nt-1)\dots(ns+n+1)}{(nr-n)!} r^{nr-n} s^{ns+n}.$$

Semplificando, si deriva:

$$\frac{M}{L} = \frac{(nr+n)(nr+n-1)\dots(nr+1)}{(ns-n+1)(ns-n+2)\dots(ns)} \frac{s^n}{r^n};$$

$$\frac{M}{\Lambda} = \frac{(ns+n)(ns+n-1)\dots(ns+1)}{(nr-n+1)(nr-n+2)\dots(nr)} \frac{r^n}{s^n}.$$

Distribuendo le potenze n -esime di r ed s all'interno delle parentesi si riscrive:

$$\frac{M}{L} = \frac{(nrs+ns)(nrs+ns-s)\dots(nrs+s)}{(nsr-nr+r)(nsr-nr+2r)\dots(nsr)},$$

$$\frac{M}{\Lambda} = \frac{(nsr+nr)(nsr+nr-r)\dots(nsr+r)}{(nrs-ns+s)(nrs-ns+2s)\dots(nrs)}.$$

Raccogliendo sia a numeratore che a denominatore n e calcolando il limite per $n \rightarrow \infty$ si trova che:

$$\frac{M}{L} \rightarrow \left(\frac{s(r+1)}{r(s-1)} \right)^n \quad \text{e} \quad \frac{M}{\Lambda} \rightarrow \left(\frac{r(s+1)}{s(r-1)} \right)^n$$

Dal momento in cui sia $\frac{s(r+1)}{r(s-1)}$ che $\frac{r(s+1)}{s(r-1)}$ sono maggiori di 1, i rapporti $\frac{M}{L}$ e $\frac{M}{\Lambda}$ diventano arbitrariamente grandi all'aumentare di n . \square

Lemma 5. Il rapporto tra la somma degli elementi inclusi tra L e Λ e la somma dei restanti termini dello sviluppo è più grande di qualsiasi quantità, per un determinato n .

Dimostrazione. Siano $L_1, L_2, \dots, L_n = L$ i termini compresi tra L ed M dove con L_i si intende il termine posto a distanza i da M e si chiamino $P_1, P_2, P_3 \dots$ i predecessori di L , con P_1 il termine che precede L . Dal Lemma 3 e in particolare dalla sua dimostrazione si deduce che:

$$\dots < P_3 < P_2 < P_1 < L_n < \dots < L_2 < L_1 < M$$

$$\frac{M}{L_1} < \frac{L_n}{P_1}, \quad \frac{L_1}{L_2} < \frac{P_1}{P_2}, \quad \frac{L_2}{L_3} < \frac{P_2}{P_3}, \quad \dots$$

Allora:

$$\frac{M}{L_n} < \frac{L_1}{P_1} < \frac{L_2}{P_2} < \frac{L_3}{P_3} < \dots$$

Per il Lemma 4, $\frac{M}{L^n}$ assume un valore arbitrariamente grande; quindi è possibile trovare c tale che $\frac{M}{L^n} \geq c(s-1)$. Di conseguenza questa disuguaglianza vale per tutti i rapporti considerati:

$$L_1 > c(s-1)P_1, \quad L_2 > c(s-1)P_2, \quad L_3 > c(s-1)P_3, \quad \dots$$

Le difficoltà si incontrano dal momento in cui il numero dei predecessori di L e il numero di termini tra L ed M sono in generale diversi: dal Lemma 1 però si rivava che il primo è $r-1$ volte il secondo. Quindi se k è il numero dei predecessori di L :

$$\frac{\sum_{i=1}^{n-1} L_i}{\sum_{i=1}^k P_i} > \frac{\sum_{i=1}^{n-1} L_i}{(r-1) \sum_{i=1}^{n-1} P_i}$$

che per le disuguaglianze viste prima si riduce a:

$$\frac{\sum_{i=1}^{n-1} L_i}{\sum_{i=1}^k P_i} > \frac{\sum_{i=1}^{n-1} L_i}{(r-1) \sum_{i=1}^{n-1} P_i} > \frac{c(s-1)}{r-1}.$$

Dunque il lemma è verificato per i predecessori di L . Conti analoghi verificano la disuguaglianza per i membri a destra di Λ . \square

Convinto che a molti poteva turbare l'utilizzo dell'infinito, Bernoulli fornisce una dimostrazione alternativa del lemma che fa uso principalmente delle proprietà del logaritmo e che sicuramente è più facile da accettare.

A questo punto Bernoulli può enunciare il teorema e fornirne una dimostrazione in cui vengono applicati i lemmi precedenti.

Teorema. *Si supponga che il rapporto tra i casi favorevoli (fertiles) e sfavorevoli (steriles) sia esattamente o approssimativamente quello tra r ed s . Allora è possibile ripetere un esperimento un numero di volte tale che sia più probabile che il rapporto $\frac{r}{t}$ sia compreso tra $\frac{r-1}{t}$ e $\frac{r+1}{t}$ rispetto ad un dato numero di osservazioni c .*

Dimostrazione. Sia nt il numero di osservazioni con $t = r + s$. In ciascuna prova ci sono t casi di cui r fertili ed s sterili, di conseguenza la probabilità che non ci siano casi sfavorevoli è $(\frac{r}{t})^{nt}$, quella che ce ne sia solo uno è $\binom{nt}{1} \frac{r^{nt-1}s}{t^{nt}}$, quella che ce ne siano due è $\binom{nt}{2} \frac{r^{nt-2}s^2}{t^{nt}}$, etc...

È possibile allora interpretare M come il caso in cui ci siano nr successi e ns insuccessi, mentre L e Λ come i casi in cui gli eventi favorevoli siano rispettivamente $nr+n$ e $nr-r$. Segue allora che il numero di casi in cui gli eventi fertili sono non minori di $nr-n$ e non maggiori di $nr+n$ corrisponde all'insieme dei termini compresi tra L e Λ . Per il Lemma 4 e il Lemma 5, è possibile determinare n tale che la somma dei termini compresi tra L e Λ sia più grande di c volte quella dei predecessori di L e dei successori di Λ , quindi è possibile considerare un n per cui sia c volte più verosimile che $\frac{r}{t}$ sia interno a $[\frac{nr-n}{nt}, \frac{nr+n}{nt}]$ cioè interno a $[\frac{r-1}{t}, \frac{r+1}{t}]$. \square

Quindi l'intuizione di Cardano non era per nulla sbagliata, infatti più il numero di osservazioni aumenta, minore è il rischio di allontanarsi dalla realtà.

Sicuramente ispirato dalle sue origini da teologo, Bernoulli decide di concludere il capitolo con una stravolgente osservazione filosofica: se si potesse avere a disposizione l'intera eternità per poter osservare un qualsiasi evento, si potrebbe dimostrare che tutto nel mondo, anche ciò che sembra più casuale, segue una legge precisa che lo rende necessario.

1.3 Poisson (1781-1840)

Chiamare *teorema* un'osservazione così importante, sembrava non rendere giustizia all'universalità che questa proponeva. Per questo motivo, il matematico francese Poisson introduce nel 1837 il termine *Legge dei Grandi Numeri*. La legge riscuote diverse critiche e risulta davvero difficile per alcuni accettarla: Bienaymé, statistico e connazionale di Poisson addirittura ne negò l'esistenza.

La versione di Poisson vuole essere un'estensione del teorema di Bernoulli in cui gli eventi considerati possono dipendere anche da cause per cui la probabilità varia in ogni prova. A tal proposito si considerino due eventi complementari E ed F , si eseguano μ ripetizioni di un esperimento aleatorio e si supponga che le probabilità che si verifichino E ed F rispettivamente siano p_1 e q_1 per la prima prova, p_2 e q_2 per la seconda e così via. Se E si verifica m volte ed F si verifica $n = \mu - m$ volte, allora quando il numero di prove va all'infinito, le quantità $\frac{m}{\mu}$ e $\frac{n}{\mu}$ tendono rispettivamente a

$$p' = \frac{\sum_{i=1}^{\mu} p_i}{\mu} \quad q' = \frac{\sum_{i=1}^{\mu} q_i}{\mu}.$$

L'evento E , e quindi anche F , può essere condizionato dalle cause C_1, \dots, C_ν di probabilità $\gamma_1, \dots, \gamma_\nu$, per cui è possibile scrivere

$$P(E|C_i) = c_i$$

In questo caso il rapporto $\frac{m}{\mu}$, per $\mu \rightarrow \infty$ tende a $\sum_{i=1}^{\nu} \gamma_i c_i$. Si deduce quindi che se si eseguissero μ^* prove e se il numero di volte in cui si presenta E è pari a m^* vengono approssimativamente soddisfatte le relazioni:

$$\frac{m}{\mu} = \frac{m^*}{\mu^*}.$$

È questa l'essenza della sua *Legge dei Grandi Numeri*.

Accanto a questa formulazione, Poisson presenta una serie di esempi che hanno lo scopo di spiegare meglio ciò che intendeva e di fornire delle dimostrazioni empiriche del suo risultato. Un esempio concreto è dato dal rapporto costante di nascite maschili sul numero di nascite totali osservato nel territorio francese in quegli anni.

1.4 Chebyshev (1821-1894)

Secondo il matematico russo Pafnuty Lvovich Chebyshev la dimostrazione della *Legge dei Grandi Numeri* proposta da Poisson scarseggiava di precisione e di rigore e per tale

motivo, nel 1864, propone una versione differente in cui fa uso della disuguaglianza, oggi conosciuta come *Disuguaglianza di Chebyshev*, ma che in realtà era già stata dimostrata nel 1853 da Bienaymé.

La dimostrazione della *Disuguaglianza*, così come la *Legge dei Grandi Numeri* nella formulazione di Chebyshev, verrà trattata nel prossimo capitolo utilizzando un linguaggio moderno, ma dal momento che il matematico non parla nè di *variabili aleatorie* nè di *indipendenza*, si ritiene opportuno inserire in questa sezione la dimostrazione originale.

Teorema. Siano a, b, c, \dots le speranze matematiche delle quantità x, y, z, \dots e a_1, b_1, c_1, \dots le speranze matematiche di x^2, y^2, z^2, \dots . Definite L ed M come:

$$L = a + b + c + \dots - \alpha \sqrt{a_1 + b_1 + c_1 + \dots - a^2 - b^2 - c^2 - \dots}$$

$$M = a + b + c + \dots + \alpha \sqrt{a_1 + b_1 + c_1 + \dots - a^2 - b^2 - c^2 - \dots}$$

allora:

$$P(L \leq x + y + z + \dots \leq M) > 1 - \frac{1}{\alpha^2} \quad \forall \alpha > 0.$$

Si osservi che nell'enunciato, Chebyshev non specifica se il numero delle quantità x, y, z, \dots è finito o numerabile e assume, ma senza scriverlo esplicitamente, che tutte le speranze matematiche considerate siano finite.

