CALCOLO DELLE PROBABILITA' — PARTE SECONDA

5. SPAZI DI PROBABILITA' e VARIABILI ALEATORIE II.

Con gli strumenti introdotti nella prima parte siamo in grado di costruire modelli matematici di esperimenti con un numero finito (o, in certi casi, numerabile) di risultati possibili. Uno dei modelli che abbiamo considerato è quello di un esperimento consistente in un numero n di lanci di un dado o di una moneta. In quest'ultimo caso, il numero di possibili risultati è N^n, se N è il numero di facce del dado o della moneta. Consideriamo ora il problema di costruire un modello probabilistico per un esperimento (ideale) in cui vi sia un numero infinito di lanci, ossia consideriamo lo spazio delle sequenze semi-infinite

$$\Omega = \Omega_\infty = \{\omega : \omega = \omega_1\omega_2\dots,\ \omega_i \in \{0, 1, \dots, N-1\}\} \tag{5.1}$$

Quanti sono in questo caso i possibili risultati? In altre parole qual'è la cardinalità di Ω? Osserviamo a questo proposito che al punto (2.11) per mezzo dello sviluppo N-adico abbiamo stabilito una corrispondenza biunivoca tra i punti di Ω e i numeri dell'intervallo $(0, 1]$. Dunque Ω ha la cardinalità del continuo. In particolare Ω non è numerabile e dunque la probabilità $p(\omega)$ di ciascun suo elemento deve essere uguale a zero. In particolare, perde di senso la possibilità di determinare la probabilità di un evento A per mezzo della formula $\mathrm{P}(A) = \sum_{\omega\in A} p(\omega)$. Vediamo così come in questo caso dobbiamo, fin dall'inizio, assegnare la probabilità non più ad eventi elementari ma a sottoinsiemi di Ω, cioè ad eventi nel senso più generale di elementi di una σ-algebra $\mathcal{F}$. La definizione di spazio di probabilità $(\Omega, \mathcal{F}, \mathrm{P})$ che abbiamo data in (1.5) acquisisce ora tutto il suo significato. La misura di probabilità P dovrà soddisfare le proprietà di positività, normalizzazione e σ-additività. E le condizioni perchè ciò avvenga sono state stabilite nel Teorema (1.4).

(5.2) Esempio. Lo spazio di probabilità $([0,1],\mathcal{B},\lambda)$**.**

Dati $-\infty \le a < b < \infty$ poniamo

$$(a,b] = \{x \in I\!\!R \,:\, a < x \le b\}.$$

L'intervallo $(a,\infty]$ sarà identificato con (a,∞), il complementare di $(-\infty,a]$. Sia $\mathcal{A}$ l'insieme dei sottoinsiemi di $I\!\!R$ formati da somme finite di intervalli disgiunti della forma $(a,b]$:

$$A \in \mathcal{A} \quad \text{se} \quad A = \cup_{i=1}^m (a_i,b_i], \quad m < \infty.$$

Se includiamo in $\mathcal{A}$ anche l'insieme vuoto $\emptyset$, allora è facile verificare che $\mathcal{A}$ forma un'algebra. Tuttavia non è una σ-algebra. Infatti se $A_n = (0, 1-1/n]$ allora $A_n \in \mathcal{A}$ ma $\cup_n A_n = (0,1) \notin \mathcal{A}$. Possiamo tuttavia costruire una σ-algebra per mezzo di un processo di limite monotòno. Diremo innanzitutto che una collezione $\mathcal{M}$ di sottoinsiemi di $I\!\!R$ forma una *classe monotòna* se per ogni sequenza monotòna A_n di sottoinsiemi di $\mathcal{M}$ che converge ad A, cioè tale che $A_n \uparrow A$ oppure $A_n \downarrow A$ (nel primo caso $A_1 \subseteq A_2 \subseteq \cdots$ e $A = \cup_n A_n$, nel secondo $A_1 \supseteq A_2 \supseteq \cdots$ e $A = \cap_n A_n$) si ha $A \in \mathcal{M}$. La σ-algebra $\mathcal{B}(I\!\!R) = \sigma(\mathcal{A})$ generata da $\mathcal{A}$, detta σ-algebra di Borel, è la più piccola classe monotòna che contiene $\mathcal{A}$. Osserviamo che

$$(a,b) = \cup_{n=1}^\infty \left(a, b-\frac{1}{n}\right], \quad a < b,$$
$$[a,b] = \cap_{n=1}^\infty \left(a-\frac{1}{n}, b\right], \quad a < b,$$
$$\{a\} = \cap_{n=1}^\infty \left(a-\frac{1}{n}, a\right].$$

Pertanto $\mathcal{B}(I\!\!R)$ contiene non solo intervalli $(a,b]$ ma anche punti isolati $\{a\}$ e tutti gli insiemi della forma (a,b), $[a,b]$, $(-\infty,b)$, $(-\infty,b]$, (a,∞). Gli elementi di $\mathcal{B}(I\!\!R)$ si dicono *boreliani*. Sia inoltre λ la *misura di Lebesgue*, cioè la misura su $I\!\!R$ che ad un intervallo $I \subseteq I\!\!R$ della forma $I = (a,b]$, $I = (a,b)$, $I = [a,b]$ o $I = [a,b)$ assegna la misura $\lambda(I) = |I| = b-a$. È facile verificare che λ soddisfa le condizioni richieste nel Teorema (1.4) sugli elementi di $\mathcal{A}$, cioè che è σ-additiva su $(I\!R, \mathcal{A})$. Per un teorema generale di teoria della misura[1] λ si estende in modo unico a $(I\!R, \mathcal{B}(I\!R))$, che diviene in tal modo uno *spazio di misura*. Poniamo ora

$$\mathcal{B}([0,1]) = \{A \cap [0,1] \,:\, A \in \mathcal{B}(I\!R)\,\}.$$

Essendo $\lambda([0,1]) = 1$, la restizione di λ a $\mathcal{B}([0,1])$ (che denoteremo ancora con λ) è una misura di probabilità. D'altra parte, oltre agli elementi di $\mathcal{B}([0,1])$ è talvolta necessario considerare i cosiddetti *sottoinsiemi misurabili* (secondo Lebesgue) di $[0,1]$. Questi sono definiti come segue: diremo che $C \in \mathcal{B} \supseteq \mathcal{B}([0,1])$ se esistono due insiemi boreliani A e B tali che $A \subseteq C \subseteq B$ e $\lambda(B \setminus A) = 0$. La misura λ può essere estesa a $\mathcal{B}$ in modo naturale: dati A, B, C come sopra poniamo $\lambda(C) = \lambda(A)$ (dove abbiamo usato lo stesso simbolo λ per indicare la misura di Lebesgue su $\mathcal{B}([0,1])$ e la sua estensione a $\mathcal{B}$). Per quanto visto, la tripletta $([0,1], \mathcal{B}, \lambda)$ forma uno spazio di probabilità nel senso di Kolmogorov.

(5.3) Esempio. Continuazione di (2.11). Riprendiamo l'Esempio (2.11). Abbiamo visto che l'applicazione (2.12) stabilisce una corrispondenza $\pi : \Omega_\infty \to [0,1]$ (che può facilmente essere resa biunivoca) tra lo spazio Ω_∞ delle sequenze infinite $\omega = \omega_1\omega_2\dots$, $\omega_i \in X = \{0, 1, \dots N-1\}$ e i punti di $[0,1]$. D'altra parte la misura P_n data in (2.16) è definita sulla σ-algebra $\mathcal{F}_n$ generata dagli insiemi cilindrici di dimensione $\le n$. Consideriamo ora lo spazio di probabilità $(\Omega, \mathcal{F}, \mathrm{P})$ dove $\Omega \equiv \Omega_\infty$ mentre $\mathcal{F}$ e P soddisfano i seguenti requisiti:

i) $\mathcal{F}$ contiene tutti gli insiemi cilindrici di dimensione finita, e si ha $\mathrm{P}(C) = \mathrm{P}_n(C)$ se $C \subset \Omega_n$ per qualche n;

ii) se C_n è una sequenza monotòna di insiemi in $\mathcal{F}$ allora $\lim_n C_n \in \mathcal{F}$ e $\mathrm{P}(\lim_n C_n) = \lim_n \mathrm{P}(C_n)$.

[1] Si tratta del *teorema di Carathéodory*: Sia Ω uno spazio, $\mathcal{A}$ un'algebra di suoi sottoinsiemi e $\sigma(\mathcal{A})$ la più piccola σ-algebra che contiene $\mathcal{A}$. Sia μ_0 una misura σ-additiva su $(\Omega, \mathcal{A})$. Allora esiste un'unica misura μ su $(\Omega, \sigma(\mathcal{A}))$ che estende μ_0, cioè tale che $\mu(A) = \mu_0(A)$ per ogni $A \in \mathcal{A}$.

Dal teorema (1.4) segue allora che P è σ-additiva su $\mathcal{F}$ e da quanto visto si deduce facilmente che $(\Omega, \mathcal{F}, \mathrm{P})$ è isomorfo, via la mappa π, allo spazio di probabilità $([0,1], \mathcal{B}, \lambda)$ considerato in (5.2). Prendiamo per semplicità $N = 2$, anche se l'argomento può essere facilmente esteso ad ogni $2 \leq N < \infty$. La (2.12) diviene allora uno sviluppo *diadico* o *binario*. Dato un numero $0 \leq p \leq 1$, possiamo chiederci quale sia la misura dell'insieme delle sequenze infinite $\omega = \omega_1\omega_2\ldots$, $\omega_i \in \{0,1\}$, per le quali il limite di $f_n^{(1)} = \nu_n^{(1)}/n$ esista e sia eguale a p. In modo equivalente, quale sia la misura (di Lebesgue) dell'insieme dei numeri $x \in [0,1]$ in cui le cifre $0, 1$ hanno una frequenza relativa limite rispettivamente $1-p$ ed p nel loro sviluppo diadico. Introduciamo a questo proposito la variabile aleatoria

$$s_n(\omega) = \sum_{i=1}^{n} r_i(\omega), \qquad r_i(\omega) = 2 \cdot 1_i^{(1)}(\omega) - 1 \tag{5.4}$$

dove le $1_i^{(1)}$ sono definite in (2.4) (qui $x_h \equiv 1$). Le funzioni r_i assumono valori in $\{-1, 1\}$ e si ha

$$\lim_n f_n^{(1)} = f \Longleftrightarrow \lim_n \frac{s_n(\omega)}{n} = 2p - 1. \tag{5.5}$$

Usando la prima disuguaglianza di Chebyshev (vedi più avanti, (5.44)-VI) con $\xi = s_n^4$ e $\alpha = n^4\epsilon^4$ otteniamo, con leggero abuso di notazione (vedi (1.53)-(1.54)),

$$\mathrm{P}(|s_n| > n\epsilon) = \mathrm{P}(s_n^4 > n^4\epsilon^4) \leq \frac{1}{n^4\epsilon^4} \int_0^1 s_n^4(x)\, dx \tag{5.6}$$

Osserviamo ora che l'integrando in (5.6) ha la forma

$$s_n^4 = \sum_{i,j,k,l} r_i\, r_j\, r_k\, r_l \tag{5.7}$$

dove gli indici i, j, k, l variano in modo indipendentente tra 1 e n. I vari termini si possono classificare nelle cinque classi seguenti, all'interno di ciascuna delle quali gli indici sono ora da considerare distinti:

$$\begin{cases} r_i^4 = 1, \\ r_i^2\, r_j^2 = 1, \\ r_i^2\, r_j\, r_k = r_j\, r_k, \\ r_i^3 r_j = r_i\, r_j, \\ r_i\, r_j\, r_k\, r_l. \end{cases}$$

Non è troppo difficile rendersi conto che l'integrale delle ultime tre classi è uguale a zero. Consideriamo ad esempio l'ultima e supponiamo che $l > i, j, k$. Allora l'integrale su $[0, 1]$ si può decomporre in 2^{l-1} integrali, ciascuno su di un intervallo diadico $(l-1)$-dimensionale del tipo $I_{a_1,\ldots,a_{l-1}}$. Ma su ciascuno di tali intervalli, il prodotto $r_i\, r_j\, r_k$ è costante, mentre la funzione r_l assume valore -1 nella metà di sinistra, $+1$ nella metà di destra. Ciò detto, vi sono n termini nella prima classe e $3n(n-1)$ termini nella seconda. Infatti, vi sono n possibili scelte per i in (5.7), a cui corrispondono tre possibili identificazioni, con j, con k o con l, e $n-1$ possibilità per il valore comune dei due rimanenti indici. Integrando la (5.7) termine a termine otteniamo quindi

$$\int_0^1 s_n^4(x)\, dx = n + 3n(n-1) \leq 3n^2. \tag{5.8}$$

Mettendo insieme (5.6) e (5.8) otteniamo

$$\mathrm{P}\left(\left|\frac{s_n}{n}\right| > \epsilon\right) \leq \frac{3}{n^2 \epsilon^4}. \tag{5.9}$$

Prendiamo ora una sequenza $\epsilon_n \to 0$ tale che $\sum_n \epsilon_n^{-4} n^{-2} < \infty$, ad esempio $\epsilon_n = n^{-1/8}$, e poniamo

$$C_n = \cup_{k \geq n} \{|s_k/k| > \epsilon_k\} \tag{5.10}$$

allora si ha $C_n \supseteq C_{n+1}$ e inoltre

$$C = \lim_n C_n = \left\{\omega \,:\, \overline{\lim}\, |s_n(\omega)/n| > 0\right\} \tag{5.11}$$