Dimostrazione. Siano $x = (x_1, \dots, x_l)$, $y = (y_1, \dots, y_m)$, $z = (z_1, \dots, z_n), \dots$ i valori che possono essere assunti dalle quantità x, y, z, \dots rispettivamente e (p_1, \dots, p_l) , (q_1, \dots, q_m) , $(r_1, \dots, r_n), \dots$ le corrispondenti probabilità. Con queste notazioni:

$$a = \sum_{\lambda=1}^l p_\lambda x_\lambda \quad b = \sum_{\mu=1}^m q_\mu y_\mu \quad c = \sum_{\nu=1}^n r_\nu z_\nu \quad \dots$$

$$a_1 = \sum_{\lambda=1}^l p_\lambda x_\lambda^2 \quad b_1 = \sum_{\mu=1}^m q_\mu y_\mu^2 \quad c_1 = \sum_{\nu=1}^n r_\nu z_\nu^2 \quad \dots$$

Si ponga poi:

$$\begin{aligned} C_{\lambda\mu\nu\dots} &= [x_\lambda + y_\mu + z_\nu + \dots - (a + b + c + \dots)]^2 \\ &= (x_\lambda^2 + y_\mu^2 + z_\nu^2 + \dots) + 2(x_\lambda y_\mu + x_\lambda z_\nu + y_\mu z_\nu + \dots) \\ &\quad - 2(a + b + c + \dots)(x_\lambda + y_\mu + z_\nu) + (a + b + c + \dots)^2 \end{aligned}$$

e si definisca

$$S := \sum_{\lambda, \mu, \nu, \dots} C_{\lambda\mu\nu\dots} p_\lambda q_\mu r_\nu \dots$$

Ricordando che $\sum_{\lambda}^l p_\lambda = 1$ e sommando su λ i singoli addendi di S , si ottengono le seguenti uguaglianze:

$$\begin{aligned} \sum_{\lambda} (x_\lambda^2 + y_\mu^2 + z_\nu^2 + \dots) p_\lambda q_\mu r_\nu \dots &= p_\lambda x_\lambda^2 q_\mu r_\nu + \sum_{\lambda} p_\lambda (y_\mu^2 + z_\nu^2 + \dots) q_\mu r_\nu \dots \\ &= a_1 q_\mu r_\nu + (y_\mu^2 + z_\nu^2 + \dots) q_\mu r_\nu \dots \end{aligned}$$

$$2 \sum_{\lambda} (x_{\lambda} y_{\mu} + x_{\lambda} z_{\nu} + y_{\mu} z_{\nu} + \dots) p_{\lambda} q_{\mu} r_{\nu} \dots = 2 q_{\mu} r_{\nu} \dots [a(y_{\mu} + z_{\nu} + \dots) + y_{\mu} z_{\nu} + \dots]$$

$$\begin{aligned} & -2 \sum_{\lambda} (a + b + c + \dots) (x_{\lambda} + y_{\mu} + z_{\nu} + \dots) p_{\lambda} q_{\mu} r_{\nu} \dots = \\ & -2(a + b + c + \dots) q_{\mu} r_{\nu} \dots (a + y_{\mu} + z_{\nu} + \dots) \end{aligned}$$

$$(a + b + c + \dots)^2 \sum_{\lambda} p_{\lambda} q_{\mu} r_{\nu} \dots = (a + b + c + \dots)^2 q_{\mu} r_{\nu} \dots$$

Su queste si svolgono le somme su μ e ν e si ottengono rispettivamente:

$$a_1 + b_1 + c_1 + \dots$$

$$2(ab + ac + bc + \dots)$$

$$-2(a + b + c + \dots)^2$$

$$(a + b + c + \dots)^2$$

Di conseguenza $S = a_1 + b_1 + c_1 + \dots - (a^2 + b^2 + c^2 + \dots)$ e quindi per ogni $\alpha > 0$:

$$\frac{S}{\alpha^2 [a_1 + b_1 + c_1 + \dots - (a^2 + b^2 + c^2 + \dots)]} = \frac{1}{\alpha^2}.$$

A questo punto Chebyshev osserva che sostituendo gli addendi per cui

$$\frac{C_{\lambda\mu\nu}}{\alpha^2 [a_1 + b_1 + c_1 + \dots - (a^2 + b^2 + c^2 + \dots)]} < 1$$

con 0 e con 1 quelli per cui

$$(*) \frac{C_{\lambda\mu\nu}}{\alpha^2 [a_1 + b_1 + c_1 + \dots - (a^2 + b^2 + c^2 + \dots)]} > 1$$

il membro a sinistra dell'equazione sopra diminuisce. Quindi considerando i termini per cui vale *

$$\sum_{\lambda\mu\nu} p_{\lambda} q_{\mu} r_{\nu} \dots < \frac{1}{\alpha^2}$$

Riscrivendo opportunamente $C_{\lambda\mu\nu}$ si deduce che la sommatoria $\sum_{\lambda\mu\nu} p_{\lambda} q_{\mu} r_{\nu} \dots$ rappresenta la probabilità che $x + y + z + \dots$ sia fuori dall'intervallo $[L, M]$. Quindi si conclude che

$$P(L \leq x + y + z + \dots \leq M) > 1 - \frac{1}{\alpha^2}.$$

□

A questo punto Chebyshev considera il caso in cui le quantità x, y, z, \dots siano in numero pari ad N e ne calcola la media aritmetica $E = \frac{x+y+z+\dots}{N}$. Poi fissa $\alpha = \frac{\sqrt{N}}{t}$ per un certo $t > 0$. Con l'ipotesi che a, b, c, \dots e $a_1, b_1, c_1 \dots$ siano limitati, applica il teorema e ne deduce che per N che tende all'infinito la probabilità $P\left(\frac{L}{N} \leq E \leq \frac{M}{N}\right)$ tende a 1.

Nel caso specifico in cui le quantità x, y, z, \dots valgono solo 0 o 1, la media E può essere interpretata come la frequenza di successo di un evento in N prove ripetute. È proprio tramite questi passaggi che il matematico mostra la *Legge dei Grandi Numeri*: aumentando il numero delle prove, la probabilità si avvicina all'unità.

Capitolo 2

Nozioni fondamentali e Legge dei Grandi Numeri

Se nel primo capitolo la *Legge dei Grandi Numeri* è stata inquadrata storicamente, in questa seconda parte verranno date alcune definizioni e mostrati alcuni risultati che saranno fondamentali per poter proseguire con l'analisi.

Oltre alla definizione di *spazio di probabilità*, ci si soffermerà sulla nozione di *variabile aleatoria* che suggerisce l'idea di una quantità data dal risultato di un esperimento casuale.

Infine verrà enunciato il teorema riguardante la *Legge dei Grandi Numeri* e verranno proposte varie dimostrazioni che variano in base alle ipotesi fissate.

2.1 Definizioni e risultati principali

σ -algebre e misure di probabilità

Definizione. Sia Ω un insieme non vuoto e $\mathcal{P}(\Omega)$ il suo insieme delle parti. Si definisce σ -algebra una famiglia $\mathcal{A} \subseteq \mathcal{P}(\Omega)$ tale che:

1. $\emptyset \in \mathcal{A}$;
2. se $A \in \mathcal{A}$ allora $A^c \in \mathcal{A}$;
3. se $(A_n)_{n \in \mathbb{N}}$ è una successione numerabile del elementi di \mathcal{A} , allora anche $\bigcup_{n \in \mathbb{N}} A_n$ è un elemento di \mathcal{A} .

Definizione. Gli elementi di \mathcal{A} vengono chiamati *sottoinsiemi misurabili* di Ω , mentre una coppia (Ω, \mathcal{A}) composta da un insieme Ω non vuoto e da una σ -algebra \mathcal{A} si dice *spazio misurabile*.

Si osserva che se \mathcal{A} è una σ -algebra su X dalle proprietà 1. e 2. segue immediatamente che $\Omega \in \mathcal{A}$. Inoltre la 3. suggerisce che \mathcal{A} è anche chiusa per unioni finite di elementi di \mathcal{A} , infatti presa una famiglia finita (A_1, \dots, A_n) di elementi di \mathcal{A} è possibile completarla ad una successione numerabile ponendo $A_k = \emptyset$ per $k > n$. Questo passaggio è lecito in

quanto l'unione non viene modificata.

Dall'identità di de Morgan $\bigcap_n A_n = (\bigcup_n A_n^c)^c$ si deduce infine che una σ -algebra è chiusa per intersezioni.

Definizione. Sia (Ω, \mathcal{A}) uno spazio misurabile. Una funzione $P: \mathcal{A} \rightarrow [0, 1]$ si dice *misura di probabilità* se:

1. $P(A) \geq P(\emptyset) = 0$ per ogni $A \in \mathcal{A}$,
2. $P(\mathcal{A})=1$
3. vale la σ -additività: per ogni successione numerabile $(A_n)_{n \in \mathbb{N}}$ di elementi disgiunti di \mathcal{A} si ha

$$P\left(\bigcup_{n \in \mathbb{N}} A_n\right) = \sum_{n \in \mathbb{N}} P(A_n),$$

Definizione. Uno *spazio di probabilità* è una tripletta (Ω, \mathcal{A}, P) in cui Ω è un insieme non vuoto, \mathcal{A} una σ -algebra e P una misura di probabilità.

L'insieme Ω viene chiamato *spazio campionario* mentre gli elementi di \mathcal{A} *eventi*.

Chiaramente in uno spazio di probabilità (Ω, \mathcal{A}, P) vale l'additività finita: dati N eventi disgiunti A_1, \dots, A_N allora:

$$P\left(\bigcup_{n=1}^N A_n\right) = \sum_{n=1}^N P(A_n).$$

Proprio come è stato fatto per mostrare la chiusura per unioni finite di elementi di una σ -algebra è possibile prolungare A_1, \dots, A_N ad una successione infinita di eventi disgiunti ponendo $A_n = \emptyset$ per $n > N$. Allora:

$$P\left(\bigcup_{n=1}^N A_n\right) = P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n) = \sum_{n=1}^N P(A_n)$$

dove la seconda uguaglianza deriva dalla σ -additività, mentre l'ultima dal fatto che $P(\emptyset) = 0$.

Nel seguente teorema vengono racchiuse alcune conseguenze della definizione di probabilità.

Teorema. Sia (Ω, \mathcal{A}, P) uno spazio di probabilità. Valgono le seguenti proprietà:

1. *monotonia:* se $A \subseteq B$ allora $P(A) \leq P(B)$;
2. *subadditività:* se $A \subseteq \bigcup_{n=1}^{\infty} A_n$ allora $P(A) \leq \sum_{n=1}^{\infty} P(A_n)$;
3. *continuità dal basso:* Se $A_n \uparrow A$ allora $P(A_i) \uparrow P(A)$;
4. *continuità dall'alto:* Se $A_n \downarrow A$, con $P(A_1) < \infty$ allora $P(A_i) \downarrow P(A)$.

Variabili aleatorie generali

Si considerino ora uno spazio di probabilità (Ω, \mathcal{A}, P) e uno spazio misurabile (E, \mathcal{E}) .

Definizione. Si definisce *variabile aleatoria*, una funzione $X: \Omega \rightarrow E$ tale per cui, se $C \in \mathcal{E}$ allora $\{X \in C\} := X^{-1}(C) \in \mathcal{A}$.

Le variabili aleatorie $X: \Omega \rightarrow \mathbb{R}^n$ vengono chiamate *vettori aleatori di dimensione n* o *variabili vettoriali*.

OSSERVAZIONE. La definizione di *variabile aleatoria a valori in E* dipende dalla scelta della σ -algebra \mathcal{E} su E.

Potendo attribuire ad una variabile aleatoria X diversi valori è necessario poter associare a X un'applicazione che attribuisca ad ogni risultato dell'esperimento la sua probabilità.

Definizione. Data una variabile aleatoria X definita sullo spazio (Ω, \mathcal{A}, P) e a valori in (E, \mathcal{E}) , si definisce *distribuzione* o *legge di X* l'applicazione

$$\begin{aligned} \mu_X: \mathcal{E} &\rightarrow [0, 1] \\ C &\mapsto P(X \in C). \end{aligned}$$

Si consideri d'ora in poi il caso in cui $E = \mathbb{R}^n$ e si prenda come \mathcal{E} la σ -algebra di Borel $\mathcal{B}(\mathbb{R}^n)$, generata dai sottoinsiemi aperti di \mathbb{R}^n .

Può succedere che se X è un vettore aleatorio di dimensione n e $g: \mathbb{R}^n \rightarrow \mathbb{R}^m$ è una funzione generica, la funzione $g(X): \Omega \rightarrow \mathbb{R}^m$ non risulta una variabile aleatoria. Perché questo sia garantito si passa alla definizione di *misurabilità*.