Usando la proprietà (ii) e la σ-additività della misura P, insieme alla (5.9), troviamo

$$\mathrm{P}(C) = \mathrm{P}(\lim_n C_n) = \lim_n \mathrm{P}(C_n) \leq \lim_n \left(3 \sum_{k \geq n} \frac{1}{k^2 \epsilon_k^4}\right) = 0. \tag{5.12}$$

Riassumendo, abbiamo trovato che, dati p e $q = 1-p$ e posto

$$M(p,q) := \{x \in (0,1] \,:\, x = \pi(\omega),\ \lim_n f_n^{(1)}(\omega) = p\} \tag{5.13}$$

si ha

$$(5.14)\qquad \lambda\left(M(\frac{1}{2},\frac{1}{2})\right)=1,\qquad \lambda\left(M(p,q)\right)=0,\quad p,q\neq 1/2.$$

I numeri $x\in M(\frac{1}{2},\frac{1}{2})$ si dicono numeri *normali* nel senso di Borel. Dunque la (5.14) afferma che l'insieme dei numeri normali ha complemento di misura nulla. In modo più pittoresco: se scegliamo 'a caso' un numero in $[0,1]$, cioè se lo 'estraiamo' usando la distribuzione uniforme su $[0,1]$, allora questo sarà certamente un numero normale. Si tratta di un caso particolare della 'legge forte dei grandi numeri', in opposizione alla 'legge debole' vista al punto (2.8). I due tipi di leggi stabiliscono due tipi diversi di convergenza delle medie artimetiche in gioco e dunque due tipi diversi di teorema ergodico. In virtù delle (2.17)-(2.18) si ha

$$(5.15)\qquad \lim_n \mathrm{P}_n\left(\{\omega\,:\,|f_n^{(1)}(\omega)-1/2|\leq\delta\}\right)=1$$

che rappresenta la convergenza in probabilità delle medie aritmetiche $f_n^{(1)}$ al valore $1/2$. D'altra parte, dalle (5.13)-(5.14) si ha

$$(5.16)\qquad \mathrm{P}\left(\{\omega\,:\,\lim_n f_n^{(1)}(\omega)=1/2\}\right)=1$$

che si legge: "l'insieme delle sequenze infinite $\omega\in\Omega$ per le quali il limite $\lim_n f_n^{(1)}(\omega)$ esiste ed è uguale a $1/2$, ha probabilità P uguale a uno". Tale convergenza si chiama *convergenza* P *-quasi ovunque*. La (5.16) può essere generalizzata in più modi. Innanzitutto si possono considerare, come già visto, sviluppi N-adici con $N\geq 2$. Inoltre, si possono considerare distribuzioni arbitrarie sull'alfabeto X. Se $p=(p_0,\ldots,p_{N-1})$ denota una tale distribuzione e $\xi_1,\xi_2,\ldots$ indica la successione di variabili aleatorie definite da $\xi_i(\omega)=\omega_i$, allora possiamo considerare la misura prodotto P su $\mathcal{F}$ definita da

$$(5.17)\qquad \mathrm{P}\left(\{\omega\,:\,\xi_i(\omega)=h\}\right)=p_h,\quad h=0,\ldots,N-1.$$

Allora, un'immediata estensione della (5.16) si ottiene come segue: poniamo

$$(5.18)\qquad M(p_0,\ldots,p_{N-1})=\{\omega\,:\,\lim_n f_n^{(h)}(\omega)=p_h,\ h=0,\ldots,N-1\}$$

allora si ha

$$\mathrm{P}\,(M(p_0, \ldots, p_{N-1})) = 1. \tag{5.19}$$

Usando un linguaggio improprio ma abbastanza evocativo, possiamo dire che le 'leggi forti' riguardano misure in spazi di dimensione infinita, mentre le 'leggi deboli' vengono formulate in termini di spazi finito-dimensionali di dimensione crescente. Si potrebbe a questo punto obiettare che, dal punto di vista delle applicazioni, l'utilità di leggi sul comportamento di sequenza *infinite* è per lo meno discutibile, mentre le 'leggi deboli' assumono evidentemente un interesse primario. D'altra parte, la trattazione del caso idealizzato infinito presenta alcuni vantaggi, tra cui quello di rendere più agevole l'introduzione di nozioni asintotiche quali: ergodicità, complessità, entropia etc. alla base della *teoria ergodica*, cioè di quella disciplina che si colloca tra la teoria della probabilità, la teoria dell'informazione e la teoria dei sistemi dinamici, e che studia le proprietà di trasformazioni che preservano la struttura di uno spazio di probabilità. Più precisamente, sullo spazio $(\Omega, \mathcal{F}, \mathrm{P}\,)$ considerato più sopra, con P definita in (5.17), possiamo introdurre una trasformazione $T : \Omega \to \Omega$ detta *traslazione*, definita da $(T\omega)_i = \omega_{i+1}$. In modo equivalente, se $\{\xi_i\}$ è la successione di variabili aleatorie definita più sopra allora si ha $\xi_i(T\omega) = \xi_{i+1}(\omega)$ e dunque $\xi_i(T^k\omega) = \xi_{i+k}(\omega)$. Ora, se C è un cilindro della forma $C^{i_1, \ldots, i_n}_{a_1, \ldots, a_n}$, allora

$$T^{-1}C = \{\omega : T\omega \in C\} = \cup_{h=0}^{N-1}\{\omega \,:\, \omega_{i_1} = h,\ \omega_{i_{j+1}} = a_j,\ j = 1, \ldots, n\,\} \tag{5.20}$$

e dunque anche $T^{-1}C$ è misurabile e si ha $\mathrm{P}\,(T^{-1}C) = \mathrm{P}\,(C)$. Queste proprietà sono facilmente estendibili ad ogni elemento della σ-algebra $\mathcal{F}$. Si dice allora che la trasformazione T definita su $(\Omega, \mathcal{F}, \mathrm{P}\,)$ *preserva la misura*. Risulta abbastanza naturale chiedersi a cosa corrisponda l'azione di T su $[0, 1]$, ossia che cosa sia $\pi(T(\omega))$. La risposta giunge immediatamente dalla (2.12):

$$\pi(T(\omega)) = \omega_2 \cdot N^{-1} + \omega_3 \cdot N^{-2} + \cdots = N \cdot x \,(\mathrm{mod}\,1) \tag{5.21}$$

In altre parole vale la corrispondenza

$$\pi \circ T = F \circ \pi, \qquad F(x) = N \cdot x \,(\mathrm{mod}\, 1). \tag{5.22}$$

Diremo a tal proposito che $T : \Omega \to \Omega$ e $F : [0,1] \to [0,1]$ sono *coniugate*. La trasformazione F preserva lo spazio di probabilità $((0,1], \mathcal{B}, \mu)$ dove μ è la misura su $\mathcal{B}$ corrispondente, via la π, alla misura P su $\mathcal{F}$. In particolare se P è la distribuzione uniforme allora $\mu = \lambda$. Osserviamo inoltre che se indichiamo ancora con $M(p_0, \ldots, p_{N-1}) \subset [0,1]$ l'immagine dell'insieme (5.18) rispetto alla mappa π, e se $p \neq (1/N, \ldots, 1/N)$, allora si ha

$$\mu\,(M(p_0, \ldots, p_{N-1})) = 1 \quad \text{ma} \quad \lambda\,(M(p_0, \ldots, p_{N-1})) = 0. \tag{5.23}$$

In altre parole le misure μ, costituiscono, al variare delle vettore p, una famiglia (non numerabile) di misure di probabilità invarianti per l'azione di F e mutuamente singolari. L'iterazione di F genera un *sistema dinamico* che ad $x \in [0,1]$ associa la sua *orbita* $x, F(x), F^2(x), \ldots$ dove $F^n = F \circ \cdots \circ F$ per n volte. Ciò determina un meccanismo per *generare* sequenze simboliche: suddividiamo l'intervallo $(0,1]$ nei sottointervalli $I_h = (h/N, (h+1)/N]$, $h = 0, \ldots, N-1$, e associamo all'orbita di $x \neq 0$ la sequenza $\omega = \omega_0\omega_1 \cdots$ determinata da $x_n = F^n(x) \in I_{\omega_n}$, $n \geq 0$. E' immediato verificare che la sequenza in tal modo generata altro non è che $\omega = \pi^{-1}(x)$. Vale osservare che proprio questo è il metodo con cui si producono nei calcolatori successioni di 'numeri casuali': si considera una trasformazione dell'intervallo unitario in sé, ad esempio la (5.22) con $N = 2$, e 'scelto a caso' un numero $x_0 \in (0,1]$, il 'seme', si definisce l'n-esimo simbolo della successione uguale a 0 se $x_n \in (0, 1/2]$, uguale a 1 se $x_n \in (1/2, 1]$. In virtù della legge forte dei grandi numeri (per la misura di Lebesgue) avremo buone speranze di ottenere una sequenza a tutti gli effetti 'indistinguibile' dalla sequenza di lanci di una moneta.

(5.24) Variabili aleatorie e funzioni di distribuzione.

Un'estensione diretta dell'Esempio discusso in (5.2) si ottiene prendendo come spazio di probabilità la tripletta $(\Omega, \mathcal{F}, \mathrm{P}\,)$ dove Ω è uno spazio metrico completo e separabile e $\mathcal{F}$ la σ-algebra di Borel, ossia la più piccola σ-algebra che contiene tutte le bocce aperte[1], cioè i sottoinsiemi della forma $B_r(\omega_0) = \{\omega, d(\omega, \omega_0) < r\}$. Tale σ-algebra sarà denotata con $\mathcal{B}(\Omega)$. Possiamo a questo punto dare un significato più preciso alla definizione di v.a. come funzione misurabile.

(5.25) Definizione. Sia $(\Omega, \mathcal{F}, \mathrm{P}\,)$ uno spazio di probabilità. Un funzione $\xi : \Omega \to I\!R$ si dice *variabile aleatoria* se è *misurabile* rispetto a $\mathcal{F}$, cioè se per ogni $B \in \mathcal{B}(I\!R)$ si ha $\xi^{-1}(B) = \{\omega : \xi(\omega) \in B\} \in \mathcal{F}$. Una funzione ξ come sopra a valori in $\overline{I\!R} = [-\infty, \infty]$ si dice *variabile aleatoria estesa.*

(5.26) Osservazione. La necessità di considerare variabili aleatorie estese emerge in particolare quando si voglia costruire nuove funzioni a partire da una data collezione $\{\xi_n\}$ di variabili aleatorie, come ad esempio $\sum_{n=1}^{\infty} |\xi_n|$, $\overline{\lim}\,\xi_n$, $\underline{\lim}\,\xi_n$, etc.

Osserviamo ora che:

$$(5.27) \qquad \xi^{-1}(I\!R) = \Omega, \quad \xi^{-1}(I\!R \setminus B) = \Omega \setminus \xi^{-1}(B), \quad \xi^{-1}(\cup_i B_i) = \cup_i \xi^{-1}(B).$$

In altri termini, la classe dei sottoinsiemi di Ω della forma $\xi^{-1}(B)$, con $B \in \mathcal{B}(I\!R)$ forma una σ-sottoalgebra di $\mathcal{F}$. Generalizzando la (prima parte della) definizione (3.18), diremo che tale σ-sottoalgebra è la σ-algebra *generata* da ξ, e la indicheremo con $\mathcal{F}_\xi$. Poniamo inoltre $\mathrm{P}_{\,\xi}(B) = \mathrm{P}\,(\xi^{-1}(B))$ per ogni $B \in \mathcal{B}(I\!R)$. È facile verificare che $\mathrm{P}_{\,\xi}$ è una misura di probabilità su $\mathcal{B}(I\!R)$.

(5.28) Definizione. $\mathrm{P}_{\,\xi}$ di dice *distribuzione di probabilità della v.a.* ξ. La funzione $F_\xi(t)$, $t \in I\!R$, definita da $F_\xi(t) = \mathrm{P}_{\,\xi}(B_t)$ dove $B_t = (-\infty, t]$, si dice *funzione distribuzione* della v.a. ξ.

[1] Il caso discusso al punto (5.3) si riduce a questo se su Ω mettiamo la metrica $d(\omega, \omega') = \theta^{\min\{i \,:\, \omega_i \neq \omega'_i\}}$ con $0 < \theta < 1$ opportuno.

(5.29) Alcune proprietà della funzione distribuzione.

1) $F_\xi(t)$ è monotòna non-decrescente: $F_\xi(t_1) \leq F_\xi(t_2)$ per $t_1 < t_2$. Ciò segue immediatamente dal fatto che $B_{t_1} \subset B_{t_2}$ e dalla (1.3.i).

2) $\lim_{t\to\infty} F_\xi(t) = 1$ e $\lim_{t\to-\infty} F_\xi(t) = 0$. Infatti, sia $\{t_i\}$ una sequenza monotòna con $t_i \to \infty$ quando $i \to \infty$, cosicchè $\{B_{t_i}\}$ forma una sequenza monotòna di eventi con $\cup_i B_{t_i} = \mathbb{R}$. In particolare ciò implica che $\{\xi^{-1}(B_{t_i})\}$ è anch'essa monotòna e per (5.27) si ha $\cup_i \xi^{-1}(B_{t_i}) = \Omega$. Da (1.4) segue allora che

$$\lim_{i\to\infty} F_\xi(t_i) = \lim_{i\to\infty} \mathrm{P}_\xi(B_{t_i}) = \lim_{i\to\infty} \mathrm{P}\,(\xi^{-1}(B_{t_i})) = \mathrm{P}\,(\Omega) = 1.$$

In modo analogo, usando (1.4.3), si mostra l'altro limite.