Definizione. Una funzione $g: \mathbb{R}^n \rightarrow \mathbb{R}^m$ si dice *misurabile* se per ogni $C \in \mathcal{B}(\mathbb{R}^m)$ si ha che $g^{-1}(C) \in \mathcal{B}(\mathbb{R}^n)$.

Per una maggiore comprensione e per sottolinearne l'importanza, conviene riassumere quanto appena visto nel seguente Teorema ed esplicitare nel Corollario successivo un risultato che verrà ampiamente utilizzato nelle prossime sezioni.

Teorema. Se X_1, \dots, X_n sono vettori aleatori di dimensione n e se $f: \mathbb{R}^n \rightarrow \mathbb{R}$ è una funzione misurabile, allora $f(X_1, \dots, X_n)$ è una variabile aleatoria.

Corollario. Se X, X_1, \dots, X_n sono variabili vettoriali e $a \in \mathbb{R}$ allora $X_1 + \dots + X_n$ e aX sono variabili aleatorie.

Indipendenza e valor medio

Durante un esperimento è possibile considerare diverse variabili aleatorie contemporaneamente: ad esempio lanciando ripetutamente un dado si può essere interessati al numero di volte X in cui esce il numero 1, ma anche alla variabile Y che corrisponde alla somma dei primi due lanci. In questi casi è normale chiedersi se la distribuzione di una variabile aleatoria influisca o meno sulla legge delle restanti variabili considerate e a tal proposito fissare la definizione di *indipendenza di variabili aleatorie* risulta necessario.

Definizione. Due variabili aleatorie X e Y definite sullo stesso spazio di probabilità (Ω, \mathcal{A}, P) e a valori in (E_X, \mathcal{E}_X) , (E_Y, \mathcal{E}_Y) rispettivamente, si dicono *indipendenti* se per ogni $A \in \mathcal{E}_X$ e per ogni $B \in \mathcal{E}_Y$ si ha:

$$\mu_{X,Y}(A, B) = P(X \in A, Y \in B) = P(X \in A)P(Y \in B) = \mu_X(A)\mu_Y(B).$$

Il significato di *indipendenza* risulta intuitivo: per due variabili aleatorie indipendenti X e Y , avere o meno informazioni su Y non modifica la probabilità di un esito possibile di X e quindi la distribuzione di X non è influenzata in alcun modo dalla distribuzione di Y .

Chiaramente la definizione può essere estesa ad una famiglia finita $(X_i)_{i \in I}$ di variabili aleatorie, mentre nel caso in cui l'insieme I sia numerabile, le $(X_i)_{i \in I}$ si dicono indipendenti se ogni sottofamiglia finita è costituita da variabili aleatorie indipendenti.

Anche in questi casi apparentemente più complicati l'indipendenza di più variabili aleatorie formalizza la stessa idea per cui avere informazioni su alcune variabili non influisce in alcun modo sull'esito delle restanti.

Si introduce ora un concetto legato ad una variabile aleatoria che è largamente utilizzato in probabilità: il *valor medio*, detto anche *valore atteso*.

Definizione. Sia X una variabile aleatoria reale definita su uno spazio di probabilità (Ω, \mathcal{A}, P) e $X^+ := \max(X, 0)$, $X^- := -\min(X, 0)$ le due variabili aleatorie positive corrispondenti alla parte positiva e la parte negativa.

- Se $X \geq 0$ si definisce *valor medio di X* la quantità $E(X) := \int_{\mathbb{R}} X dP$.
- In generale si dice che X *ammette valor medio* se almeno una tra $E(X^+)$ e $E(X^-)$ è finita e in tal caso si scrive $E(X) := E(X^+) - E(X^-)$.

Nel caso discreto la formula del valore atteso equivale a $\sum_{x \in X(\Omega)} x \mu_X(x)$. Questa scrittura rende più evidente il significato di $E(X)$: vedendo le probabilità $\mu_X(x)$ come dei pesi, il valor medio $E(X)$ lo si può interpretare come il baricentro dei possibili esiti x di X , infatti $\sum_{x \in X(\Omega)} \mu_X(x) = 1$.¹

OSSERVAZIONE. Siano X_1, \dots, X_n n variabili aleatorie indipendenti con la stessa distribuzione (i.i.d.) che rappresentano n esiti di uno stesso esperimento aleatorio e si consideri la loro media aritmetica $\bar{X}_n := \frac{1}{n}(X_1, \dots, X_n)$. Poichè le funzioni di somma e di divisione per un numero n sono misurabili, \bar{X}_n è una variabile aleatoria ed è possibile calcolarne il valor medio. La *Legge dei Grandi Numeri*, come verrà approfondito nella prossima sezione, afferma che con probabilità elevata e per n sufficientemente grande $E(\bar{X}_n)$ assume valori vicini a \bar{X}_n : la legge allora suggerisce che quando esiste, il valore medio può essere interpretato come la media aritmetica dei risultati ottenuti.

¹La formula per calcolare il baricentro di un insieme finito di numeri reali (x_1, \dots, x_n) di pesi (m_1, \dots, m_n) rispettivamente è: $\frac{x_1 m_1 + \dots + x_n m_n}{m_1 + \dots + m_n}$.

Definizione. Sia $p \in (0, \infty)$. Si definisce l'insieme delle variabili aleatorie X di norma p finita l'insieme

$$\begin{aligned} L^p(\Omega, P) &:= \{X : \Omega \rightarrow \mathbb{R} \text{ tali che } \|X\|_p < +\infty\} \\ &= \{X : \Omega \rightarrow \mathbb{R} \text{ tali che } E(|X|^p) < +\infty\} \end{aligned}$$

dove $\|X\|_p := [E(|X|^p)]^{1/p}$.

Si osservi che $|X|^p$ è una variabile aleatoria positiva quindi il suo valor medio è sempre ben definito ed eventualmente vale $+\infty$.

Dalle proprietà degli integrali discendono le seguenti proprietà del valor medio.

Teorema. Siano X e Y due variabili aleatorie definite sullo stesso spazio di probabilità (Ω, \mathcal{A}, P) positive o tali che $X, Y \in L^1(\Omega, P)$. Allora:

1. *monotonia:* Se $X \geq Y$ allora $E(X) \geq E(Y)$;
2. $E(X + Y) = E(X) + E(Y)$;
3. $E(aX + b) = aE(X) + b$.

Definizione. Se $X \in L^2(\Omega, P)$, viene definita la *varianza di X* come la quantità

$$\text{Var}(X) := E((X - E(X))^2) = E(X^2) - E(X)^2.$$

Nel seguito vengono enunciati alcuni risultati fondamentali che legano le definizioni di indipendenza e di valor medio e che mostrano alcune proprietà della varianza.

Teorema. Siano X_1, \dots, X_n variabili aleatorie indipendenti definite nello stesso spazio di probabilità (Ω, \mathcal{A}, P) . Se $X_i \geq 0$ oppure $X_i \in L^1(\Omega, P) \forall i$ allora

$$E\left(\prod_{i=1}^n X_i\right) = \prod_{i=1}^n E(X_i).$$

In alcune situazioni accade che $E(XY) = E(X)E(Y)$ anche senza supporre che X e Y siano indipendenti. In questi casi si parla di *scorrelazione*.

Definizione. Una famiglia $(X_i)_{i \in \{1, \dots, n\}}$ di variabili aleatorie definite su (Ω, \mathcal{A}, P) è detta *scorrelata* se $X_i \in L^2(\Omega, P)$ e $E(X_i X_j) = E(X_i)E(X_j) \forall i \neq j$.

Teorema. Siano X_1, \dots, X_n n variabili aleatorie scorrelate e tali che $E(X_i^2) < +\infty$. Allora:

$$\text{Var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \text{Var}(X_i).$$

Teorema. Se X è una variabile aleatoria allora $\text{Var}(aX) = a^2 \text{Var}(X)$.

Prima di procedere con l'enunciato e la dimostrazione della *Legge dei Grandi Numeri* è importante mostrare un'importante disuguaglianza che mostra come i valori di una variabile aleatoria X si "disperdono" attorno al valor medio.

Teorema (Disuguaglianza di Chebyshev). *Si consideri una variabile aleatoria $X \in L^2(\Omega, P)$ definita sullo spazio di probabilità (Ω, \mathcal{A}, P) . Allora $\forall \varepsilon > 0$:*

$$P(|X - E(X)| \geq \varepsilon) \leq \frac{\text{Var}(X)}{\varepsilon^2}.$$

La dimostrazione di questo teorema utilizza un'altra importante disuguaglianza, quella di Markov.

Lemma (Disuguaglianza di Markov). *Se X è una variabile aleatoria reale e positiva su (Ω, \mathcal{A}, P) , allora $\forall \varepsilon > 0$:*

$$P(X \geq \varepsilon) \leq \frac{E(x)}{\varepsilon}.$$

Dimostrazione (Lemma). Si fissi ε e si definisca la variabile

$$\mathbb{1}_{\{X \geq \varepsilon\}} := \begin{cases} 1 & \text{se } X \geq \varepsilon \\ 0 & \text{altrimenti} \end{cases}$$

Dalla condizione $X \geq 0$ si deduce che $X \geq \varepsilon \mathbb{1}_{\{X \geq \varepsilon\}}$, mentre dalla monotonia e dalla linearità del valor medio si ricava che:

$$E(X) \geq E(\varepsilon \mathbb{1}_{\{X \geq \varepsilon\}}) = \varepsilon E(\mathbb{1}_{\{X \geq \varepsilon\}}) = \varepsilon P(\mathbb{1}_{\{X \geq \varepsilon\}})$$

da cui segue la tesi. □

Dimostrazione (Teorema). Siccome la variabile aleatoria X non è in generale positiva, non è possibile applicare direttamente la disuguaglianza di Markov. Si osserva però che gli insiemi $\{|X - E(X)| \geq \varepsilon\}$ e $\{|X - E(X)|^2 \geq \varepsilon^2\}$ sono equivalenti: infatti preso $\omega \in \Omega$ questo appartiene alla prima famiglia di eventi se e solo se ω appartiene alla seconda. A questo punto, è possibile applicare il Lemma alla variabile $(X - E(X))^2$ perchè questa è sempre positiva:

$$P(|X - E(X)| \geq \varepsilon) = P((X - E(X))^2 \geq \varepsilon^2) \leq E((X - E(X))^2) = \frac{\text{Var}(X)}{\varepsilon^2}.$$

Quindi la tesi è dimostrata. □

2.2 Legge dei Grandi Numeri

Si supponga di dover ripetere uno stesso esperimento fisico n volte e si denoti con X_i il risultato ottenuto all' i -esima prova. A causa di imprecisioni degli strumenti e/o di fenomeni esterni, l'esperienza è da considerarsi aleatoria.

Per come è stato posto il problema è naturale supporre che le variabili $(X_i)_{i \in \{1, \dots, n\}}$ siano

indipendenti e con la stessa distribuzione, ma si tralasci momentaneamente questa ipotesi. Quello che si spera è che per n molto grande, la media campionaria

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$$

sia più precisa delle singole X_i , ovvero che tenda a oscillare meno e che si concentri attorno ad un valore specifico.

Ma il modello matematico che si sta costruendo rispecchia tali aspettative? Per poter rispondere correttamente a questa domanda e per poter presentare in maniera più rigorosa la *Legge dei Grandi Numeri* è opportuno capire cosa si intende per *convergenza* in ambito probabilistico.

Convergenza in probabilità

Definizione. Siano X, X_1, \dots, X_n n variabili aleatorie definite sullo spazio di probabilità (Ω, \mathcal{A}, P) . Si dice che la successione $(X_i)_{i \in \{1, \dots, n\}}$ converge a X in probabilità se:

$$\forall \varepsilon > 0 \quad P(|X_n - X| > \varepsilon) \rightarrow 0 \text{ per } n \rightarrow +\infty.$$

Lemma (Proprietà della convergenza in probabilità). *Siano $(X_n)_{n \in \mathbb{N}}$ e $(Y_n)_{n \in \mathbb{N}}$ due successioni di variabili aleatorie reali e sia $(a_n)_{n \in \mathbb{N}} \in \mathbb{R}$. Allora:*

1. *se $(X_n)_n$ e $(Y_n)_n$ convergono a 0 in probabilità allora $(X_n + Y_n)_n$ converge a 0 in probabilità;*
2. *se $(X_n)_n$ converge a 0 in probabilità e $(a_n)_n$ è limitata allora anche $(a_n X_n)_n$ converge a 0 in probabilità.*

Dimostrazione. Sia $\varepsilon > 0$. Per le proprietà della misura P si ricava che

$$P(|X_n + Y_n| > \varepsilon) \leq P\left(|X_n| > \frac{\varepsilon}{2}\right) + P\left(|Y_n| > \frac{\varepsilon}{2}\right).$$

Quindi $(X_n + Y_n)_n$ converge a 0 in probabilità.

Se poi $|a_n| \leq c$ segue che

$$P(|a_n X_n| > \varepsilon) \leq P(c|X_n| > \varepsilon) = P\left(|X_n| > \frac{\varepsilon}{c}\right)$$

Quindi anche $(a_n X_n)_n$ converge a 0 in probabilità. □

Legge Debole dei Grandi Numeri

La domanda posta in precedenza ha come risposta la *Legge debole dei Grandi Numeri* che ha diverse formulazioni in base alle ipotesi che si decidono di fissare.

L'assunzione più comune, come si è già accennato, è quella di considerare una successione di variabili aleatorie indipendenti e identicamente distribuite.

Teorema (Legge debole dei Grandi Numeri in L^1). Sia $(X_i)_{i \in \mathbb{N}}$ una successione di variabili aleatorie reali, indipendenti e identicamente distribuite sullo spazio di probabilità (Ω, \mathcal{A}, P) . Se esiste finita la media $E(X_i) = \mu$, ovvero se $(X_i) \in L^1(\Omega, P) \forall i \in \mathbb{N}$, allora:

$$\forall \varepsilon > 0 \quad P(|\bar{X}_n - \mu| > \varepsilon) \rightarrow 0 \text{ per } n \rightarrow +\infty.$$

Un'altra versione della *Legge* considera variabili aleatorie in L^2 . Questo secondo caso richiede ipotesi più deboli: le variabili aleatorie non devono essere indipendenti e identicamente distribuite, bensì scorrelate e con lo stesso valor medio e con la stessa varianza.

La sua dimostrazione procede utilizzando le due importanti proprietà della varianza di variabili aleatorie viste sopra.

Teorema (Legge debole dei Grandi Numeri in L^2). Siano X_1, \dots, X_n delle variabili aleatorie reali in $L^2(\Omega, P)$, scorrelate e tali che $E(X_i) = \mu \forall i \in \{1, \dots, n\}$. Se $Var(X_i) = \sigma^2 \forall i \in \{1, \dots, n\}$ allora \bar{X}_n converge a μ in probabilità.

Dimostrazione. Siccome le variabili aleatorie sono scorrelate, per i teoremi visti in precedenza valgono le seguenti uguaglianze:

$$\begin{aligned} Var(\bar{X}_n) &= Var\left(\frac{X_1 + \dots + X_n}{n}\right) \\ &= \frac{1}{n^2} Var(X_1 + \dots + X_n) \\ &= \frac{1}{n^2} (Var(X_1) + \dots + Var(X_n)) \\ &= \frac{1}{n^2} (\sigma^2 + \dots + \sigma^2) \\ &= \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}. \end{aligned}$$

Applicando la disuguaglianza di Chebyshev a \bar{X}_n si ottiene che $\forall \varepsilon > 0$

$$P(|\bar{X}_n - \mu| \geq \varepsilon) \leq \frac{Var(\bar{X}_n)}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2} \rightarrow 0 \text{ per } n \rightarrow +\infty.$$

□

OSSERVAZIONE. La richiesta che le (X_i) abbiano la stessa varianza può essere indebolita ipotizzando che $\exists c$ tale che $Var(X_i) \geq c \forall i \in \{1, \dots, n\}$.

La potenza di questo risultato è particolarmente evidente osservando il grafico di una passeggiata aleatoria. Si consideri la passeggiata data da $S_n = \sum_{i=1}^n X_i$, dove le variabili aleatorie X_i rappresentano gli incrementi, o passi del cammino e si supponga che le X_i abbiano tutte la stessa distribuzione e la stessa media μ . Avendo introdotto S_n , allora è possibile riscrivere la media campionaria come $\bar{X}_n = \frac{1}{n} S_n$. La *Legge debole dei Grandi Numeri* indica che $\frac{1}{n} S_n \rightarrow \mu$ in probabilità, ovvero $S_n \approx \mu \cdot n + o(n)$. Nel seguente grafico la linea tratteggiata rappresenta la funzione costante, $E(X_i) = \mu$, mentre la linea continua il valore di \bar{X}_n al variare di n .

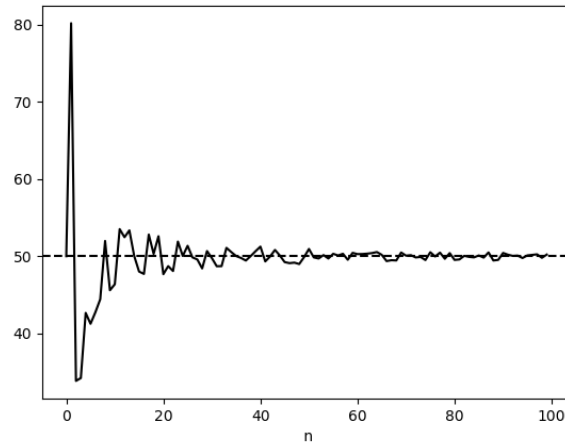


Figura 2.1: Simulazione di una passeggiata aleatoria in cui $\mu = 50$.

Come già è stato accennato, la disuguaglianza di Chebyshev può essere fondamentale per calcolare la velocità di convergenza della probabilità. Per mostrarlo, si consideri il seguente esempio.

Esempio. Si supponga di avere una moneta disonesta e che si voglia verificare con una precisione del 95% che la probabilità che esca croce sia 0.53. A tale scopo si definisca la variabile aleatoria

$$X := \begin{cases} 1 & \text{se esce croce} \\ 0 & \text{altrimenti} \end{cases}$$

Allora:

$$E(X) = 0.53 \qquad \sigma^2 = 0.53 \cdot 0.47 = 0.2491$$

Fissato l'errore $\varepsilon = \frac{1}{100}$ si ripeta il lancio un numero n di volte. Usando la *Legge debole dei Grandi Numeri* si impone che:

$$P\left(\left|\frac{X_1 + \dots + X_n}{n} - 0.53\right| \geq \frac{1}{100}\right) \leq \frac{0.2491 \cdot 100^2}{n} \leq 0.05.$$

Di conseguenza per avere una precisione del 95%, n deve essere almeno $\frac{0.2491 \cdot 100^2}{0.05} = 49820$.

Legge forte dei Grandi Numeri

La sola *Legge debole dei Grandi Numeri* non basta a soddisfare completamente le esigenze che vengono imposte dall'esperienza. Se ad esempio si lanciasse una moneta onesta un centinaio di volte potrebbe succedere con probabilità molto bassa che le frequenze relative siano diverse da $1/2$, mentre questa differenza tenderebbe a diminuire mano a mano che si aumentano i lanci. Una spiegazione può essere data introducendo una nuova definizione di convergenza.

Definizione. Siano $X, (X_n)_{n \in \mathbb{N}}$ delle variabili aleatorie su (Ω, \mathcal{A}, P) . Si dice che la successione $(X_n)_{n \in \mathbb{N}}$ converge a X *P-quasi certamente* se

$$P(\omega \in \Omega : X_n(\omega) \rightarrow X(\omega)) = 1.$$

OSSERVAZIONE. La convergenza quasi certa implica la convergenza in probabilità. Infatti $\forall \varepsilon > 0$:

$$P(|X_n - X| \geq \varepsilon) \leq P(\sup_{k \geq n} |X_k - X| \geq \varepsilon)$$

e passando al limite per $n \rightarrow +\infty$ si ottiene che $P(|X_n - X| \geq \varepsilon) \rightarrow 0$ perchè per ipotesi $(X_n)_{n \in \mathbb{N}}$ convergono P - quasi certamente.

Teorema (Lemma di Borel-Cantelli). *Si consideri una successione di eventi $(A_k)_{k \geq 1}$ nello spazio di probabilità (Ω, \mathcal{A}, P) e sia*

$$A = \limsup_{k \rightarrow +\infty} A_k.$$

1. Se $\sum_{k \geq 1} P(A_k) < \infty$, allora $P(A) = 0$.
2. Se $\sum_{k \geq 1} P(A_k) = \infty$ e se gli eventi A_k sono indipendenti, allora $P(A) = 1$.

Dimostrazione.

1. Sia $m \in \mathbb{N}$. Si ha che $A \subset \bigcup_{k \geq m} A_k$. Quindi per la monotonia e per la subadditività di P :

$$P(A) \leq \sum_{k \geq m} P(A_k).$$

Siccome per ipotesi $\sum_{k \geq 1} P(A_k)$ converge, allora anche $\sum_{k \geq m} P(A_k)$ converge. In particolare per $m \rightarrow \infty$ si ha che $\sum_{k \geq m} P(A_k) \rightarrow 0$. Quindi $P(A) = 0$.

2. Si nota che

$$P(A) = P(\limsup_{k \rightarrow \infty} A_k) = \lim_{k \rightarrow \infty} P\left(\bigcup_{k \geq m} A_k\right).$$

Usando ora le ipotesi di indipendenza, di divergenza della somma e ricordando che $\forall x \ e^{-x} \geq 1 - x$, si calcola:

$$\begin{aligned} P\left(\bigcup_{k \geq m} A_k\right) &= 1 - P\left(\bigcap_{k \geq m} A_k^c\right) = 1 - \lim_{n \rightarrow \infty} P\left(\bigcap_{k=m}^n A_k^c\right) \\ &= 1 - \lim_{n \rightarrow \infty} \left(\prod_{k=m}^n P(A_k^c)\right) = 1 - \lim_{n \rightarrow \infty} \left(\prod_{k=m}^n (1 - P(A_k))\right) \\ &\geq 1 - \lim_{n \rightarrow \infty} \prod_{k=m}^n e^{-P(A_k)} = 1 - \lim_{n \rightarrow \infty} e^{-\sum_{k=m}^n P(A_k)} = 1. \end{aligned}$$

Dunque $P(A) \geq \lim_{n \rightarrow \infty} 1 = 1$, cioè $P(A) = 1$.

□

Con il seguente teorema si vuole mostrare una versione più forte della *Legge dei Grandi Numeri* in cui la convergenza finale è quasi certa.

Teorema (Legge forte dei Grandi Numeri). *Siano $(X_i)_{i \geq 1}$ delle variabili aleatorie sullo spazio di probabilità (Ω, \mathcal{A}, P) , scorrelate e appartenenti a $L^2(\Omega, P)$. Se $\text{Var}(X_i) = \sigma^2 \forall i \geq 1$, allora:*

$$\frac{1}{n} \sum_{i=1}^n (X_i - E(X_i)) \rightarrow 0 \text{ quasi certamente per } n \rightarrow \infty.$$

Dimostrazione. Senza perdita di generalità si fissi n e si supponga che $E(X_i) = 0 \forall i \geq 1$: se così non fosse si può considerare $X'_i := X_i - E(X_i)$. Si ponga poi $Y_n = \frac{1}{n} \sum_{i=1}^n X_i$. La dimostrazione procede in due Step.

Step 1. Si mostra che $Y_{n^2} \rightarrow 0$.