3) $F_\xi(t)$ è continua da destra. In effetti, sia $t_i \downarrow t^*$. Allora $\{B_{t_i}\}$ è una sequenza decrescente con $\cap_i B_{t_i} = B_{t^*}$. Pertanto, da (1.4) si ha

$$\lim_{i\to\infty} F_\xi(t_i) = \lim_{i\to\infty} \mathrm{P}\,(\xi^{-1}(B_{t_i})) = \mathrm{P}\,(\cap \xi^{-1}(B_{t_i})) = F_\xi(t^*).$$

4) Per ogni t^* la probabilità $\mathrm{P}\,(\xi = t^*)$ si può calcolare come il limite $P(\xi = t^*) = F_\xi(t^*) - \lim_{i\to\infty} F_\xi(t_i)$ per ogni sequenza $\{t_i\}$ tale che $t_i \uparrow t^*$. Infatti si ha $\cup_i B_{t_i} = (-\infty, t^*)$ e quindi

$$\lim_{i\to\infty} F_\xi(t_i) = \mathrm{P}\,(\xi < t^*),$$

e

$$F_\xi(t^*) - \lim_{i\to\infty} F_\xi(t_i) = \mathrm{P}\,(\xi \leq t^*) - \mathrm{P}\,(\xi < t^*) = \mathrm{P}\,(\xi = t^*).$$

Da questa proprietà segue che $\mathrm{P}\,(\xi = t^*) = 0$ se e solo se $F_\xi(t)$ è continua nel punto t^*. Osserviamo inoltre che per ogni intervallo semi aperto $(t_1, t_2]$ si ha

(5.30)
$$\begin{aligned}\mathrm{P}_\xi((t_1,t_2]) &= \mathrm{P}\,(\{\omega | t_1 < \xi(\omega) \leq t_2\}) \\ &= \mathrm{P}\,(\{\omega | \xi(\omega) \leq t_2\}) - \mathrm{P}\,(\{\omega | \xi(\omega) \leq t_1\}) = F_\xi(t_2) - F_\xi(t_1).\end{aligned}$$

Pertanto la funzione distribuzione $F_\xi(t)$ determina univocamente i valori di P_ξ sugli intervalli semiaperti, e per il teorema di Carathéodory ciò determina univocamente la misura P_ξ su $\mathcal{B}(\mathbb{R})$.

(5.31) Tipi di funzioni di distribuzione. Ogni funzione di distribuzione si può rappresentare in modo unico come somma di funzioni di distribuzione dei tre tipi seguenti.

Tipo discreto. Supponiamo che ξ assuma un insieme finito o al più numerabile di valori $x_1, x_2 \ldots$ con probabilità $p_1, p_2, \ldots$. In particolare, i valori $\{x_i\}$ si possono mettere in corrispondenza con gli elementi di una partizione $\{D_i\}$ di Ω in modo tale che $\xi(\omega) = \sum_i x_i 1_{D_i}(\omega)$. Variabili aleatorie di questo tipo sono state studiate nella Prima Parte e verrano qui chiamate *semplici*. In questo caso $F_\xi(t) = \sum_{i\,:\,x_i \le t} p_i$, cioè una funzione a gradini, dunque discontinua.
Tipo assolutamente continuo. Anche questo questo caso è già stato discusso nella Prima Parte (seppure in forma euristica), e si verifica ogni volta che esista una funzione *densità* $p_\xi(t) \ge 0$ tale che $F_\xi(t) = \int_{-\infty}^t p_\xi(s)ds$.
Tipo singolare continuo. In questo caso $F_\xi(t)$ è continua ma non può essere rappresentata come un integrale di una funzione densità. Un esempio classico si costruisce come segue: poniamo innanzitutto $F_\xi(0) = 0$ e $F_\xi(1) = 1$. Dividiamo l'intervallo $[0, 1]$ in tre parti uguali e poniamo $F_\xi(t) = (0 + 1)/2 = 1/2$ su $[1/3, 2/3]$, mentre a questo stadio resta ancora non definita sui due intervalli aperti $(0, 1/3)$ e $(2/3, 1)$. A questo punto procediamo iterativamente: al passo n-esimo avremo 2^n intervallini aperti di lunghezza 3^{-n} su cui $F_\xi(t)$ non è ancora definita, mentre è definita sui loro punti di frontiera. Dividiamo ciascuno di questi intervallini in tre parti uguali e poniamo sull'intervallino centrale (chiuso) $F_\xi(t) =$ costante, uguale alla media artimetica dei due valori assunti sui suddetti punti di frontiera. La funzione limite si chiama *scala del diavolo*.
Vedremo ora che una qualsiasi v.a. (la cui funzione di distribuzione sarà in generale una sovrapposizione dei tre tipi sopra elencati) può essere costruita come limite di v.a. semplici (la cui funzione di distribuzione è per definizione discreta).

(5.32) Teorema. Per ogni v.a. (incluse le v.a. estese) esiste una sequenza si v.a. semplici $\xi_1, \xi_2 \ldots$ tale che $|\xi_n| \le |\xi|$ e $\xi_n(\omega) \to \xi(\omega)$, $n \to \infty$, per ogni

$\omega \in \Omega$. Se inoltre $\xi(\omega) \geq 0$ allora la sequenza ξ_n può essere scelta in modo che $\xi_n(\omega) \uparrow \xi(\omega)$, $n \to \infty$, per ogni $\omega \in \Omega$.

Dimostrazione. Mostriamo prima la seconda affermazione. Poniamo a questo scopo

$$\xi_n(\omega) = \sum_{k=0}^{n \cdot 2^n - 1} \frac{k}{2^n} 1_{\{k \leq 2^n \cdot \xi(\omega) < (k+1)\}}(\omega) + n\, 1_{\{\xi(\omega) \geq n\}}(\omega).$$

È immediato verificare che $\xi_n(\omega) \uparrow \xi(\omega)$ per ogni $\omega \in \Omega$. La prima affermazione segue ora dalla seconda e dal fatto che un'arbitraria ξ può essere rappresentata nella forma $\xi = \xi^+ - \xi^-$ con $\xi^+ = \xi \cdot 1_{\{\omega : \xi(\omega) \geq 0\}}$ e $\xi^- = -\,\xi \cdot 1_{\{\omega : \xi(\omega) < 0\}}$. q.e.d.

(5.33) Teorema. Sia $\phi : \mathbb{R} \to \mathbb{R}$ una funzione $\mathcal{B}(\mathbb{R})$-misurabile[1] e ξ una v.a. Allora la composizione $\eta = \phi \circ \xi$ è anch'essa una v.a. Inversamente, se η è una v.a. $\mathcal{F}_\xi$-misurabile allora esiste una funzione boreliana ϕ tale che $\eta = \phi \circ \xi$.

Dimostrazione. La prima parte segue subito osservando che per ogni $B \in \mathcal{B}(\mathbb{R})$

$$\{\omega : \eta(\omega) \in B\} = \{\omega : \phi(\xi(\omega)) \in B\} = \{\omega : \xi(\omega) \in \phi^{-1}(B)\} \in \mathcal{F}$$

perchè per ipotesi $\phi^{-1}(B) \in \mathcal{B}(\mathbb{R})$. Per la seconda parte, consideriamo $A \in \mathcal{F}_\xi$ e $\eta(\omega) = 1_A(\omega)$. Mostriamo che η è rappresentabile nella forma $\phi \circ \xi$ per qualche funzione boreliana ϕ. In effetti, se $A \in \mathcal{F}_\xi$ allora possiamo trovare $B \in \mathcal{B}(\mathbb{R})$ tale che $A = \{\omega : \xi(\omega) \in B\}$. Pertanto $\eta(\omega) = 1_A(\omega) = 1_B(\xi(\omega))$. La stessa cosa può evidentemente essere detta per ogni funzione semplice $\mathcal{F}_\xi$-misurabile della forma $\eta(\omega) = \sum_i y_i 1_{A_i}(\omega)$, $A_i \in \mathcal{F}_\xi$. Sia ora η un'arbitraria funzione $\mathcal{F}_\xi$-misurabile. Per (5.32) esiste una sequenza $\{\eta_n\}$ di funzioni semplici $\mathcal{F}_\xi$-misurabili tali che $\eta_n(\omega) \to \eta(\omega)$ per ogni $\omega \in \Omega$ e, per quanto visto, si può trovare una sequenza di funzioni boreliane $\{\phi_n\}$ tale che $\eta_n = \phi_n \circ \xi$ e dunque $\phi_n(\xi(\omega)) \to \eta(\omega)$. L'insieme $B = \{x \in \mathbb{R} : \lim_n \phi_n(x) \text{ esiste}\}$ è boreliano, e dunque la funzione $\phi(x) = \lim_n \phi_n(x)$ se $x \in B$, $\phi(x) = 0$ se $x \notin B$, è boreliana. D'altra parte si ha $\eta(\omega) = \lim_n \phi_n(\xi(\omega)) = \phi(\xi(\omega))$ per ogni $\omega \in \Omega$. q.e.d.

[1] Una funzione $\mathcal{B}(\mathbb{R}^n)$-misurabile per qualche $n \geq 1$ viene detta anche *funzione boreliana*.

Una immediata estensione del teorema appena dimostrato è che data una collezione di v.a. $\xi_1, \ldots, \xi_n$ e una funzione boreliana ϕ, la funzione $\eta = \phi(\xi_1, \ldots, \xi_n)$ è una v.a. $\mathcal{F}_{\xi_1 \ldots \xi_n}$-misurabile. In particolare lo è ogni combinazione lineare $\eta = \sum_i a_i \xi_i$, lo è il prodotto $\eta = \prod_i \xi_i$, e, date ξ_1, ξ_2, con $\xi_2(\omega) \neq 0$ per ogni $\omega \in \Omega$, lo è anche $\eta = \xi_1/\xi_2$. In accordo con quanto stabilito per le v.a. semplici, diamo ora la seguente definizione.

(5.34) Definizione. Le v.a. $\xi_1, \ldots, \xi_n$ si dicono *mutuamente indipendenti* se per ogni collezione $C_1, \ldots, C_k$ di elementi di $\mathcal{B}(I\!R)$ si ha

$$\mathrm{P}\left(\xi_1 \in C_1, \cdots, \xi_n \in C_n\right) = \prod_{i=1}^{n} \mathrm{P}\left(\xi_i \in C_i\right).$$

Se inoltre sono identicamente distribuite e μ indica la comune distribuzione di probabilità su $\mathcal{B}(I\!R)$, allora l'espressione scritta sopra è uguale a $\prod_{i=1}^{n} \mu(C_i)$. Sia inoltre

$$F_{\xi_1 \ldots \xi_n}(t_1, \ldots, t_n) = \mathrm{P}\left(\xi_1 \leq t_1, \ldots, \xi_n \leq t_n\right)$$

la funzione distribuzione congiunta di $\xi_1, \ldots, \xi_n$ (detta anche distribuzione n-dimensionale del *vettore aleatorio* $\xi = (\xi_1, \ldots, \xi_n)$). Ragionando come al punto (5.30) si mostra che condizione necessaria e sufficiente affinchè $\xi_1, \ldots, \xi_n$ siano mutuamente indipendenti è che per ogni $(t_1, \ldots, t_n) \in I\!R^n$ si abbia

$$F_{\xi_1 \ldots \xi_n}(t_1, \ldots, t_n) = \prod_{i=1}^{n} F_{\xi_i}(t_i). \tag{5.35}$$

(5.36) Esempio. Se ξ e η sono v.a. con funzione distribuzione congiunta $F_{\xi\eta}(x, y)$ e $\phi = \phi(x, y)$ è una funzione boreliana allora, posto $\zeta = \phi(\xi, \eta)$ si ha

$$F_\zeta(z) = \int_{\{x,y\,:\,\phi(x,y) \leq z\}} dF_{\xi\eta}(x, y).$$

Ad esempio, se $\phi(x, y) = x + y$ e ξ ed η sono indipendenti (cosicchè $F_{\xi\eta}(x, y) =$

$F_\xi(x) \cdot F_\eta(y))$, allora per il teorema di Fubini[1] si ha

$$\begin{aligned}
F_\zeta(z) &= \int_{\{x,y\,:\,x+y\le z\}} dF_\xi(x)\cdot dF_\eta(y) \\
&= \int_{\mathbb{R}^2} 1_{\{x+y\le z\}}(x,y)\, dF_\xi(x)\cdot dF_\eta(y) \\
&= \int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty} 1_{\{x+y\le z\}}(x,y)\, dF_\eta(y)\right) dF_\xi(x) = \int_{-\infty}^{\infty} F_\eta(z-x)\, dF_\xi(x)
\end{aligned}$$

e, simmetricamente,

$$F_\zeta(z) = \int_{-\infty}^{\infty} F_\xi(z-y)\, dF_\eta(y).$$

Ora, se F e G sono funzioni distribuzione, la funzione

$$H(z) = \int_{-\infty}^{\infty} F(z-x)\, dG(x)$$

si chiama *convoluzione* di F e G e si indica con $F * G$. Abbiamo così mostrato che la funzione distribuzione F_ζ della somma di due v.a. indipendenti ξ e η è la convoluzione delle rispettive funzioni distribuzione F_ξ e F_η: $F_\zeta = F_\xi * F_\eta$, dove è chiaro che $F_\xi * F_\eta = F_\eta * F_\xi$. Supponiamo ora che le distribuzioni di ξ e η abbiano densità p_ξ e p_η. Allora, usando ancora il teorema di Fubini, l'espressione