Per ogni $\varepsilon > 0$ la *Legge debole dei Grandi Numeri* implica che $P(|Y_{n^2}| > \varepsilon) \geq \frac{\sigma^2}{n^2 \varepsilon^2}$. Inoltre la somma delle probabilità dei $|Y_{n^2}|$ costituisce una serie convergente per cui è possibile applicare il Lemma di Borel-Cantelli. Segue che:

$$P\left(\limsup_{n \rightarrow \infty} |Y_{n^2}| > \varepsilon\right) = 0$$

e per $\varepsilon \rightarrow 0$ si trova che $P(\limsup_{n \rightarrow \infty} |Y_{n^2}| > 0) = 0$, da cui $|Y_n| \rightarrow 0$.

Step 2. Sia ora $m \in \mathbb{N}$ e sia $n = n(m)$ tale che $n^2 \leq m < (n+1)^2$ e si definisca $S_k = kY_k = \sum_{i=1}^k X_i$. Applicando la disuguaglianza di Chebyshev si ha che:

$$P(|S_m - S_{n^2}| > \varepsilon n^2) \leq \frac{\sigma^2(m - n^2)}{\varepsilon^2 n^4}.$$

Di conseguenza:

$$\begin{aligned} \sum_{m \geq 1} P(|S_m - S_{nm^2}| > \varepsilon nm^2) &\geq \frac{\sigma^2}{\varepsilon^2} \sum_{n \geq 1} \sum_{m=n^2}^{(n+1)^2-1} \frac{m - n^2}{n^4} \\ &= \frac{\sigma^2}{\varepsilon^2} \sum_{n \geq 1} \sum_{k=1}^{2n} \frac{k}{n^4} \\ &= \frac{\sigma^2}{\varepsilon^2} \sum_{n \geq 1} \frac{2n(2n+1)}{2n^4} < \infty. \end{aligned}$$

Per il lemma di Borel-Cantelli si ottiene che $P\left(\left|\frac{S_m}{nm^2} - Y_{nm^2}\right| \rightarrow 0\right) = 1$ per $m \rightarrow \infty$ e unendo questo risultato con quanto visto nello Step 1 si deduce che

$$P\left(\frac{S_m}{nm^2} \rightarrow 0\right) = 1 \text{ per } m \rightarrow \infty.$$

Infine da $|Y_m| = \frac{|S_m|}{m} \geq \frac{|S_m|}{nm^2}$ si conclude che $P(Y_m \rightarrow 0) = 1$.

□

Capitolo 3

Una legge per variabili aleatorie correlate

Le formulazioni della *Legge dei Grandi Numeri* fino ad ora trattate ruotano attorno ad una principale ipotesi: la scorrelazione di variabili aleatorie. Chiaramente, questa assunzione impedisce di poter applicare la Legge in qualsiasi caso di dipendenza e sarà scopo di questo capitolo estenderne la validità anche in queste situazioni.

Il caso più semplice di correlazione è rappresentato dalle *Catene di Markov* per cui, la previsione riguardo il comportamento di una variabile aleatoria dipende solo dallo stato attuale e non dall'evoluzione passata.

3.1 Catene di Markov

Definizioni e proprietà

Si considerino le variabili aleatorie discrete $(X_n)_{n \in \mathbb{N}}$ definite sullo stesso spazio di probabilità (Ω, \mathcal{A}, P) a valori in S . La successione $(X_n)_{n \in \mathbb{N}}$ prende il nome di *processo stocastico*, mentre l'insieme S si definisce *spazio degli stati* e può essere finito oppure numerabile.

Al variare di n , il processo può passare da uno stato j ad uno stato k e nel caso in cui $j \neq k$ si dice che il processo ha compiuto una *transizione*.

Definizione. Un processo stocastico discreto $(X_n)_{n \in \mathbb{N}}$ si dice *catena di Markov* su S se vale la *proprietà di Markov*:

$$P(X_n = k | X_{n-1} = j, X_h = j_h \forall h \in [0, n-2]) = P(X_n = k | X_{n-1} = j) =: p_{j,k}$$

Le $p_{j,k}$ si dicono *probabilità di transizione* e sono tali che:

$$\begin{cases} p_{j,k} \geq 0 \\ \sum_{k \in S} p_{j,k} = 1 \quad \forall j \in S. \end{cases}$$

La distribuzione su S :

$$\begin{aligned} \lambda : S &\rightarrow [0, 1] \\ k &\rightarrow \lambda_k \end{aligned}$$

viene chiamata *distribuzione iniziale* ed è tale per cui: $P(X_0 = k) = \lambda_k \forall k \in S$.
 Infine si definisce la *probabilità di transizione in n passi* come $p_{j,k}^{(n)} = P(X_n = k | X_0 = j)$ e rappresenta la probabilità che il processo, dallo stato j giunga allo stato k in n passi.

La proprietà di Markov suggerisce il fatto che la distribuzione di X_{n+1} condizionata dai valori X_0, \dots, X_n dipende solo dal valore dello stato presente X_n e non dal valore delle variabili precedenti. Inoltre l'omogeneità rende la distribuzione indipendente da n : $P(X_{n+1} = k | X_n = j) = P(X_1 = k | X_0 = j) = p_{j,k}$.

Di seguito si mostrano alcune semplici proprietà.

Lemma. Se $(X_n)_{n \in \mathbb{N}}$ è una catena di Markov con distribuzione iniziale

$$\begin{aligned} \lambda : \mathbb{N} &\rightarrow [0, 1] \\ k &\rightarrow \lambda_k \end{aligned}$$

e probabilità di transizione $p_{j,k}$, allora:

1. $P(X_0 = k_0, \dots, X_n = k_n) = \lambda_{k_0} \cdot p_{k_0, k_1} \cdots p_{k_{n-1}, k_n}$

2. equazione di Chapman-Kolmogorov:

$$p_{j,k}^{(n+m)} = \sum_{i=0}^{\infty} p_{j,i}^{(n)} p_{i,k}^{(m)} \quad \forall n, m \geq 0, j, k \in \mathbb{N}.$$

3. $P(X_n = k) = \sum_{j \in \mathbb{N}} \lambda_j p_{j,k}^{(n)}$

Dimostrazione.

1. Dal momento in cui $P(A \cap B) = P(A|B)P(B)$:

$$\begin{aligned} P(X_0 = k_0, \dots, X_n = k_n) &= \\ P(X_0 = k_0, \dots, X_{n-1} = k_{n-1})P(X_n = k_n | X_0 = k_0, \dots, X_{n-1} = k_{n-1}) &= \\ P(X_0 = k_0, \dots, X_{n-1} = k_{n-1})P(X_n = k_n | X_{n-1} = k_{n-1}) &= \\ P(X_0 = k_0, \dots, X_{n-1} = k_{n-1})p_{k_{n-1}, k_n} & \end{aligned}$$

Iterando il ragionamento per $P(X_0 = k_0, \dots, X_{n-1} = k_{n-1})$, si ottiene che:

$$P(X_0 = k_0, \dots, X_n = k_n) = \lambda_{k_0} \cdot p_{k_0, k_1} \cdots p_{k_{n-1}, k_n}$$

2.

$$\begin{aligned}
 p_{j,k}^{(n+m)} &= P(X_{n+m} = k | X_0 = j) \\
 &= \sum_{i=0}^{\infty} P(X_{n+m} = k, X_n = i | X_0 = j) \\
 &= \sum_{i=0}^{\infty} P(X_{n+m} = k | X_0 = j, X_n = i) P(X_n = i | X_0 = j) \\
 &= \sum_{i=0}^{\infty} P(X_{n+m} = k | X_n = i) p_{j,i}^{(n)} \\
 &= \sum_{i=0}^{\infty} P(X_m = k | X_0 = i) p_{j,i}^{(n)} \\
 &= \sum_{i=0}^{\infty} p_{i,k}^{(m)} p_{j,i}^{(n)}
 \end{aligned}$$

3. Si applica 2. con $m = 0$.

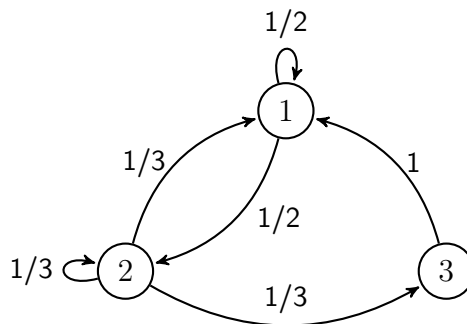
□

Una catena di Markov può essere rappresentata attraverso una matrice stocastica i cui elementi sono le probabilità di transizione, e di conseguenza può essere illustrata tramite un grafo orientato.

Si consideri ad esempio, la catena di Markov data dalla matrice

$$P = (p_{j,k})_{j,k \in S} = \begin{pmatrix} 1/2 & 1/2 & 0 \\ 1/3 & 1/3 & 1/3 \\ 1 & 0 & 0 \end{pmatrix}$$

sullo spazio $S = \{1, 2, 3\}$. Allora il grafo orientato associato è:



Classificazione degli stati

Sia $(X_n)_{n \in \mathbb{N}}$ una catena di Markov sullo spazio S .

Definizione. Si dice che j comunica con k se esiste $n \geq 0$ tale che $p_{j,k}^{(n)} > 0$. In tal caso si scrive $j \rightarrow k$.

Se $j \rightarrow k$ e $k \rightarrow j$ si dice che gli stati j e k sono *equivalenti* e l'equivalenza tra i due stati si indica con: $j \leftrightarrow k$.

Capire se due stati comunicano può essere molto semplice esaminando il grafo di transizione e cercando un cammino che colleghi j con k . Ritornando all'esempio precedente si può dedurre che: $1 \leftrightarrow 2, 2 \rightarrow 3, 3 \rightarrow 1$.

Chiaramente ogni stato comunica con se stesso e questo fatto rende " \leftrightarrow " una relazione di equivalenza. Se ne deduce che lo spazio S può essere partizionato in classi di equivalenza.

Definizione. Se tutti gli stati di una catena di Markov comunicano tra loro, ovvero se $\forall j, k > 0 \exists n \geq 1$ tale che $p_{j,k}^{(n)} > 0$, la catena è detta *irriducibile*.

Equivalentemente, una catena di Markov si dice *irriducibile* se ha un'unica classe di equivalenza e quindi se il grafo di transizione è semplicemente connesso.

Definizione. Per ogni $j \in S$ si pone $N_j = \{n \geq 1 : p_{j,j}^{(n)} > 0\}$ e si definisce il *periodo* di j come: $d(j) = MCD(N_j)$.

Definizione. Lo stato $j \in S$ si dice:

- *aperiodico*: se $d(j) = 1$;
- *assorbente*: se $p_{j,k} = 0 \forall k \neq j$;
- *ricorrente*: se la probabilità che il processo ritorni in j è pari a 1, ovvero se $P(\exists n \in \mathbb{N} : X_n = j) = 1$;
- *transiente*: se non è ricorrente, ovvero se $P(\exists n \in \mathbb{N} : X_n = j) < 1$;
- *ricorrente positivo*: se è ricorrente e se $E(T_j) < +\infty$, dove $T_j := \min\{n \geq 1 : X_n = j\}$;
- *ergodico*: se è aperiodico e ricorrente positivo.

Si osservi che se uno stato è ricorrente, allora il processo visiterà tale stato infinite volte quindi: uno stato j di una catena di Markov è ricorrente se e soltanto se il processo transita in j infinite volte. Analogamente, uno stato j di una catena di Markov è transiente se e soltanto se esiste uno stato k tale che $j \rightarrow k$ e $k \not\rightarrow j$.

3.2 Teorema ergodico

Il prossimo teorema permette di studiare l'andamento asintotico di una catena di Markov su uno spazio i cui stati sono irriducibili ed ergodici e mostra quale sia la velocità di convergenza.