[1] (*Teorema di Fubini*) Sia $(\Omega, \mathcal{F}, \mu)$ uno spazio di misura con $\Omega = \Omega_1 \times \Omega_2$, $\mathcal{F} = \mathcal{F}_1 \otimes \mathcal{F}_2$ e $\mu = \mu_1 \times \mu_2$ il prodotto diretto di due misure μ_1 e μ_2 (cioè tale che $\mu(A \times B) = \mu_1(A)\mu_2(B)$, per ogni $A \in \mathcal{F}_1$, $B \in \mathcal{F}_2$). Sia $\xi = \xi(\omega_1, \omega_2)$ una funzione misurabile rispetto a $\mathcal{F}_1 \otimes \mathcal{F}_2$ e integrabile rispetto alla misura μ, cioè $\int_\Omega |\xi| d\mu < \infty$. Allora gli integrali $\int_{\Omega_1} \xi(\omega_1,\omega_2)\mu_1(d\omega_1)$ e $\int_{\Omega_2} \xi(\omega_1,\omega_2)\mu_2(d\omega_2)$ sono definiti per ogni ω_1 e ω_2 e si ha

$$\begin{aligned}
\int_\Omega \xi(\omega_1,\omega_2) d\mu(\omega_1,\omega_2) &= \int_{\Omega_1}\left[\int_{\Omega_2} \xi(\omega_1,\omega_2)\mu_2(d\omega_2)\right]\mu_1(d\omega_1) \\
&= \int_{\Omega_2}\left[\int_{\Omega_1} \xi(\omega_1,\omega_2)\mu_1(d\omega_1)\right]\mu_2(d\omega_2).
\end{aligned}$$

scritta sopra diviene

$$
\begin{aligned}
F_\zeta(z) &= \int_{-\infty}^{\infty} F_\eta(z-x)\, dF_\xi(x) = \int_{-\infty}^{\infty} \left(\int_{-\infty}^{z-x} p_\eta(u) du \right) p_\xi(x) dx \\
&= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{z} p_\eta(u-x) du \right) p_\xi(x) dx \\
&= \int_{-\infty}^{z} \left(\int_{-\infty}^{\infty} p_\eta(u-x) p_\xi(x) dx \right) du = \int_{-\infty}^{z} p_\zeta(u) du
\end{aligned}
$$

con

$$
p_\zeta(x) = \int_{-\infty}^{\infty} p_\eta(x-u) p_\xi(u) du = \int_{-\infty}^{\infty} p_\xi(x-v) p_\eta(v) dv.
$$

Vediamo come queste formule possono essere usate in casi concreti. Consideriamo innanzitutto due v.a ξ e η, indipendenti e identicamente distribuite con distribuzione uniforme su $[0,1]$, cioè $p_\xi(x) = p_\eta(x) = 1_{[0,1]}(x)$. Allora si trova

$$
p_{\xi+\eta}(x) = \int_0^1 1_{\{x-1 \le u \le x\}}(u)\, du = \begin{cases} x, & 0 \le x \le 1, \\ 2-x, & 1 \le x \le 2, \\ 0, & x > 2. \end{cases}
$$

Se invece si ha

$$
p_\xi(x) = \frac{1}{\sqrt{2\pi}\sigma_\xi}\, e^{-(x-m_\xi)^2/2\sigma_\xi^2}, \qquad p_\eta(x) = \frac{1}{\sqrt{2\pi}\sigma_\eta}\, e^{-(x-m_\eta)^2/2\sigma_\eta^2}
$$

allora un semplice calcolo dà

$$
p_{\xi+\eta}(x) = \frac{1}{\sqrt{\sigma_\xi^2 + \sigma_\eta^2}} e^{-(x-m_\xi-m_\eta)^2/2(\sigma_\xi^2+\sigma_\eta^2)}.
$$

In altre parole la somma di due v.a. gaussiane indipendenti è ancora una v.a. gaussiana con media $m_\xi + m_\eta$ e varianza $\sigma_\xi^2 + \sigma_\eta^2$.

(5.37) Aspettazione e sue proprietà. Se ξ è un v.a. semplice possiamo procedere come in (1.23) e definire la sua aspettazione come

$$
\mathrm{E}\,\xi = \sum_k x_k\, \mathrm{P}\,(C_k) \tag{5.38}
$$

dove $C_k = \{\omega \in \Omega : \xi(\omega) = x_k\}$ deve ora essere pensato come un sottoinsieme di uno spazio di probabilità arbitrario Ω. Per mezzo della (5.38) e del teorema

(5.32) definiremo ora l'aspettazione $\mathrm{E}\,\xi$ di un'arbitraria v.a. ξ. Nel linguaggio dell'analisi $\mathrm{E}\,\xi$ altro non è che *integrale di Lebesgue* della funzione $\mathcal{F}$-misurabile $\xi = \xi(\omega)$ rispetto alla misura P. Oltre a $\mathrm{E}\,\xi$ useremo talvolta la notazione $\int_\Omega \xi(\omega)\mathrm{P}\,(d\omega)$. Sia dunque $\xi \geq 0$ una v.a. non negativa e $\xi_n \geq 0$ una sequenza di v.a. semplici tali che $\xi_n(\omega) \uparrow \xi(\omega)$, $n \to \infty$, per ogni $\omega \in \Omega$. Osserviamo che essendo E dato dalla (5.38) un funzionale lineare e positivo sullo spazio delle v.a. semplici, da $\xi_{n+1} \geq \xi_n$ segue $\mathrm{E}\,\xi_{n+1} \geq \mathrm{E}\,\xi_n$. Pertanto il limite $\lim_n \mathrm{E}\,\xi_n$ esiste (possibilmente infinito).

(5.39) Definizione. L'integrale di Lebesgue della v.a. $\xi \geq 0$, o anche la sua aspettazione, è il limite

$$\mathrm{E}\,\xi = \lim_n \mathrm{E}\,\xi_n.$$

Non è difficile rendersi conto che la definizione è consistente, nel senso che il limite non dipende dalla scelta della sequenza $\{\xi_n\}$ che approssima ξ[1]. Supponiamo infatti che $\xi_n \uparrow \xi$ e $\eta_n \uparrow \xi$. Mostreremo che per ogni k, si ha $\mathrm{E}\,\eta_k \leq \lim_n \mathrm{E}\,\xi_n$. Per simmetria si allora anche $\mathrm{E}\,\xi_k \leq \lim_n \mathrm{E}\,\eta_n$, da cui passando al limite $k \to \infty$ segue $\lim_k \mathrm{E}\,\xi_k = \lim_k \mathrm{E}\,\eta_k$. Ora, fissato k, è evidente che $\eta_k \leq \xi$. Allora, se per $\epsilon > 0$ poniamo $A_n = \{\omega : \xi_n \geq \eta_k - \epsilon\}$, si ha $A_n \uparrow \Omega$ e $\xi_n = \xi_n\,1_{A_n} + \xi_n\,1_{A_n^c} \geq \xi_n 1_{A_n} \geq (\eta_k - \epsilon)1_{A_n}$. Per cui, dalle proprietà di monotonicità dell'aspettazione per v.a. semplici si ha che $\mathrm{E}\,\xi_n \geq \mathrm{E}\,(\eta_k - \epsilon)1_{A_n} = \mathrm{E}\,\eta_k 1_{A_n} - \epsilon\,\mathrm{P}\,(A_n) = \mathrm{E}\,\eta_k - \mathrm{E}\,\eta_k 1_{A_n^c} - \epsilon\,\mathrm{P}\,(A_n) \geq \mathrm{E}\,\eta_k - \max_\omega \eta_k(\omega)\mathrm{P}\,(A_n^c) - \epsilon$, ovvero $\lim_n \mathrm{E}\,\xi_n \geq \mathrm{E}\,\eta_k - \epsilon$ e la disuguaglianza cercata segue dall'arbitrarietà di ϵ. q.e.d.

(5.40) Definizione. Diremo che l'aspettazione $\mathrm{E}\,\xi$ di una v.a. ξ *esiste*, o che è *definita*, se almeno uno tra $\mathrm{E}\,\xi^+$ e $\mathrm{E}\,\xi^-$ è finito, ovvero se $\min(\mathrm{E}\,\xi^+, \mathrm{E}\,\xi^-) < \infty$. In tal caso definiremo

$$\mathrm{E}\,\xi \equiv \mathrm{E}\,\xi^+ - \mathrm{E}\,\xi^-.$$

Diremo infine che l'aspettazione di ξ è *finita* se $\mathrm{E}\,\xi^+ < \infty$ e $\mathrm{E}\,\xi^- < \infty$.

[1] Si tratta in essenza del *teorema della convergenza monotòna*.

Essendo poi $|\xi| = \xi^+ + \xi^-$, la finitezza di $\mathrm{E}\,\xi$, cioè $|\mathrm{E}\,\xi| < \infty$, equivale a $\mathrm{E}\,|\xi| = \mathrm{E}\,\xi^+ + \mathrm{E}\,\xi^- < \infty$ (e diremo in tal caso che l'integrale di Lebesgue è assolutamente covergente). Lo spazio delle v.a. $\xi : \Omega \to I\!R$ che verificano $\mathrm{E}\,|\xi| < \infty$ viene talvolta denotato con $L^1(\Omega, \mathcal{F}, \mathrm{P}\,)$. Oltre all'aspettazione, altre caratteristiche significative di una variabile aleatoria ξ sono i numeri $\mathrm{E}\,\xi^p$ (il *momento* di ordine p) e $\mathrm{E}\,|\xi|^p$ (il *momento assoluto* di ordine p. Lo spazio delle v.a. $\xi : \Omega \to I\!R$ che verificano $\mathrm{E}\,|\xi|^p < \infty$ viene talvolta denotato con $L^p(\Omega, \mathcal{F}, \mathrm{P}\,)$. La *varianza* di ξ sarà poi definita nel modo usuale: $\mathrm{Var}\,\xi = \mathrm{E}\,(\xi - \mathrm{E}\,\xi)^2$. La *covarianza* di due v.a. ξ e η è invece definita da $\mathrm{Cov}(\xi, \eta) = \mathrm{E}\,(\xi - \mathrm{E}\,\xi)(\eta - \mathrm{E}\,\eta)$ e il loro *coefficiente di correlazione* è il numero $\rho(\xi, \eta) = \mathrm{Cov}(\xi, \eta)/\sqrt{\mathrm{Var}\,\xi \cdot \mathrm{Var}\,\eta}$. Un fatto importante in teoria della probabilità è che tutte le caratteristiche numeriche associate ad una variabile aleatoria ξ, come l'aspettazione, i momenti, la varianza e così via si possono ottenere dalla misura $\mathrm{P}\,_\xi$ o, alternativamente, dalla funzione distribuzione F_ξ.

(5.41) Teorema. Sia ξ una v.a. e $\phi : I\!R \to I\!R$ una funzione boreliana. Per la v.a. $\eta = \phi \circ \xi$ si ha

$$\mathrm{E}\,\eta = \int_{I\!R} \phi(x)\,\mathrm{P}\,_\xi(dx) = \int_{I\!R} \phi(x)\,dF_\xi(x).$$

(5.42) Corollario. In particolare,

$$\mathrm{E}\,\xi = \int_{I\!R} x\ \mathrm{P}\,_\xi(dx), \quad \mathrm{E}\,\xi^p = \int_{I\!R} x^p\ \mathrm{P}\,_\xi(dx), \quad \mathrm{Var}\,\xi = \int_{I\!R} (x - \mathrm{E}\,\xi)^2\ \mathrm{P}\,_\xi(dx).$$

Dimostrazione di (5.41). Al solito, consideriamo dapprima il caso in cui ϕ è una funzione semplice e non negativa. Siano $C_i = \{x : \phi(x) = \phi_i\}$ la collezione dei suoi insiemi di livello e $B_i = \phi^{-1}(C_i) \in \mathcal{F}$. Allora avremo $\mathrm{E}\,\eta = \sum_i \phi_i \mathrm{P}\,(B_i) = \sum_i \phi_i \mathrm{P}\,_\xi(C_i) = \int \phi(x) \mathrm{P}\,_\xi(dx)$. Il caso $\phi \geq 0$ ma per altro generica si tratta passando al limite in una sequenza di funzioni semplici non negative. Infine nel caso generale si pone $\phi = \phi^+ - \phi^-$, con $\phi^+, \phi^- \geq 0$ e si dimostra l'asserto separatamente per ϕ^+ e ϕ^-. q.e.d.

Un'altra espressione interessante, valida per v.a. non negative, è data dal seguente,

(5.43) Lemma. Se $\xi \geq 0$ allora

$$\mathrm{E}\,\xi = \int_0^\infty \mathrm{P}\left(\xi(\omega) \geq x\right) dx.$$

Dimostrazione. Supponiamo innazitutto che ξ sia semplice con valori $0 \leq x_1 < x_2 < \cdots$ Allora si ha

$$\begin{aligned} \mathrm{E}\,\xi &= \sum_i x_i \mathrm{P}\left(\xi = x_i\right) \\ &= \sum_i x_i \left(\mathrm{P}\left(\xi \geq x_i\right) - \mathrm{P}\left(\xi \geq x_{i+1}\right)\right) \\ &= \sum_i (x_i - x_{i-1})\,\mathrm{P}\left(\xi \geq x_i\right) \end{aligned}$$

dove $(x_0 \equiv 0)$. Ora, essendo $\mathrm{P}(\xi \geq x) = \mathrm{P}(\xi \geq x_1)$ se $0 \leq x \leq x_1$, e $\mathrm{P}(\xi \geq x) = \mathrm{P}(\xi \geq x_i)$ se $x_{i-1} < x \leq x_i$, l'ultima somma scritta sopra può essere riguardata come l'integrale di Riemann della funzione a gradini $\mathrm{P}(\xi \geq x)$. Se ora $\xi \geq 0$ è una v.a. arbitraria e $\xi_n \uparrow \xi$ allora $\mathrm{E}\,\xi_n \uparrow \mathrm{E}\,\xi$; inoltre, evidentemente, $\mathrm{P}(\xi_n \geq x) \uparrow \mathrm{P}(\xi \geq x)$ e dunque $\int_0^\infty \mathrm{P}(\xi_n \geq x)dx \uparrow \int_0^\infty \mathrm{P}(\xi \geq x)dx$, per il teorema della convergenza monotòna. q.e.d.

(5.44) Proprietà dell'aspettazione. L'aspettazione $\mathrm{E}\,\xi$ definita in (5.39) soddisfa alcune proprietà, alcune delle quali analoghe a quelle introdotte nel caso di variabli aleatorie semplici (cfr la discussione che segue (1.23)).

I. Sia $\mathrm{E}\,\xi$ definita e c una costante, allora esiste anche $\mathrm{E}\,(c\xi)$ e $\mathrm{E}\,(c\xi) = c\,\mathrm{E}\,\xi$.

II. Se $\xi \leq \eta$ allora $\mathrm{E}\,\xi \leq \mathrm{E}\,\eta$.

III. Se $\mathrm{E}\,\xi$ esiste allora $|\mathrm{E}\,\xi| \leq \mathrm{E}\,|\xi|$ (che segue da $-|\xi| \leq \xi \leq |\xi|$ e dalla II).

IV. Se $\mathrm{E}\,\xi$ esiste allora esiste anche $\mathrm{E}\,(\xi\, 1_A)$ per ogni $A \in \mathcal{F}$; se poi $\mathrm{E}\,\xi$ è finita, tale è $\mathrm{E}\,(\xi\, 1_A)$; inoltre $\mathrm{E}\,(\xi\, 1_A) \leq \mathrm{E}\,\xi$ se $\xi \geq 0$;

V. Se ξ e η sono non negative, oppure tali che $\mathrm{E}\,|\xi| < \infty$ e $\mathrm{E}\,|\eta| < \infty$, allora $\mathrm{E}\,(\xi + \eta) = \mathrm{E}\,\xi + \mathrm{E}\,\eta$.

VI. La prima disuguaglianza di Chebyshev dimostrata in (1.28) per v.a. semplici diviene ora

$$\mathrm{P}\,(\xi \geq \alpha) = \mathrm{E}\,1_{\{\xi \geq \alpha\}} \leq \frac{\mathrm{E}\,\left(\xi \cdot 1_{\{\xi \geq \alpha\}}\right)}{\alpha} \leq \frac{\mathrm{E}\,\xi}{\alpha}.$$

La (1.32) e le disuguaglianze in (1.56) seguono allo stesso modo.

Vi è poi un gruppo di proprietà dell'aspettazione che si verificano "con probabilità 1" o anche "P -quasi ovunque" (P -q.o.), o semplicemente "quasi-ovunque" (q.o.) ove sia chiaro il riferimento alla misura di probabilità P . Ricordiamo che una proprietà vale P -quasi ovunque se si può trovare un insieme $N \in \mathcal{F}$ con $\mathrm{P}\,(N) = 0$ tale che la proprietà in questione vale per ogni punto $\omega \in \Omega \setminus N$.

VII. Se $\xi = 0$ (q.o.) allora $\mathrm{E}\,\xi = 0$. (Se ξ è semplice allora se $x_k \neq 0$ deve essere $\mathrm{P}\,(C_k) = 0$ e dunque dalla (5.41) si evince subito che $\mathrm{E}\,\xi = 0$. Se poi $\xi \geq \xi_n \geq 0$ con ξ_n semplice allora $\xi_n = 0$ (q.o.) e dunque $\mathrm{E}\,\xi_n = 0$. Pertanto $\mathrm{E}\,\xi = \sup_{\{\xi_n : \xi_n \leq \xi\}} \mathrm{E}\,\xi_n = 0$. Il caso generale segue al solito ponendo $\xi = \xi^+ - \xi^-$ e osservando che $\xi^+ \leq |\xi|$, $\xi^- \leq |\xi|$ e $|\xi| = 0$ (q.o.)).

VIII. Se $\xi = \eta$ (q.o.) e $\mathrm{E}\,|\xi| < \infty$, allora $\mathrm{E}\,|\eta| < \infty$ e $\mathrm{E}\,\xi = \mathrm{E}\,\eta$. (Segue facilmente da VII).

IX. Se $\xi \geq 0$ e $\mathrm{E}\,\xi = 0$ allora $\xi = 0$ (q.o.). (Sia $A = \{\omega : \xi(\omega) > 0\}$, $A_n = \{\omega : \xi(\omega) > 1/n\}$. È evidente che $A_n \uparrow A$ e $0 \leq \xi \cdot 1_{A_n} \leq \xi \cdot 1_A$. Per la proprietà II si ha allora $0 \leq \mathrm{E}\,\xi 1_{A_n} \leq \mathrm{E}\,\xi 1_A = 0$. Quindi $0 = \mathrm{E}\,\xi 1_{A_n} \geq \mathrm{P}\,(A_n)/n$ e $\mathrm{P}\,(A_n) = 0$ $\forall n \geq 1$. Ma $\mathrm{P}\,(A) = \lim_n \mathrm{P}\,(A_n)$ e dunque $\mathrm{P}\,(A) = 0$).

X. Sia ξ una v.a. estesa con $\mathrm{E}\,|\xi| < \infty$. Allora $|\xi| < \infty$ (q.o.). (Se fosse $\mathrm{P}\,(A) > 0$ con $A = \{\omega : \xi(\omega) = \infty\}$ allora si avrebbe $\mathrm{E}\,|\xi| \geq \mathrm{E}\,(|\xi| 1_A) = \infty \cdot \mathrm{P}\,(A) = \infty$).

(5.45) Probabilità e aspettazione condizionate rispetto a una σ-algebra. Sia $(\Omega, \mathcal{F}, \mathrm{P}\,)$ uno spazio di probabilità, $\mathcal{G}$ una σ-algebra, $\mathcal{G} \subseteq \mathcal{F}$ (cioè $\mathcal{G}$ è una σ-sottoalgebra di $\mathcal{F}$).

(5.46) Definizione. L'*aspettazione condizionata* $\mathrm{E}\,(\xi|\mathcal{G})$ *di una v.a.* $\xi = \xi(\omega) \geq 0$ *rispetto alla σ-algebra* $\mathcal{G}$ è la v.a. (estesa) che verifica:

a) $\mathrm{E}\,(\xi|\mathcal{G})(\omega)$ è $\mathcal{G}$-misurabile;

b) per ogni $A \in \mathcal{G}$ si ha

$$\int_A \xi(\omega) \mathrm{P}\,(d\omega) = \int_A \mathrm{E}\,(\xi|\mathcal{G})(\omega) \mathrm{P}\,(d\omega).$$

L'aspettazione condizionata di una v.a. $\xi = \xi(\omega)$ arbitraria è definita se

$$\min\big(\mathrm{E}\,(\xi^+|\mathcal{G}), \mathrm{E}\,(\xi^-|\mathcal{G})\big) < \infty \qquad (\mathrm{P}\text{-q.o.}),$$

e in tal caso vale

$$\mathrm{E}\,(\xi|\mathcal{G}) \equiv \mathrm{E}\,(\xi^+|\mathcal{G}) - \mathrm{E}\,(\xi^-|\mathcal{G}),$$

ove, sull'insieme di probabilità zero dove la condizione sopra non è verficata, si assegni a $\mathrm{E}\,(\xi|\mathcal{G})$ un valore arbitrario, ad esempio 0. Infine, dato $C \in \mathcal{F}$, l'aspettazione condizionata $\mathrm{E}\,(1_C|\mathcal{G})$ si indica anche con $\mathrm{P}\,(C|\mathcal{G})$ e si chiama *probabilità condizionata dell'evento C rispetto alla σ-algebra $\mathcal{G}$*. Infine, se η è una v.a. porremo

$$\mathrm{E}\,(\xi|\eta) \equiv \mathrm{E}\,(\xi|\mathcal{F}_\eta).$$

(5.47) Osservazione. L'esistenza dell'aspettazione condizionata segue dal teorema di Radon-Nicodym. In effetti, la funzione $\mathrm{Q}\,(A) = \int_A \xi(\omega)\mathrm{P}\,(d\omega)$ è una funzione d'insieme σ-additiva su $\mathcal{G}$. Inoltre se $\mathrm{P}\,(A) = 0$ allora anche $\mathrm{Q}\,(A) = 0$, ossia Q è assolutamente continua rispetto a P. Pertanto, per il teorema di Radon-Nicodym si può trovare una funzione (estesa) $\mathcal{G}$-misurabile $\mathrm{E}\,(\xi|\mathcal{G})(\omega) = (d\mathrm{Q}\,/d\mathrm{P}\,)(\omega)$ tale che $\mathrm{Q}\,(A) = \int_A \mathrm{E}\,(\xi|\mathcal{G})(\omega)\mathrm{P}\,(d\omega)$.

(5.48) Osservazione. La definizione $\mathrm{E}\,(\xi|\mathcal{G}) \equiv \mathrm{E}\,(\xi^+|\mathcal{G}) - \mathrm{E}\,(\xi^-|\mathcal{G})$ si riduce alla definizione di $\mathrm{E}\,\xi$ quando $\mathcal{G}$ è la σ-algebra banale, $\mathcal{G} = \{\emptyset, \Omega\}$, ma non presuppone l'esistenza di $\mathrm{E}\,\xi$. Se infatti prendiamo $\mathcal{G} = \mathcal{F}$ e ξ è una v.a. tale che $\mathrm{E}\,\xi^+ = \mathrm{E}\,\xi^- = \infty$, allora $\mathrm{E}\,\xi$ non è definita ma $\mathrm{E}\,(\xi|\mathcal{F})$ esiste ed è semplicemente $\mathrm{E}\,(\xi|\mathcal{F}) = \xi = \xi^+ - \xi^-$. In generale però non sarà vero che $\mathrm{E}\,(\xi|\mathcal{G}) = \xi$, perchè ξ può non essere $\mathcal{G}$-misurabile. Ad esempio se $C \in \mathcal{F}$ ma $C \notin \mathcal{G}$ allora $\xi = 1_C$ non è $\mathcal{G}$-misurabile (perchè $\{\omega : \xi(\omega) = 1\} \notin \mathcal{G}$). Tuttavia in questo caso $\mathrm{E}\,(\xi|\mathcal{G}) = \mathrm{P}\,(C|\mathcal{G})$ è $\mathcal{G}$-misurabile e si ha $\mathrm{P}\,(A \cap C) = \int_A \mathrm{P}\,(C|\mathcal{G})(\omega)\mathrm{P}\,(d\omega)$, per ogni $A \in \mathcal{G}$. In modo simile a quanto fatto per l'aspettazione di possono qui mostrare proprietà analoghe a quelle dell'aspettazione condizionata rispetto a una partizione. La sola differenza è che proprietà come la (3.26), (3.28), (3.32), (3.34), (3.35) e (3.36) saranno ora proprietà valide P-q.o.

(5.49) Osservazione. Un caso importante si ha quando $\mathcal{G} = \sigma(\mathcal{D})$ e $\mathcal{D} = \{D_i\}$ è una partizione finita o numerabile di Ω che sia misurabile rispetto a P. Ciò significa che $\mathrm{P}(D_i) > 0 \ \forall i$ e se $A \subseteq D_i$ allora $\mathrm{P}(A) = 0$ oppure $\mathrm{P}(D_i \setminus A) = 0$, e implica che la distribuzione condizionata $\mathrm{P}(\,\cdot\,|\mathcal{D})$ (vedi (3.22)) esiste P-q.o. In questo caso si può dimostrare che se ξ è una v.a. per cui $\mathrm{E}\,\xi$ è definita, allora

$$\mathrm{E}\,(\xi|\mathcal{G}) = \mathrm{E}\,(\xi|D_i) = \frac{\mathrm{E}\,(\xi \cdot 1_{D_i})}{P(D_i)}, \qquad (\mathrm{P}\text{-q.o. su } D_i).$$

Ciò segue dal fatto generale che se una v.a. η è $\sigma(\mathcal{D})$-misurabile allora è rappresentabile nella forma $\eta(\omega) = \sum_i x_i 1_{D_i}(\omega)$, cioè assume valore costante sugli atomi della partizione $\mathcal{D}$.