Teorema (Teorema ergodico). *Si consideri una catena di Markov $(X_n)_{n \in \mathbb{N}}$ su uno spazio S e con matrice di transizione $P = (p_{j,k})_{j,k \in S}$. Se $(X_n)_{n \in \mathbb{N}}$ è irriducibile ed ergodica, allora esiste un'unica distribuzione $\pi = (\pi_i)_{i \in S}$ su S tale che $\forall j, k \in S$:*

$$\lim_{n \rightarrow \infty} p_{j,k}^{(n)} = \pi_k$$

Inoltre la velocità di convergenza è esponenziale, ovvero:

$$\exists c, \delta > 0 \text{ tali che } |p_{j,k}^{(n)} - \pi_k| \leq ce^{-\delta n}$$

Una prima importante osservazione viene dedotta applicando l'equazione di Chapman-Kolmogorov:

$$\begin{aligned} \pi_k &= \lim_{n \rightarrow \infty} p_{j,k}^{(n)} = \lim_{n \rightarrow \infty} \sum_{i \in S} p_{j,i}^{(n-1)} p_{i,k} \\ &= \sum_{i \in S} p_{i,k} \lim_{n \rightarrow \infty} p_{j,i}^{(n-1)} \\ &= \sum_{i \in S} p_{i,k} \pi_i = (\pi \cdot P)_k \end{aligned}$$

dove $(\pi \cdot P)_k$ indica la componente k -esima del prodotto tra il vettore riga π e la matrice P .

Si può dunque scrivere che $\pi = \pi P$ e questa uguaglianza suggerisce che in realtà π è l'unico autovettore sinistro associato all'autovalore 1 della matrice P . Per tale motivo la ricerca di π si riduce a determinare l'unica soluzione del sistema

$$\begin{cases} \pi_k = (\pi \cdot P)_k & \forall k \in S \\ \sum_{k \in S} \pi_k = 1 \end{cases}$$

composto da $|S| + 1$ equazioni e $|S|$ incognite.

Definizione. Una distribuzione P che soddisfa $\pi = \pi P$ viene detta *stazionaria* oppure *invariante*.

In secondo luogo, si osservi che il limite non dipende in alcun modo dallo stato iniziale j scelto. Infine dalla proprietà 3. delle catene di Markov si conclude che per n sufficientemente grande, la distribuzione di X_n può essere approssimata con π e tale approssimazione non dipende dalla distribuzione iniziale λ . Infatti:

$$P(X_n = k) = \sum_{j \in S} \lambda_j p_{j,k}^{(n)} \longrightarrow \sum_{j \in S} \lambda_j \pi_k = \pi_k \quad \text{per } n \rightarrow \infty$$

Si ritorni all'esempio della catena di Markov con matrice di transizione $P = \begin{pmatrix} 1/2 & 1/2 & 0 \\ 1/3 & 1/3 & 1/3 \\ 1 & 0 & 0 \end{pmatrix}$.

Risolvendo il sistema $\pi = \pi P$ si trova che $\pi = (\frac{1}{2}, \frac{3}{8}, \frac{1}{8})$ e applicando il teorema ergodico:

$$P^n \longrightarrow \begin{pmatrix} 1/2 & 3/8 & 1/8 \\ 1/2 & 3/8 & 1/8 \\ 1/2 & 3/8 & 1/8 \end{pmatrix} \text{ per } n \rightarrow \infty.$$

Prima di concludere la sezione è bene dare un'ultima definizione.

Definizione. Una catena di Markov $(X_n)_{n \in \mathbb{N}}$ ergodica e con distribuzione stazionaria $\pi = (p_i)_{i \in S}$, è detta reversibile se

$$\forall k, j \in S \quad \pi_j p_{j,k} = \pi_k p_{k,j}.$$

Una catena di Markov è reversibile se mantiene le stesse probabilità di transizione, anche se viene percorsa al contrario.

3.3 Metodo Monte Carlo (MCMC)

Si consideri la catena di Markov $(X_n)_{n \in \mathbb{N}}$ e si fissi un k nello spazio degli stati S e si definisca la successione di variabili aleatorie

$$Y_j^{(k)} = \mathbb{1}_{\{X_j = k\}} := \begin{cases} 1 & \text{se } X_j = k \\ 0 & \text{altrimenti.} \end{cases}$$

Teorema. Il teorema ergodico suggerisce che per $n \gg 1$, la distribuzione di X_n tende a π e di conseguenza $Y_n^{(k)} \approx Be(\pi_k)$. Per la Legge dei Grandi Numeri, la media campionaria $\bar{Y}_n^{(k)}$ tende a π_k , ovvero:

$$\bar{Y}_n^{(k)} = \frac{Y_1^{(k)} + Y_2^{(k)} + \dots + Y_n^{(k)}}{n} \longrightarrow \pi_k \text{ per } n \gg 1$$

Dal momento in cui $Y_1^{(k)} + Y_2^{(k)} + \dots + Y_n^{(k)}$ coincide con il numero di volte per cui lungo il processo transita in k , è possibile interpretare la distribuzione π_k come la probabilità che il processo si trovi nello stato k .

Questo modo di intendere π è alla base del *metodo Monte Carlo* basato sulle Catene di Markov, che permette di risolvere problemi e simulare risultati a partire dalla generazione di variabili casuali ed è particolarmente adatta nel caso in cui S ha una cardinalità elevata.

Si supponga di dover calcolare il valor medio ϑ di una funzione $\varphi : S \rightarrow \mathbb{R}$ per cui è complicato, o computazionalmente dispendioso, conoscere i valori che può assumere.

L'idea del *Metodo Monte Carlo* consiste nel generare una famiglia di variabili aleatorie $(X_n)_{n \in \mathbb{N}}$, con probabilità stazionarie $\pi = (\pi_x)_{x \in S}$ note e nell'approssimare ϑ con il $\lim_{n \rightarrow \infty} \sum_{k=1}^n \frac{1}{n} \varphi(X_k)$. Infatti, per la *Legge dei Grandi Numeri*:

$$\vartheta = E(\varphi(X)) = \sum_{x \in S} \varphi(x) \pi_x = \lim_{n \rightarrow \infty} \sum_{k=1}^n \frac{\varphi(X_k)}{n}.$$

Algoritmo di Hastings-Metropolis

Nel seguito, viene presentato l'*algoritmo di Hastings-Metropolis* che permette di costruire una catena di Markov $(X_n)_n$ reversibile in cui le probabilità stazionarie valgono

$$\pi_k = \frac{b_k}{\sum_{k=1}^{\infty} b_k},$$

dove i b_k sono numeri positivi di somma finita.

L'algoritmo comincia considerando una matrice di transizione irriducibile $Q = (q_{j,k})_{j,k}$ e inizializzando la variabile X_0 ad un valore iniziale x_0 . Procede poi, definendo $(X_n)_n$ in maniera ricorsiva: se all'iterazione i -esima X_i vale j , allora si pone

$$X_{i+1} = \begin{cases} j & \text{con probabilità } \alpha_{j,k}, \\ i & \text{con probabilità } 1 - \alpha_{j,k}. \end{cases}$$

Di conseguenza le probabilità di transizione $p_{j,k}$ sono le seguenti:

$$p_{j,k} = P(X_{i+1} = k | X_i = j) = q_{j,k} \alpha_{j,k} \quad \text{per } j \neq k$$

$$p_{j,j} = 1 - \sum_{j \neq k} p_{j,k} = q_{j,j} + \sum_{j \neq k} q_{j,k} \alpha_{j,k}.$$

Ponendo $\alpha_{j,k} = \min\left(\frac{\pi_k q_{k,j}}{\pi_j q_{j,k}}\right)$, la catena di Markov che è stata generata è reversibile e ha distribuzione stazionaria. Infatti, la condizione di reversibilità equivale a:

$$\pi_j q_{j,k} \alpha_{j,k} = \pi_k q_{k,j} \alpha_{k,j},$$

da cui se $\frac{\pi_k q_{k,j}}{\pi_j q_{j,k}} < 1$, allora $\alpha_{j,k} = \frac{\pi_k q_{k,j}}{\pi_j q_{j,k}}$ e $\alpha_{k,j} = 1$. Quindi:

$$\pi_j q_{j,k} \frac{\pi_k q_{k,j}}{\pi_j q_{j,k}} = \pi_k q_{k,j} \cdot 1$$

Mentre se $\frac{\pi_k q_{k,j}}{\pi_j q_{j,k}} > 1$, allora $\alpha_{j,k} = 1$ e $\alpha_{k,j} = \frac{\pi_j q_{j,k}}{\pi_k q_{k,j}}$ e:

$$\pi_j q_{j,k} \cdot 1 = \pi_k q_{k,j} \cdot \frac{\pi_j q_{j,k}}{\pi_k q_{k,j}}$$

In entrambi i casi la condizione di reversibilità è soddisfatta.

Capitolo 4

Applicazioni

Anche se nel capitolo precedente è stato approfondito un algoritmo specifico, i metodi Monte Carlo in realtà, sono diversi e dipendono dal problema.

Nelle prossime sezioni, verranno presentate alcune applicazioni del metodo.

4.1 Calcolo di integrali

È noto che non per tutte le funzioni limitate è possibile calcolare con esattezza l'integrale; è possibile però, sviluppare un algoritmo che permette di approssimarne il valore.

Si consideri allora una funzione limitata $f : [a, b] \rightarrow \mathbb{R}$ e si supponga di voler calcolare $\int_a^b f(x)dx$.

L'algoritmo comincia generando delle variabili aleatorie casuali X_1, \dots, X_n uniformi in $[a, b]$: nel caso in cui la generazione dei numeri casuali in un intervallo generico $[a, b]$ non sia possibile, questa può essere realizzata generando dei numeri casuali U_1, \dots, U_n con $U_i \sim U(0, 1)$ e ottenendo da questi le $X_i = a + (b - a)U_i$.

Si procede poi calcolando i valori $f(X_1), \dots, f(X_n)$ e osservando che $E(f(X)) = \frac{1}{b-a} \int_a^b f(x)dx$. Siccome le X_i sono indipendenti e identicamente distribuite ed f è limitata, per ogni $i = 1, \dots, n$, $f(X_i) \in L^2(\mathbb{R}, P)$ e quindi le ipotesi della *Legge debole dei Grandi Numeri* sono soddisfatte. Pertanto, $\forall \varepsilon > 0$:

$$\lim_{n \rightarrow +\infty} P \left(\left| \frac{1}{n} \sum_{i=1}^n f(X_i) - \frac{1}{b-a} \int_a^b f(x)dx \right| > \varepsilon \right) = 0$$

Si deduce allora, che per un n opportunamente grande, $\frac{b-a}{n} \sum_{i=1}^n f(X_i)$ rappresenta una buona approssimazione dell'integrale $\int_a^b f(x)dx$.

Attraverso questo metodo, è inoltre possibile calcolare quanti numeri è opportuno generare per poter avere con una probabilità elevata, ma non certa, un errore non superiore ad un fissato ε .

Definendo $M := \sup_{x \in [a, b]} |f(x)| < +\infty$, si nota che $\forall i = 1, \dots, n$:

$$Var(f(X_i)) \leq E((f(X))^2) \leq M^2.$$

Di conseguenza, fissata la soglia $\delta > 0$:

$$P\left(\left|\frac{1}{n}\sum_{i=1}^n f(X_i) - \frac{1}{b-a}\int_a^b f(x)dx\right| > \varepsilon\right) \leq \frac{\text{Var}(f(X_i))}{n\varepsilon^2} \leq \frac{M^2}{n\varepsilon^2} \leq \delta,$$

da cui $n \geq \frac{M^2}{\delta\varepsilon^2}$.

Tuttavia questo modo non è l'unico per risolvere integrali. Viene presentato ora un altro metodo tramite un esempio.