(5.50) Esempio. Supponiamo che la durata di un qualche apparecchio (ad es. una lampadina), cioè il tempo in cui cui tale apparecchio continua a funzionare, sia descritto da una v.a. non negativa $\xi = \xi(\omega)$ la cui funzione distribuzione $F_\xi(t)$ ha densità $p_\xi(t) \geq 0$, cosicchè $F_\xi(t) = \int_0^t p_\xi(s)ds$ (perchè essendo $\xi \geq 0$ si ha $F_\xi(t) = p_\xi(t) = 0$ se $t < 0$). Determiniamo l'aspettazione condizionata $\mathrm{E}\,(\xi - \tau|\xi \geq \tau)$, cioè la durata media dell'apparecchio sotto l'ipotesi che abbia già funzionato precedentemente per un tempo τ. Sia $\mathrm{P}\,(\xi \geq \tau) > 0$. Allora si ha, per (5.41),

$$\begin{aligned}\mathrm{E}\,(\xi - \tau|\xi \geq \tau) &= \frac{\mathrm{E}\,((\xi - \tau) \cdot 1_{\{\xi \geq \tau\}})}{\mathrm{P}\,(\xi \geq \tau)} = \frac{\int_\Omega (\xi - \tau) \cdot 1_{\{\xi \geq \tau\}} \mathrm{P}\,(d\omega)}{\mathrm{P}\,(\xi \geq \tau)} \\ &= \frac{\int_\tau^\infty (s - \tau) p_\xi(s)ds}{\int_\tau^\infty p_\xi(s)ds}\,.\end{aligned}$$

In particolare, se ξ è distribuita esponenzialmente, cioè se $p_\xi(t) = \lambda e^{-\lambda t}$ per $t \geq 0$ e $p_\xi(t) = 0$ per $t < 0$, allora si trova (la primitiva di $\lambda\, t\, e^{-\lambda t}$ è $-t\, e^{-\lambda t} - e^{-\lambda t}/\lambda$):

$$\mathrm{E}\,\xi = \int_0^\infty \lambda\, t\, e^{-\lambda t} dt = \frac{1}{\lambda}$$

e anche

$$\mathrm{E}\,(\xi-\tau|\xi\geq\tau)=\frac{\int_\tau^\infty\lambda\,(s-\tau)\,e^{-\lambda s}ds}{\int_\tau^\infty\lambda\,e^{-\lambda s}(s)ds}=\frac{1}{\lambda}$$

per ogni $\tau > 0$. In altre parole, in questo caso la durata media dell'apparecchio sotto l'ipotesi che abbia già funzionato precedentemente per un tempo τ, è indipendente da τ, ed è semplicemente uguale alla sua durata media. Possiamo anche calcolare la funzione distribuzione condizionata $\mathrm{P}\,(\xi-\tau\leq t\,|\,\xi\geq\tau)$. Si trova, usando (5.30),

$$\begin{aligned}\mathrm{P}\,(\xi-\tau\leq t\,|\,\xi\geq\tau)&=\frac{\mathrm{P}\,(t\leq\xi\leq\tau+t)}{\mathrm{P}\,(\xi\geq\tau)}=\frac{F_\xi(t+\tau)-F_\xi(t)+\mathrm{P}\,(\xi=\tau)}{1-F_\xi(\tau)+\mathrm{P}\,(\xi=\tau)}\\&=\frac{(1-e^{-\lambda(t+\tau)})-(1-e^{-\lambda\tau})}{1-(1-e^{-\lambda\tau})}\\&=\frac{e^{-\lambda\tau}(1-e^{-\lambda t})}{e^{-\lambda\tau}}=1-e^{-\lambda t}.\end{aligned}$$

Dunque la distribuzione condizionata $\mathrm{P}\,(\xi-\tau\leq t\,|\,\xi\geq\tau)$ coincide con la funzione distribuzione non condizionata $\mathrm{P}\,(\xi\leq t)$. Questa notevole proprietà è un'esclusiva della distribuzione esponenziale.

(5.51) Problema. Dato un foglio a righe, qual'è la probabilità un ago di lunghezza pari alla metà della distanza tra le righe cada su una di esse quando venga lasciato cadere 'a caso' sul foglio? Per formalizzare questo problema dobbiamo innanzitutto definire cosa significa lasciar cadere l'ago 'a caso'. A questo scopo possiamo considerare il foglio come la ripetizione periodica della regione delimitata tra due righe consecutive assegnate. Denotiamo poi con $0\leq\xi\leq 1$ la distanza del punto di mezzo dell'asticella da una qualisasi delle due righe e con $\theta\in[-\pi/2,\pi/2]$ l'angolo formato dall'ago con la retta passante per il suddetto punto di mezzo e perpendicolare alle due righe. Assumeremo allora che sia ξ che θ siano uniformemente distribuite nei rispettivi intervalli e mutuamente indipendenti. Sia A l'evento in cui l'asticella interseca una delle due rette. Una breve riflessione mostra che $A=\{\omega:(\theta,\xi)\in B\}$ dove

$$B=\left\{a\in[-\pi/2,\pi/2],\ b\in\left[0,\frac{1}{4}\cos a\right]\cup\left[1-\frac{1}{4}\cos a,1\right]\right\},$$

e dunque la probabilità cercata è

$$\mathrm{P}\,(A) = \mathrm{E}\,1_A(\omega) = \mathrm{E}\,1_B(\theta(\omega), \xi(\omega))\,.$$

Ora usando la versione della (3.28) per l'aspettazione condizionata di una v.a. arbitraria rispetto a una σ-algebra si ha

$$\begin{aligned}\mathrm{E}\,1_B(\theta(\omega), \xi(\omega)) &= \mathrm{E}\,(\mathrm{E}\,(1_B(\theta(\omega), \xi(\omega))\,|\,\theta(\omega))) \\ &= \int_\Omega \mathrm{E}\,(1_B(\theta(\omega), \xi(\omega))\,|\,\theta(\omega))\,\mathrm{P}\,(d\omega). \\ &= \int_{-\pi/2}^{\pi/2} \mathrm{E}\,(1_B(\theta(\omega), \xi(\omega))\,|\,\theta(\omega) = a)\,\mathrm{P}_{\theta}(da).\end{aligned}$$

D'altra parte, dalla (versione generale della) (3.34) si ha che se $B \in \mathcal{B}(I\!R^2)$ e ξ e η sono indipendenti allora P-q.o. si ha $\mathrm{E}\,(1_B(\xi,\eta)|\eta = y) = \mathrm{E}\,1_B(\xi, y)$. Inoltre si ha $\mathrm{P}_{\theta}(da) = da/\pi$ e dunque l'integrale scritto sopra diviene

$$\frac{1}{\pi}\int_{-\pi/2}^{\pi/2} \mathrm{E}\,1_B(a, \xi(\omega))\,da = \frac{1}{2\pi}\int_{-\pi/2}^{\pi/2} \cos a\,da = \frac{1}{\pi},$$

perchè

$$\mathrm{E}\,1_B(a, \xi(\omega)) = \mathrm{P}\,\left(\xi \in \left[0, \frac{1}{4}\cos a\right] \cup \left[1 - \frac{1}{4}\cos a, 1\right]\right) = \cos a.$$

Dunque $\mathrm{P}\,(A) = 1/\pi$. Questo risultato può essere usato per una determinazione 'sperimentale' di π: supponendo di poter realizzare una successione di 'lanci' dell'ago mutuamente indipendenti e ponendo $\xi_i = 1$ se l'ago interseca all'i-esimo lancio e $\xi_i = 0$ se non interseca, avremo, per la legge dei grandi numeri,

$$\mathrm{P}\left\{\left|\frac{\xi_1 + \cdots + \xi_n}{n} - \mathrm{P}\,(A)\right| > \delta\right\} \to 0, \quad n \to \infty,$$

per ogni $\delta > 0$. In altre parole, usando il risultato ottenuto sopra, avremo che in una sequenza 'tipica' di lanci $\frac{n}{\xi_1+\cdots+\xi_n} \simeq \pi$. Questa formula, ottenuta per la prima volta nel 1777 da Georges Louis Leclerc, conte di Buffon, è poi stata effettivamente usata per la determinazione statistica di π. Ad esempio, nel 1850, un astronomo di Zurigo, R. Wolf, ha lanciato un ago 5000 volte ottenendo una stima di π pari a 3.1596. Si trattava in effetti di una della prime applicazioni delle regolarità statistiche all'analisi numerica (applicazioni oggi note come *metodi Monte Carlo*).

5.52 Appendice. Entropia e dimensione.

Riprendendo le notazioni dell'Esempio (5.3), sia $p = (p_0, \ldots, p_{N-1})$ e denotiamo con $M(p_0, \ldots, p_{N-1})$ l'insieme dei punti dell'intervallo unitario che nel loro sviluppo N-adico contengono la cifra h nella proporzione p_h, $h = 0, \ldots, N-1$. Allora abbiamo visto che, se $p \neq (1/N, \ldots, 1/N)$, allora $\lambda(M(p_0, \ldots, p_{N-1})) = 0$. Dati $p \neq p' \neq (1/N, \ldots, 1/N)$, una domanda naturale è: in che modo confrontare i due insiemi $M(p_0, \ldots, p_{N-1})$ e $M(p'_0, \ldots, p'_{N-1})$, entrambi di misura di Lebesgue uguale a zero, per vedere 'quale è più grande'? Per rispondervi possiamo procedere come segue. Supponiamo di ricoprire il nostro insieme con un certo numero di palle di raggio ρ e consideriamo il limite $\rho \to 0$. Il numero minimo $N(\rho)$ di palle necessarie a ricoprire l'insieme avrà un andamento del tipo $N(\rho) \simeq \rho^{-D}$ dove D può essere intero o non intero. Più precisamente, dato un insieme non vuoto M, su cui è definita una metrica, sia $N(\rho, M)$ il minimo numero di palle aperte di raggio ρ necessarie a ricoprire M. Allora la *capacità* o *dimensione di Kolmogorov* di M è definita da:

$$\dim_K(M) = \overline{\lim_{\rho\to 0}} \frac{\log N(\rho, M)}{\log(1/\rho)}. \tag{5.53}$$

(5.54) Esercizio. Mostrare che se M è l'insieme di Cantor dei terzi di mezzo allora $\dim_K(M) = \log 2/\log 3 < 1$.

Un'altra caratterizzazione 'dimensionale' di un insieme è data dalla *dimensione di Hausdorff*, introdotta nel 1919. Sia σ un ricoprimento di M con una famiglia numerabile di palle aperte B_i di diametro $\leq \rho$. Dato un numero reale $\alpha > 0$ definiamo la *misura di Hausdorff in dimensione* α di M come il limite

$$m^\alpha(M) = \lim_{\rho\to 0} m^\alpha_\rho(M), \tag{5.55}$$

dove

$$m^\alpha_\rho(M) = \inf_\sigma \sum_i (\operatorname{diam} B_i)^\alpha. \tag{5.56}$$

E' evidente che il limite (5.55) esiste, finito o infinito. Infatti, quando $\rho \to 0$, l'estremo inferiore che definisce $m^\alpha_\rho(M)$ viene preso su ricoprimenti sempre più fini e dunque $m^\alpha_\rho(M)$ cresce, o almeno non decresce. Osserviamo inoltre che $m^\alpha(M)$, come funzione di M, è una misura esterna[1].

La dimensione di Hausdorff di un insieme M si definisce considerando il comportamento di $m^\alpha(M)$ non in funzione di M, ma in funzione di α. Supponiamo che $\{B_i\}$ sia un ρ-ricoprimento di M t.c. $\sum_i(\operatorname{diam} B_i)^\alpha \leq m^\alpha_\rho(M) + 1 \leq m^\alpha(M) + 1 = K < \infty$. Allora se $\alpha' > \alpha$ si ha $m^{\alpha'}_\rho(M) \leq \sum_i(\operatorname{diam} B_i)^{\alpha'} \leq \rho^{\alpha'-\alpha}\sum_i(\operatorname{diam} B_i)^\alpha < \rho^{\alpha'-\alpha}K$ che tende a zero quando $\rho \to 0$. Dunque, se $m^\alpha(M)$ è finita per qualche valore di α allora è nulla per tutti gli α più grandi. Esiste dunque un unico valore α_0, tale che $m^\alpha(M) = +\infty$ per $\alpha < \alpha_0$ e $m^\alpha(M) = 0$ per $\alpha > \alpha_0$. Il valore di $m^\alpha(M)$ in $\alpha = \alpha_0$ può essere 0, positivo e finito, o ∞. Questo numero è la dimensione di Hausdorff di M, che dunque risulta definita da

$$\dim_H(M) = \sup\{\alpha \,:\, m^\alpha(M) = +\infty\} = \inf\{\alpha \,:\, m^\alpha(M) = 0\}. \tag{5.57}$$

(5.58) Esercizio. Mostrare che per ogni insieme compatto M vale la relazione $\dim_H(M) \leq \dim_K(M)$.