Esempio. Si supponga di voler calcolare il seguente integrale:

$$\int_0^1 \frac{e^{\sin(x^3)}}{3(1+5x^8)} dx$$

e si indichi con I l'area che sottostà al grafico della funzione in $f = \frac{e^{\sin(x^3)}}{3(1+5x^8)}$, ovvero $I = \int_0^1 \frac{e^{\sin(x^3)}}{3(1+5x^8)} dx$. Si vuole costruire una variabile aleatoria X tale che $E(X) = I$. Per farlo si prendano due numeri casuali x_1, y_1 nell'intervallo $[0, 1]$ e si definisca X_1 nel seguente modo:

$$X_1 := \begin{cases} 1 & \text{se } y_1 \leq f(x_1), \\ 0 & \text{se } y_1 \geq f(x_1). \end{cases}$$

Equivalentemente, la variabile X_1 vale 1 se il punto (x, y) , appartenente al quadrato $Q := [0, 1] \times [0, 1]$, si trova nella porzione sottostante al grafico di f , mentre vale 0 se (x, y) è in quella sovrastante. Allo stesso modo si generano le variabili X_2, \dots, X_n . Allora, se $A = 1$ è l'area di Q :

$$P(X = 1) = P(y \leq f(x)) = \frac{I}{A} = I.$$

Si conclude, dunque che:

$$E(X) = P(X = 1) = I,$$

$$\text{Var}(X) = I(1 - I)$$

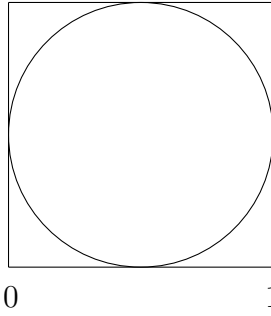
e applicando la *Legge dei Grandi Numeri* e la *Disuguaglianza di Chebyshev*, $\forall \varepsilon > 0$

$$P\left(\left|\frac{X_1 + \dots + X_n}{n} - I\right| \geq \varepsilon\right) \leq \frac{I(1 - I)}{n\varepsilon^2}$$

4.2 Algoritmo per l'approssimazione di π

Un algoritmo simile a quello utilizzato nell'esempio precedente, permette di approssimare il valore di π .

Si consideri un quadrato di lato 1 e la circonferenza unitaria ad esso inscritta.



Il metodo comincia generando due numeri casuali x_1 e y_1 nell'intervallo $[0, 1]$, ovvero generando un punto casuale all'interno del quadrato $[0, 1] \times [0, 1]$. Se $x_1^2 + y_1^2 \leq 1$ si pone $X_1 = 1$, altrimenti $X_1 = 0$ e si procede allo stesso modo per determinare X_2, \dots, X_n . Di conseguenza:

$$P(X = 1) = P(x^2 + y^2 \leq 1) = \frac{\text{area della circonferenza unitaria}}{\text{area del quadrato di lato 1}} = \frac{\pi}{4}.$$

Quindi $E(X) = \frac{\pi}{4}$ e $Var(X) = \frac{\pi}{4} \left(1 - \frac{\pi}{4}\right)$. Allora:

$$P\left(\left|\frac{X_1 + \dots + X_n}{n} - \frac{\pi}{4}\right| \geq \varepsilon\right) \leq \frac{\frac{\pi}{4} \left(1 - \frac{\pi}{4}\right)}{n\varepsilon^2} \text{ con } \varepsilon > 0.$$

Siccome non si conosce il valore esatto di $\frac{\pi}{4} \left(1 - \frac{\pi}{4}\right)$ si studia la funzione $p(1 - p)$ nell'intervallo $[0, 1]$ e si osserva che questa ha un massimo in $p = \frac{1}{2}$, in cui vale $\frac{1}{4}$. Se si volesse approssimare il valore di π con una tolleranza $\frac{1}{1000}$, ovvero se si volesse approssimare $\frac{\pi}{4}$ con una precisione di $\varepsilon = \frac{1}{4000}$, si ha che:

$$P\left(\left|\frac{X_1 + \dots + X_n}{n} - \frac{\pi}{4}\right| \geq \varepsilon\right) \leq \frac{1/4}{n(1/4.000)^2} = \frac{4.000.000}{n}$$

Di conseguenza se si volesse avere una precisione del 99%, si impone la condizione:

$$\frac{4.000.000}{n} \leq \frac{1}{100}$$

che determina il numero minimo di iterazioni $n \geq 400 \times 10^6$.

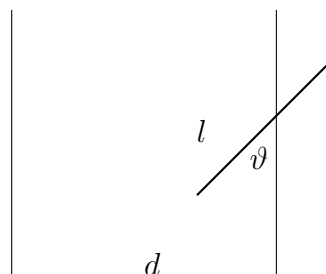
4.3 Il problema dell'ago di Buffon

Nel XVIII secolo, il matematico e naturalista francese Georges-Louis Leclerc, conte di Buffon propose il seguente problema. Si consideri un pavimento a strisce verticali di larghezza d e si lasci cadere da una certa altezza un ago di lunghezza l : quale è la probabilità che l'ago una volta atterrato, intersechi una delle linee che compongono il motivo del pavimento?

I modi di procedere sono molteplici e possono utilizzare approcci completamente diversi. In questa sezione ne verranno presentati due, in cui nel primo si fa uso della geometria e del calcolo integrale, mentre nel secondo del valore atteso.

Soluzione geometrica

Nel momento in cui atterra, l'ago occupa una determinata posizione sul pavimento che può essere descritta tramite una coordinata orizzontale e un angolo acuto di inclinazione ϑ , rispetto alla linea verticale più vicina.



Per un fissato angolo ϑ , la probabilità che l'ago intersechi la linea coincide con il rapporto:

$$P(\vartheta) = \frac{\text{componente orizzontale dell'ago}}{\text{distanza tra le due linee verticali}} = \frac{l \sin \vartheta}{d}.$$

Di conseguenza, usando il *teorema della media integrale*, è possibile calcolare la probabilità richiesta per un qualsiasi angolo acuto ϑ :

$$P = \frac{1}{\frac{\pi}{2} - 0} \int_0^{\frac{\pi}{2}} \frac{l \sin \vartheta}{d} d\vartheta = \frac{2l}{\pi d}.$$

Prova di Barbier

La risposta al quesito di Leclerc suggerita dal matematico francese Barbier (1839-1889), fa uso del valor medio per variabili aleatorie ed è a dir poco geniale: la soluzione infatti, è immediata se si lancia il giusto ago.

Quando cade al suolo, a seconda della sua lunghezza l , l'ago può intersecare le linee anche più di una volta e per tale motivo, il numero atteso di intersezioni è:

$$E(l) = 1 \cdot p_1 + 2 \cdot p_2 + 3 \cdot p_3 + \dots$$

dove le p_i indicano le probabilità che l'ago inersechi i volte le linee che compongono il motivo del pavimento.

Nel caso in cui l'ago abbia lunghezza pari a metà della distanza delle linee, si ha che $p_2 = p_3 = \dots = 0$ e quindi che $E(1) = p_1$ e il problema viene tradotto in termini del valor medio.

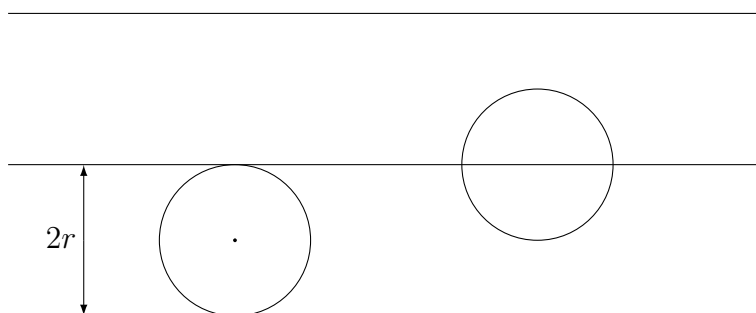
Anche se apparentemente più complicata, questa formulazione propone una furba osservazione. Si supponga di dividere l'ago in due parti di lunghezza x e y , come in figura.



Per linearità, $E(1) = E(x + y) = E(x) + E(y)$ e questa formula continua a valere anche se l'ago fosse composto da due segmenti spezzati. Si deduce allora, che il numero atteso di intersezioni dipende solo dal perimetro dell'ago e non dalla sua forma. In particolare $E(n) = n \cdot E(1)$ e per un poligono P a n lati vale $E(P) = \text{perimetro}(P)E(1)$.

Si lancino ora, un ago circolare C di raggio $r = d/2$ e due aghi poligonali ad n lati P_n e P^n , uno inscritto e l'altro circoscritto a C . Per monotonia del valor medio: $E(P_n) \leq E(C) \leq E(P^n)$.

Come si vede nella figura, indifferentemente dalla sua posizione, l'ago circolare interseca sempre due volte le linee che compongono il motivo.



Di conseguenza $E(P_n) \leq 2 \leq E(P^n)$ e calcolando esplicitamente il valore dei due aghi polinomiali si trova che:

$$\text{perimetro}(P_n)E(1) \leq 2 \leq \text{perimetro}(P^n)E(1).$$

Chiaramente, mandando n all'infinito e supponendo per semplicità che r sia pari a 1, i poligoni convergono alla circonferenza unitaria e il loro perimetro tende a 2π . Allora:

$$2\pi E(1) \leq 2 \leq 2\pi E(1)$$

da cui, dividendo per 2π , si conclude che:

$$E(1) = p_1 = \frac{1}{\pi}.$$

Il problema dell'ago di Buffon, permette di sviluppare un altro metodo Monte Carlo per approssimare il valore di π , per cui le variabili aleatorie casuali coincidono con le posizioni di N aghi.

Si torni al caso generale. Dal momento in cui la probabilità che un'ago di lunghezza l intersechi una linea che forma una striscia del pavimento di larghezza d , vale $P = \frac{2l}{\pi d}$, è possibile calcolare: $\pi = \frac{2l}{Pd}$.

Si supponga allora di lasciar cadere N aghi e si contino quelli che toccano una linea. Se questi sono in numero h , allora $P = h/N$. Quindi:

$$\pi \approx \frac{2lN}{hd}.$$

Nel caso particolare in cui $d = 2, l = 1$ si trova che $\pi \approx \frac{N}{h}$.

4.4 Applicazioni Teoriche

Il teorema di approssimazione di Weierstrass

La *Legge dei Grandi Numeri* propone anche alcune applicazioni teoriche, ad esempio permette di dimostrare in maniera alternativa il *Teorema di approssimazione di Weierstrass* (1885).

Teorema. *Data una qualsiasi funzione continua $f : [0, 1] \rightarrow \mathbb{R}$, esiste una successione di polinomi che approssima f in modo uniforme.*

OSSERVAZIONE. L'enunciato può essere esteso ad una qualsiasi funzione definita su un intervallo generico $[a, b]$ tramite opportune trasformazioni.

Nel 1911, Bernstein associa ad ogni funzione $f : [0, 1] \rightarrow \mathbb{R}$ la successione di *polinomi di Bernstein*:

$$p_n^f(x) := \sum_{k=0}^n \binom{n}{k} f\left(\frac{k}{n}\right) x^k (1-x)^{n-k}.$$

e dimostra che:

Teorema. *Se $f : [0, 1] \rightarrow \mathbb{R}$ è una funzione continua allora la successione dei polinomi di Bernstein converge uniformemente a f , ovvero:*

$$\forall \varepsilon > 0 \exists n_\varepsilon \in \mathbb{N} : \sup_{x \in [0,1]} |f(x) - p_n^f(x)| \leq \varepsilon, \quad \forall n \geq n_\varepsilon.$$

Per farlo, osserva che per il teorema di Heine-Cantor una funzione continua in uno spazio compatto è uniformemente continua. Quindi:

$$\forall \varepsilon > 0 \exists \delta_\varepsilon > 0 : \forall x, y \in [0, 1], \quad |y - x| \leq \delta_\varepsilon \implies |f(y) - f(x)| \leq \frac{\varepsilon}{2}.$$

Inoltre f ammette massimo e minimo, cioè esiste $M := \sup_{x \in [0,1]} |f(x)| < +\infty$.