Se $M \subset [0,1]$ allora ovviamente $m^1(M) \leq 1$. Dunque $\dim_H(M)$ sta tra 0 e 1. Ad esempio, nel caso di una famiglia di intervalli si ha $m^1(M) > 0$ e quindi $\dim_H(M) = 1$. Al contrario, ogni insieme finito o numerabile di punti ha dimensione di Hausdorff uguale a zero. Tra i due troviamo, ad esempio, l'insieme di Cantor dei terzi di mezzo, per il quale si trova il valore $\dim_H(M) = \dim_K(M) = \log 2/\log 3$ (vedi più avanti). Mostriamo ora che la dimensione di Hausdorff ha interessanti connessioni con l'entropia. A questo scopo, osserviamo che se $M \subseteq [0,1]$, allora le palle B_i sono intervalli e il loro diametro è la loro lunghezza. Consideriamo in particolare gli intervalli N-adici (o insiemi cilindrici)

[1] Una misura esterna è una funzione m^* definita su tutti i sottoinsiemi di uno spazio Ω, con le proprietà: (i) $m^*(C) \in [0,\infty]$ per ogni $C \subset \Omega$; (ii) $m^*(\emptyset) = 0$; (iii) $A \subset B$ implica $m^*(A) \leq m^*(B)$; (iv) $m^*(\cup_i C_i) \leq \sum_i m^*(C_i)$.

della forma

$$I = \left(\frac{h}{N^n}, \frac{h+1}{N^n}\right], \qquad n \geq 1, \quad h = 0, 1, \ldots, N^n - 1, \tag{5.59}$$

Se ora poniamo

$$\ell_\rho^\alpha(M) = \inf \sum_i |I_i|^\alpha, \tag{5.60}$$

dove l'estremo inferiore è ristretto ai ricoprimenti di M con intervalli della forma (5.59) di lunghezza $\leq \rho$, allora non è difficile verificare che $m_\rho^\alpha(M) \leq \ell_\rho^\alpha(M) \leq 2N m_\rho^\alpha(M)$ e dunque possiamo effettuare il calcolo della dimensione di M usando la (5.60) al posto della (5.56). Generalizziamo ulteriormente la definizione data sopra ponendo, per ogni misura di probabilità μ su $\mathcal{B}$,

$$\mu_\rho^\alpha(M) = \inf \sum_i \mu(I_i)^\alpha, \tag{5.61}$$

dove l'estremo inferiore è ristretto ai ricoprimenti di M con intervalli I_i della forma (5.59) e tali che $\mu(I_i) \leq \rho$. Di nuovo, quando $\rho \to 0$, $\mu_\rho^\alpha(M)$ tende monotonamente a un limite $\mu^\alpha(M)$, che come funzione di M è una misura esterna, e per M fissato esiste α_0 t.c. $\mu^\alpha(M) = \infty$ se $\alpha < \alpha_0$ mentre $\mu^\alpha(M) = 0$ se $\alpha > \alpha_0$. Definiamo allora α_0 come la *dimensione di M rispetto alla misura μ*, e indichiamolo con $\dim_\mu(M)$. Evidentemente $\dim_\ell(M) = \dim_H(M)$. Ora, se $I_n(x)$ denota il cilindro n-dimensionale che contiene x, allora si ha il

(5.62) Lemma.

$$M \subset \left\{ x \,:\, \lim_{n\to\infty} \left(-\frac{1}{n} \log \mu(I_n(x)) \right) = \theta \right\} \Longrightarrow \dim_H(M) = \frac{\theta}{\log N} \dim_\mu(M).$$

Pseudo-dimostrazione. (Per una dimostrazione completa si veda [2]) Supponiamo che $\log \mu(I_n(x)) = -n\theta$ per ogni n. Se $\{I_i\}$ è un ricoprimento di M con cilindri della forma (5.59), ciascuno dei quali interseca M, allora ogni I_i ha la forma $I_i = I_n(x)$ per qualche n e $x \in M$, dunque $\mu(I_i) = e^{-n\theta} = |I_i|^{\theta/\log N}$, perchè $|I_i| = N^{-n}$. Pertanto $\sum_i \mu(I_i)^\alpha = \sum_i |I_i|^{\theta\alpha/\log N}$ e dunque $\mu^\alpha(M) = \ell^{\alpha'}(M)$ con $\alpha' = \theta\alpha/\log N$, che implica l'asserto. q.e.d.

(5.63) Corollario. Se μ soddisfa l'ipotesi del Lemma 5.62 e $\mu(M) = \mu^1(M) > 0$ allora $\dim_\mu(M) = 1$ e dunque $\dim_H(M) = \frac{\theta}{\log N}$.

Sia ora μ la misura su $\mathcal{B}$ corrispondente alla (5.17). Allora si ha

$$-\frac{1}{n}\log\mu(I_n(x)) = -\sum_{h=0}^{N-1}\frac{\nu_n^{(h)}(x)}{n}\log p_h = -\sum_{h=0}^{N-1} f_n^{(h)}(x)\log p_h \tag{5.64}$$

D'altra parte è chiaro che $M = M(p_0,\ldots,p_{N-1})$ soddisfa l'ipotesi del Lemma 5.62 con $\theta = h = -\sum_h p_h \log p_h$. Ora dalla (5.23) si ha che $M(p_0,\ldots,p_{N-1})$ ha μ-misura uguale a 1 e dunque, usando (5.63), abbiamo mostrato la seguente relazione tra entropia e dimensione:

(5.65) Teorema.

$$\dim_H(M(p_0,\ldots,p_{N-1})) = -\frac{1}{\log N}\sum_h p_h \log p_h.$$

Dunque la dimensione di Hausdorff di $M(p_0,\ldots,p_{N-1})$ è uguale a uno se e solo se $p_h = 1/N$, $\forall h$, altrimenti è < 1.

Un'altra applicazione di (5.62) è la seguente. Prendiamo $N = 3$, $p = (1/2, 0, 1/2)$ e μ come sopra. Ora $-\frac{1}{n}\log\mu(I_n(x))$ è uguale a $\log 2$ se le prime n cifre dello sviluppo 3-adico di x sono 0 o 2. Così

$$M \subset \left\{x \,:\, \lim_{n\to\infty}\left(-\frac{1}{n}\log\mu(I_n(x))\right) = \log 2\right\} \tag{5.66}$$

se M è l'insieme di Cantor (dei terzi di mezzo). Di nuovo, $\mu(M) = 1$ e (5.63) implicano $\dim_H(M) = \log 2/\log 3$.

5.67 Dimensione e matrici stocastiche. Concludiamo con un'ulteriore applicazione. Consideriamo di nuovo la corrispondenza π, tra sequenze simboliche semi-infinite ω, con alfabeto $X = (0, 1, \ldots, N-1)$, e numeri x in $[0, 1]$ determinata dallo sviluppo N-adico. Sia

$$M(\boldsymbol{Q}) = \{x \in [0,1] \,:\, \lim_n f_n^{(ij)}(\omega) = q_{ij}\}, \quad i, j = 1, \ldots, N,$$

dove $\mathcal{Q} = (q_{ij})$ è una matrice $N \times N$ di elementi non-negativi e tali che se $\pi_i = \sum_j q_{ij}$ allora $\sum_i q_{ij} = \pi_i$ e $\sum_i \pi_i = 1$. Sia $p_{ij} = q_{ij}/\pi_i$; allora $\mathrm{P} = (p_{ij})$ è una matrice stocastica con distribuzione stazionaria (π_i). Supponiamo per semplicità che tutti i p_{ij} siano positivi e consideriamo la misura di probabilità P definita sui cilindri da

$$\mathrm{P}\left(\{\omega \,:\, \omega_0 = x_{i_0}, \omega_1 = x_{i_1}, \ldots, \omega_n = x_{i_n}\}\right) = \mu_{i_0}\, p_{i_0 i_1} \cdots p_{i_{n-1} i_n},$$

per qualche distribuzione inziale (μ_i). La legge forte dei grandi numeri afferma in questo caso che

$$\mathrm{P}\,\left(\{\omega \,:\, \lim_n f_n^{(i)}(\omega) = \pi_i\}\right) = 1, \qquad 1 \le i \le N,$$

e

$$\mathrm{P}\,\left(\{\omega \,:\, \lim_n f_n^{(ij)}(\omega) = \pi_i\, p_{ij}\}\right) = 1, \qquad 1 \le i, j \le N.$$

Sia μ la misura di probabilità su $([0,1], \mathcal{B})$ che corrisponde alla misura di Markov definita sopra. Allora, per ogni $x \in M(\mathcal{Q})$ si ha,

$$\lim_{n\to\infty} \left(-\frac{1}{n} \log \mu(I_n(x))\right) = -\sum_{ij} \pi_i\, p_{ij}\, \log p_{ij},$$

dove $I_n(x)$ è il cilindro n-dimensionale che contiene x. D'altra parte $\mu(M(\mathcal{Q})) = 1$ e dunque, per (5.63)-(5.64), si ha

$$\dim_H(M(\mathcal{Q})) = -\frac{1}{\log N} \sum_{ij} \pi_i\, p_{ij}\, \log p_{ij}.$$

6. SEQUENZE DI VARIABILI ALEATORIE INDIPENDENTI.

La serie $\sum_{n\geq 1}(1/n)$ diverge ma $\sum_{n\geq 1}(-1)^n(1/n)$ converge. Che cosa possiamo dire sulla convergenza o la divergenza della serie $\sum_{n\geq 1}(\xi_n/n)$, dove $\xi_1, \xi_2 \ldots$ è una sequenza di v.a.i.i.d. bernoulliane con $\mathrm{P}(\{\xi_i = 1\}) = \mathrm{P}(\{\xi_i = -1\}) = 1/2$? Detto altrimenti, che cosa possiamo dire sulla convergenza di una serie il cui termine generale sia $\pm 1/n$, con il segno scelto a caso in accordo con la sequenza di v.a. $\xi_1, \xi_2 \ldots$? Sia

$$C_1 = \left\{ \omega \,:\, \sum_{n\geq 1}(\xi_n(\omega)/n) \;\text{ converge} \right\} \tag{6.1}$$

l'insieme delle realizzazioni (cioè dei punti di Ω) per cui $\sum_{n\geq 1}(\xi_n/n)$ converge a un numero finito e consideriamo la probabilità di questo insieme, $\mathrm{P}(C_1)$. A prima vista non sembra affatto chiaro quale valore tale probabilità possa assumere. Vedremo tra poco la 'legge zero-uno' di Kolmogorov che implica che $\mathrm{P}(C_1)$ possa assumere soltanto due valori, 0 o 1. Ma prima ci occorrono alcune nozioni preliminari. Consideriamo uno spazio di probabilità $(\Omega, \mathcal{F}, \mathrm{P})$ e una successione $\{A_n\}_{n\geq 1}$ di eventi, con $A_n \in \mathcal{F}$, $\forall n$. Se $A_1, A_2, \ldots$ hanno probabilità 0, allora lo stesso vale per $\cup_n A_n$. Se, all'opposto, $A_1, A_2, \ldots$ hanno probabilità 1, allora lo stesso vale per $\cap_n A_n$. Data $\{A_n\}_{n\geq 1}$, definiamo l'evento limite

$$A^\infty = \overline{\lim}\, A_n = \cap_{n=1}^\infty \cup_{k=n}^\infty A_k. \tag{6.2}$$

Evidentemente si ha $A^\infty \in \mathcal{F}$ e $\cup_{k=n}^\infty A_k \downarrow A^\infty$, per cui, essendo $A_n \subseteq \cup_{k=n}^\infty A_k$, si ha $\mathrm{P}(A_n) \leq \mathrm{P}(\cup_{k=n}^\infty A_k) \to \mathrm{P}(A^\infty)$. In particolare, se $A_n \to A$, ad esempio se A_n è monotona, allora $A = A^\infty$ e $\lim \mathrm{P}(A_n) = \mathrm{P}(A^\infty)$. Ora, $\omega \in A^\infty$ se e solo se per ogni n si può trovare almeno un $k \geq n$ per cui $\omega \in A_k$; in altre parole, $\omega \in A^\infty$ se e solo se esiste una sequenza infinita $n_i = n_i(\omega)$ tale che $\omega \in A_{n_i}$, $i = 1, 2, \ldots$; se l'indice n viene pensato come un indice temporale allora si può scrivere (6.2) nella forma

$$A^\infty = \{\, A_n \text{ i. s.}\}, \qquad \text{i. s.} \equiv \text{infinitamente spesso} \tag{6.3}$$

6.4 Esempio. Nel contesto dell'esempio (5.3), con $N = 2$, sia $l_n(\omega)$ il numero di 1 *consecutivi* a partire dall'n-esimo simbolo della parola ω. In altre parole $l_n(\omega) = k$ se $\xi_n(\omega) = 1, \ldots, \xi_{n+k-1}(\omega) = 1$ e $\xi_{n+k}(\omega) = 0$. Inoltre porremo $l_n(\omega) = 0$ se $\xi_n(\omega) = 0$. Ora, l'evento $\{\omega : l_n(\omega) = k\}$ è l'unione di 2^{n-1} intervalli (cilindri) disgiunti di lunghezza 2^{-n-k}, e ha dunque probabilità 2^{-k-1}. Se ora fissiamo un intero positivo r e poniamo

$$A_n = \{\omega \,:\, l_n(\omega) \geq r\} = \{\omega \,:\, \xi_i(\omega) = 1,\ n \leq i < n + r\},$$

allora si ha $\mathrm{P}(A_n) = \sum_{k \geq r} 2^{-k-1} = 2^{-r}$ e, per quanto visto sopra $\mathrm{P}(\{ A_n \text{ i. s.}\}) \geq 2^{-r}$. Vedremo tra breve un risultato più forte.

È conveniente formulare a questo punto due risultati semplici ma importanti che risultano essenziali nello studio di proprietà che sono soddisfatte con probabilità 1.

6.5 Lemma. Borel-Cantelli I. Se $\sum_n \mathrm{P}(A_n) < \infty$ allora $\mathrm{P}(\{ A_n \text{ i. s.}\}) = 0$.

Dimostrazione. Da $\{ A_n \text{ i. s.}\} \subset \cup_{k=n}^{\infty} A_k$ segue che $\mathrm{P}(\{ A_n \text{ i. s.}\}) \leq \mathrm{P}(\cup_{k=n}^{\infty} A_k) \leq \sum_{k \geq n} \mathrm{P}(A_k)$, e nelle ipotesi del lemma questa somma tende a 0 quando $n \to \infty$. q.e.d.

6.6 Osservazione. L'argomento usato in (5.4)-(5.12) per dimostrare le legge forte dei grandi numeri può essere riguardato come un'applicazione di (6.5). In effetti, con le notazioni dell'Esempio (5.3), sia $A_n = \left\{|s_n/n| > n^{-1/8}\right\}$. Dalla (5.9) si ha $\sum_n \mathrm{P}(A_n) < \infty$ e dunque $\mathrm{P}(\{ A_n \text{ i. s.}\}) = 0$. Ma per ω nel complementare di $\{ A_n \text{ i. s.}\}$ si ha $s_n(\omega)/n \to 0$.