Fissa poi x in $[0, 1]$, ε e δ_ε e definisce n variabili aleatorie di Bernoulli indipendenti X_1, \dots, X_n su uno spazio di probabilità (Ω, \mathcal{A}, P) e deduce che $\forall i \in \{0, \dots, n\}, E(X_i) = x$, mentre $Var(X_i) = x(1-x)$.

Considerando poi la media campionaria delle variabili $\bar{X}_n = \frac{1}{n} \sum_{k=0}^n X_k$, trova che:

$$E(f(\bar{X}_n)) = p_n^f(x).$$

Se inoltre viene introdotto l'evento: $A_{n,\varepsilon} = \{|\bar{X}_n - x| \geq \delta_\varepsilon\}$, allora per la *Disuguaglianza di Chebyshev*:

$$P(A_n) \leq \frac{x(1-x)}{n\delta_\varepsilon^2} \leq \max_{x \in [0,1]} \frac{x(1-x)}{n\delta_\varepsilon^2} = \frac{1}{4n\delta_\varepsilon^2}.$$

Scomponi poi:

$$E(|f(x) - f(\bar{X}_n)|) = E(|f(x) - f(\bar{X}_n)| \cdot \mathbb{1}_{A_{n,\varepsilon}}) + E(|f(x) - f(\bar{X}_n)| \cdot \mathbb{1}_{A_{n,\varepsilon}^c})$$

e maggiore:

$$|f(x) - f(\bar{X}_n)| \leq |f(x)| + |f(\bar{X}_n)| \leq 2M;$$

da cui:

$$E(|f(x) - f(\bar{X}_n)| \cdot \mathbb{1}_{A_{n,\varepsilon}}) \leq 2ME(\mathbb{1}_{A_{n,\varepsilon}}) = 2MP(A_{n,\varepsilon}) \leq \frac{M}{2n\delta_\varepsilon^2}.$$

Infine osserva che se $\omega \in A_{n,\varepsilon}^c$, allora $|\bar{X}_n(\omega) - x| < \delta_\varepsilon$, quindi per continuità $|f(x) - f(\bar{X}_n)| \leq \varepsilon/2$. Di conseguenza:

$$E(|f(x) - f(\bar{X}_n)| \cdot \mathbb{1}_{A_{n,\varepsilon}^c}) \leq \frac{\varepsilon}{2}E(\mathbb{1}_{A_{n,\varepsilon}^c}) = \frac{\varepsilon}{2}P(A_{n,\varepsilon}^c) \leq \frac{\varepsilon}{2}.$$

Unendo le stime precedenti, deduce quindi che:

$$|f(x) - p_n^f(x)| = |f(x) - E(f\bar{X}_n)| \leq E(|f(x) - f(\bar{X}_n)|) \leq \frac{M}{2n\delta_\varepsilon^2} + \frac{\varepsilon}{2}$$

e questo vale per ogni $x \in [0, 1]$, $\varepsilon > 0$. Allora ottiene la tesi ponendo $n_\varepsilon = \frac{M}{\delta_\varepsilon^2 \varepsilon}$.

Entropia

Il calcolo delle probabilità attribuisce ad ogni fenomeno il suo grado di fiducia, ovvero permette di stabilire il “prezzo” che un individuo è disposto a pagare per ottenere una vincita nel caso in cui un evento si verifica. Parrebbe quindi esserci uno sfondo caotico dietro alla legge di ogni variabile aleatoria.

È il concetto di *entropia* che permette di misurare questa quantità di *disordine*.

Definizione. Sia X una variabile aleatoria definita su un insieme finito F e si indichi con p_X la densità $P(X = x)$, con $x \in X$. Si definisce l'*entropia di X* il valore:

$$H_X := E(-\log p_X(X)) = -\sum_{x \in F} p_x(x) \log p_X(x).$$

Coerentemente con sua interpretazione vale la seguente:

Proposizione. *Date due variabili aleatorie indipendenti X e Y definite sugli insiemi finiti F e G , si consideri la variabile (X, Y) definita su $F \times G$. Allora:*

$$H_{(X,Y)} = H_X + H_Y.$$

Il prossimo risultato mostra quali siano gli estremi dell'intervallo in cui è contenuta H_X e quando l'entropia, cioè il “disordine” assume il valore massimo e minimo.

Proposizione. *Data una variabile aleatoria X a valori in F , con $|F| < \infty$ allora:*

$$0 \leq H_X \leq \log |F|.$$

Inoltre l'entropia è pari a 0 se e solo se la variabile aleatoria è costante quasi certamente, mentre vale $\log |F|$ se e solo se X ha distribuzione uniforme.

La *Legge dei Grandi Numeri* entra in gioco nella dimostrazione del *Teorema di equipartizione asintotica* il quale descrive lo spazio in cui si trova un vettore aleatorio (X_1, \dots, X_n) . Nello specifico, più la quantità H_X è grande più questo spazio ha dimensione maggiore e i valori di (X_1, \dots, X_n) si disperdono, viceversa ad una entropia minore corrisponde uno spazio di dimensione inferiore e i valori di (X_1, \dots, X_n) sono “meno disordinati”.

Teorema (Teorema di equipartizione asintotica). *Data una variabile aleatoria X definita nell'insieme F di cardinalità finita, e delle variabili aleatorie indipendenti $(X_i)_{i \in \mathbb{N}}$ che hanno la stessa distribuzione di X , allora $\forall \varepsilon > 0$:*

- esiste una famiglia $A_n \subseteq F^n$ con $|A_n| \leq e^{n(H_X + \varepsilon)}$ per cui

$$\lim_{n \rightarrow \infty} P((X_1, \dots, X_n) \in A_n) = 1$$

- esiste una famiglia $B_n \subseteq F^n$ con $|B_n| \leq e^{n(H_X - \varepsilon)}$ per cui

$$\lim_{n \rightarrow \infty} P((X_1, \dots, X_n) \in B_n) = 0$$

Dimostrazione. Sia $\varepsilon > 0$ e si definisca la famiglia

$$A_n = A_n^\varepsilon := \left\{ (x_1, \dots, x_n) \in F^n : H_X - \varepsilon \leq -\frac{1}{n} \sum_{i=1}^n \log p_X(x_i) \leq H_X + \varepsilon \right\}.$$

Poichè le variabili sono indipendenti, è possibile riscrivere $P((X_1, \dots, X_n) \in A_n)$ come segue:

$$\begin{aligned} P((X_1, \dots, X_n) \in A_n) &= \sum_{(x_1, \dots, x_n) \in A_n} P((X_1, \dots, X_n) = (x_1, \dots, x_n)) \\ &= \sum_{(x_1, \dots, x_n) \in A_n} p_X(x_1) \cdots p_X(x_n) \end{aligned}$$

Per come è stato definito A_n , vale $p_X(x_1) \cdots p_X(x_n) \geq e^{-n(H_X + \varepsilon)}$, $\forall (x_1, \dots, x_n) \in A_n$. Allora

$$1 \geq \sum_{(x_1, \dots, x_n) \in A_n} e^{-n(H_X + \varepsilon)} = |A_n| e^{-n(H_X + \varepsilon)},$$

quindi $|A_n| \leq e^{n(H_X + \varepsilon)}$. Si indichi ora con $Y_i := -\log p_X(X_i)$, che risulta essere una successione di variabili aleatorie indipendenti e identicamente distribuite di media H_X . Si osserva che:

$$\begin{aligned} P((X_1, \dots, X_n) \in A_n) &= P\left(H_X - \varepsilon \leq -\frac{1}{n} \sum_{i=1}^n Y_i \leq H_X + \varepsilon\right) \\ &= P\left(\left|\frac{1}{n} \sum_{i=1}^n Y_i - H_X\right| \leq \varepsilon\right), \end{aligned}$$

quindi applicando la *Legge dei Grandi Numeri* si conclude che:

$$P((X_1, \dots, X_n) \notin A_n) = P\left(\left|\frac{1}{n} \sum_{i=1}^n Y_i - H_X\right| \geq \varepsilon\right) \rightarrow 0.$$

Sia data una famiglia $B_n \subseteq F^n$, $n \in \mathbb{N}$ e sia $\varepsilon > 0$. Si definisca $\bar{A}_n := A_n^{\varepsilon/2}$ e si scriva:

$$P((X_1, \dots, X_n) \in B_n) = P((X_1, \dots, X_n) \in B_n \cap \bar{A}_n) + P((X_1, \dots, X_n) \in B_n \cap \bar{A}_n^c).$$

Ora, $P((X_1, \dots, X_n) \in B_n \cap \bar{A}_n^c) \leq P((X_1, \dots, X_n) \in \bar{A}_n^c)$ e sempre per la *Legge dei Grandi Numeri* $P((X_1, \dots, X_n) \notin \bar{A}_n) \rightarrow 0$. Inoltre è possibile stimare:

$$\begin{aligned} P((X_1, \dots, X_n) \in B_n \cap \bar{A}_n) &= \sum_{(x_1, \dots, x_n) \in B_n \cap \bar{A}_n} p_X(x_1) \cdots p_X(x_n) \\ &\leq \sum_{(x_1, \dots, x_n) \in B_n \cap \bar{A}_n} e^{(-n(H_X - \varepsilon/2))} \\ &= |B_n \cap \bar{A}_n| e^{(-n(H_X - \varepsilon/2))}. \end{aligned}$$

Infine, supponendo che $|B_n| \leq e^{n(H_X - \varepsilon)}$ si trova:

$$P((X_1, \dots, X_n) \in B_n \cap \bar{A}_n) \leq |B_n| e^{(-n(H_X - \varepsilon/2))} \leq e^{-\frac{1}{2}n\varepsilon} \rightarrow 0$$

e ciò conclude la dimostrazione. □

Il *Teorema di equipartizione asintotica* suggerisce che preso comunque un esito di un esperimento aleatorio realizzato eseguendo tante prove ripetute ed indipendenti, questo appartiene con probabilità asintoticamente pari ad uno ad un sottoinsieme dello spazio prodotto la cui cardinalità è esponenziale nell'entropia di una singola prova. Questo è un risultato particolarmente utile nella *teoria dell'informazione* che permette di sviluppare delle tecniche per comprimere dati senza perdere informazioni.

Bibliografia

- [1] Jakob Bernoulli. *Ars Conjectandi*, opus posthumum, 1713.
- [2] Francesco Caravenna, Paolo Dai Pra, and Quentin Berger. *Probabilità: Un'introduzione attraverso modelli e applicazioni*. Springer, 2021.
- [3] Gerolamo Cardano. *Liber de ludo aleae*, 1663.
- [4] Rick Durrett. *Probability: theory and examples*, volume 49. Cambridge University press, 2019.
- [5] Hans-Otto Georgii. Stochastics. In *Stochastics*. de Gruyter, 2012.
- [6] University of Massachusetts Amherst. The Law of Large Numbers and the Monte Carlo Method. http://people.math.umass.edu/~lr7q/ps_files/teaching/math456/lecture17.pdf, 2014.
- [7] Enrico Rivari. *La mente di Girolamo Cardano*. Zanichelli, 1906.
- [8] Sheldon M Ross. *Introduction to Probability Models, ISE*. Academic press, 2006.
- [9] Riccardo Rosso. *Storia della Probabilità*. Università di Pavia, 2022.
- [10] Kelly Sedor. The Law of Large Number and its Application, 2015. Lakehead University.
- [11] Oscar Sheynin. Chebyshev's Lecture on the Theory of Probability. In *Archive for History of Exact Sciences*, volume 46. 1994.
- [12] Presh Talwalkar. Buffon's needle problem. <https://mindyourdecisions.com/blog/2016/03/13/buffons-needle-problem-sunday-puzzle/>, 2016.