6.7 Definizione. Gli eventi $A_1, A_2, \ldots$ in $(\Omega, \mathcal{F}, \mathrm{P})$ si diranno indipendenti se tali sono le variabili aleatorie $\xi_1 = 1_{A_1}, \xi_2 = 1_{A_2}, \ldots$.

6.8 Lemma. Borel-Cantelli II. Se $\{A_n\}$ è una sequenza di eventi indipendenti e $\sum_n \mathrm{P}(A_n) = \infty$ allora $\mathrm{P}(\{ A_n \text{ i. s.}\}) = 1$.

Dimostrazione. Poniamo $\overline{C} = \Omega \setminus C$ ed osserviamo che $\overline{\{ A_n \text{ i. s.}\}} = \cup_{n=1}^{\infty} \cap_{k=n}^{\infty} \overline{A}_k$. Sarà dunque sufficiente provare che $\mathrm{P}(\cap_{k=n}^{\infty} \overline{A}_k) = 0$. Ora, usando $1 - x \leq$

e^{-x}, si trova

$$\mathrm{P}\left(\cap_{k=n}^{n+l}\overline{A}_k\right) = \prod_{k=n}^{n+l}(1-\mathrm{P}\left(A_k\right)) \leq \exp\left(-\sum_{k=n}^{n+l}\mathrm{P}\left(A_k\right)\right) \to 0, \quad l\to\infty$$

e dunque $\mathrm{P}\left(\cap_{k=n}^{\infty}\overline{A}_k\right) = \lim_{l\to\infty}\mathrm{P}\left(\cap_{k=n}^{n+l}\overline{A}_k\right) = 0$. q.e.d.

6.9 Esempio. Nella situazione dell'esempio (5.3) sia $\mathbf{s}$ una sequenza prefissata di 0 e 1 di lunghezza k. Consideriamo i cilindri $A_n = \{\omega : (\omega_n \ldots \omega_{n+k-1}) = \mathbf{s}\}$. Evidentemente si ha $\mathrm{P}\left(A_n\right) = 2^{-k}$. A causa delle sovrapposizioni tra, ad esempio, A_n, A_{n+1}, $\cdots$, A_{n+k-1}, gli eventi A_n non sono indipendenti, tuttavia lo sono gli eventi $B_n = A_{(n-1)k+1}$, per i quali vale ancora $\mathrm{P}\left(B_n\right) = 2^{-k}$. Inoltre si ha $\{A_n \text{ i. s.}\} \supset \{B_n \text{ i. s.}\}$. Ora $\sum_{n\geq 1}\mathrm{P}\left(B_n\right) = \infty$, e dunque, per (6.8), si ha $\mathrm{P}\left(\{A_n \text{ i. s.}\}\right) = 1$. Cosí ad esempio, in una sequenza 'tipica' (infinita) di lanci di una moneta, ovvero nello sviluppo diadico di un numero scelto 'a caso' in $[0,1]$, un'arbitraria 'parola' $\mathbf{s}$ di 0 e 1, comparirà un numero infinito di volte.

6.10 Esempio. Riconsideriamo l'esempio (6.4) e invece di prendere r fissato prendiamo una successione non-decrescente r_n di reali positivi. Abbiamo visto che $\mathrm{P}\left(\{\omega : l_n(\omega) \geq r_n\}\right) = 2^{-q_n} \leq 2^{-r_n}$ dove q_n è l'arrotondamento di r_n al primo intero $\geq r_n$. Se ad esempio $r_n = (1+\epsilon)\log_2 n$ con $\epsilon > 0$ allora $\sum_n 2^{-r_n}$ converge, perchè $2^{-r_n} = 1/n^{1+\epsilon}$, e da (6.5) segue che

$$\mathrm{P}\left(\{l_n \geq (1+\epsilon)\log_2 n \ \text{ i. s.}\}\right) = 0. \tag{6.11}$$

Il limite superiore del rapporto $l_n(\omega)/\log_2 n$ supera 1 se e solo se ω appartiene all'evento di probabilità nulla considerato in (6.11) per qualche ϵ razionale positivo. Ora, l'unione di questa classe numerabile di eventi ha anch'essa probabilità nulla e dunque $\mathrm{P}\left(\overline{\lim}_n\, l_n/\log_2 n > 1\right) = 0$ o, che è lo stesso,

$$\mathrm{P}\left(\{\omega : \overline{\lim}\,\frac{l_n(\omega)}{\log_2 n} \leq 1\}\right) = 1. \tag{6.12}$$

Ora, se $r_n = \log_2 n$ allora $\sum_n 2^{-r_n}$ diverge. D'altra parte, per poter applicare Borel-Cantelli II, dobbiamo avere una sequenza di eventi indipendenti. A questo scopo, scegliamo una sottosequenza n_k definita da $n_1 = 1$ e $n_{k+1} = n_k + [r_{n_k} + 1]$, $k \geq 1$. Così, se ad esempio $r_n = \log_2 n$, si trova $n_2 = 2$, $n_3 = 4$, $n_4 = 7$, etc. In altre parole gli indici n_k sono tutti distinti e gli eventi $A_k = \{\omega : l_{n_k}(\omega) \geq r_{n_k}\} = \{\omega : \xi_i(\omega) = 1,\ n_k \leq i < n_{k+1}\}$ sono indipendenti. Inoltre si ha

$$\sum_{k\geq 1} 2^{-r_{n_k}} \geq \sum_{k\geq 1} \frac{2^{-r_{n_k}}}{r_{n_k}+1}(n_{k+1}-n_k) \geq \sum_{k\geq 1} \sum_{n_k \leq n < n_{k+1}} \frac{2^{-r_n}}{r_n+1} = \sum_{n\geq 1} \frac{2^{-r_n}}{r_n+1}.$$

Così, la divergenza di $\sum_{n\geq 1} 2^{-r_n}/r_n$ implica quella di $\sum_{k\geq 1} 2^{-r_{n_k}}$. Per il nostro esempio $\sum_{n\geq 1} 2^{-r_n}/r_n = \sum_{n\geq 1} 1/(n \log_2 n)$, che ancora diverge. Pertanto (6.8) implica

$$\mathrm{P}\left(\{\, l_n \geq \log_2 n \ \text{ i. s.}\}\right) = 1. \tag{6.13}$$

La (6.12) insieme alla (6.13) dà

$$\mathrm{P}\left(\{\omega : \overline{\lim}\, \frac{l_n(\omega)}{\log_2 n} = 1\}\right) = 1. \tag{6.14}$$

In altre parole, per tutti gli ω in un insieme di probabilità 1, $\log_2 n$ costituisce una sorta di inviluppo superiore per la funzione $l_n(\omega)$.

Abbiamo visto fin qui alcuni esempi di eventi, ad esempio quello definito in (6.2)-(6.3), con la curiosa proprietà che il fatto che un punto ω vi appartenga o meno non dipende dalle sue prime n coordinate, cioè dai valori assunti da $\xi_i(\omega) = \omega_i$, $i \leq n$, per quanto grande si possa prendere n.

6.15 Definizione. Sia $\xi_1, \xi_2 \ldots$ una sequenza di v.a. su $(\Omega, \mathcal{F}, \mathrm{P})$ e consideriamo le σ-sottoalgebre $\mathcal{F}_n = \mathcal{F}(\xi_n, \xi_{n+1}, \ldots)$. La loro intersezione $\mathcal{F}_\infty = \cap_{n\geq 1}\mathcal{F}_n$ si chiama σ-algebra all'∞ e ogni suoi elemento si dice evento all'∞, o anche evento asintotico, ed è determinato, in qualche senso, solo dai valori assunti dai termini 'infinitamente lontani' della sequenza $\xi_1, \xi_2 \ldots$

Ad esempio, per l'evento C_1 definito in (6.1) si ha, per ogni $n \geq 1$, $n \geq 1$,

$$C_1 = \left\{\omega \ : \ \sum_{i\geq 1}(\xi_i(\omega)/n) \ \text{ converge}\right\} = \left\{\omega \ : \ \sum_{i\geq n}(\xi_i(\omega)/n) \ \text{ converge}\right\} \in \mathcal{F}_n$$

con $\mathcal{F}_n = \mathcal{F}(\xi_n, \xi_{n+1}, \ldots)$. Pertanto $C_1 \in \mathcal{F}_\infty = \cap_{n\geq 1}\mathcal{F}_n$. Similmente si vede che, data un'arbitraria sequenza I_i di boreliani di $I\!R$,

$$C_2 = \{\omega \,:\, \xi_i(\omega) \in I_i \;\text{ i. s.}\} \in \mathcal{F}_\infty,$$

con $\mathcal{F}_\infty$ come sopra. Un altro esempio è l'evento definito in (5.11):

$$C_3 = \left\{\omega \,:\, \overline{\lim_k}\left|\frac{r_1(\omega)+\cdots+r_k(\omega)}{k}\right| > 0\right\}.$$

In effetti, per ogni $n \geq 1$,

$$C_3 = \left\{\omega \,:\, \overline{\lim_k}\left|\frac{r_n(\omega)+\cdots+r_k(\omega)}{k}\right| > 0\right\},$$

e dunque, se poniamo $A_n = \{\omega : \xi_n(\omega) = 1\}$, allora $C_3 \in \mathcal{F}(A_n, A_{n+1}, \ldots)$ per ogni $n \geq 1$, ovvero $C_3 \in \mathcal{F}_\infty$. Allo stesso modo

$$\overline{C}_3 = \left\{\omega \,:\, \lim_k \frac{r_1(\omega)+\cdots+r_k(\omega)}{k} = 0\right\} \in \mathcal{F}_\infty.$$

Al contrario,

$$\begin{aligned}
B_1 &= \{\omega \,:\, \xi_i(\omega) = 0 \;\forall i \geq 1\}, \\
B_2 &= \left\{\omega \,:\, \lim_n(\xi_1 + \cdots + \xi_n) \;\text{ esiste ed è }\; \leq c\right\}, \\
B_3 &= \{\omega \,:\, \xi_1 + \cdots + \xi_n = 0 \;\text{ i. s. }\},
\end{aligned}$$

sono esempi di eventi che *non* appartengono a $\mathcal{F}_\infty = \cap_{n\geq 1}\mathcal{F}(\xi_n, \xi_{n+1}, \ldots)$. Abbiamo visto nell'Esempio (5.3) che $\mathrm{P}(C_3) = 0$ mentre $\mathrm{P}(\overline{C}_3) = 1$. Inoltre, se supponiamo che le variabili aleatorie ξ_i usate nella definizione dell'evento C_2 siano indipendenti, allora per Borel-Cantelli I e II si ha

$$\begin{aligned}
\mathrm{P}(C_2) = 0 &\Longleftrightarrow \sum_i \mathrm{P}(\{\xi_i \in I_i\}) < \infty, \\
\mathrm{P}(C_2) = 1 &\Longleftrightarrow \sum_i \mathrm{P}(\{\xi_i \in I_i\}) = \infty.
\end{aligned}$$

Quindi la probabilità di C_2 può assumere solo i valori 0 o 1 a seconda del carattere convergente o divergente della serie $\sum_i \mathrm{P}(\{\xi_i \in I_i\})$. Più in generale, le probabilità degli eventi all'∞ è descritta dalla *legge zero-uno di Kolmogorov*:

6.16 Teorema. Se $\xi_1, \xi_2, \ldots$ è una sequenza di v.a. indipendenti e $A \in \mathcal{F}_\infty$, allora $\mathrm{P}(A) = 1$ oppure $\mathrm{P}(A) = 0$.

Dimostrazione. L'idea della prova è che se $A \in \mathcal{F}_\infty$ allora A è indipendente da se stesso e dunque $\mathrm{P}(A) = \mathrm{P}(A \cap A) = \mathrm{P}^2(A)$. Più precisamente, se $A \in \mathcal{F}_\infty$ allora $A \in \mathcal{F}_1^\infty = \mathcal{F}(\xi_1, \xi_2, \ldots) = \mathcal{F}(\cup_n \mathcal{F}_1^n)$, dove $\mathcal{F}_1^n = \mathcal{F}(\xi_1, \ldots, \xi_n)$. Ma allora non è difficile rendersi conto che si può determinare una sequenza di eventi $A_n \in \mathcal{F}_1^n$, $n \geq 1$, tale che $\mathrm{P}(A \triangle A_n) \to 0$, e dunque si ha $\mathrm{P}(A_n) \to \mathrm{P}(A)$ e $\mathrm{P}(A \cap A_n) \to \mathrm{P}(A)$. Ma A non dipende da $\xi_n, \ldots, \xi_n$ per ogni n finito e dunque A e A_n sono indipendenti $\forall n \geq 1$. Pertanto si ha che $\mathrm{P}(A \cap A_n) = \mathrm{P}(A_n) \cdot \mathrm{P}(A) \to \mathrm{P}^2(A) = \mathrm{P}(A)$ e quindi $\mathrm{P}(A) = 0$ o 1. q.e.d.

6.17 Convergenza di serie aleatorie. Sia $\xi_1, \xi_2, \ldots$ una sequenza di v.a. indipendenti e $s_n = \xi_1 + \cdots + \xi_n$. Sia A l'insieme degli eventi elementari ω tali che $s_n(\omega)$ converge a un limite finito. Dalla legge zero-uno segue che $\mathrm{P}(A) = 0$ oppure 1. Cioè la serie $\sum \xi_n$ converge o diverge con probabilità 1. Diamo ora dei criteri che consentano di decidere tra le due eventualità.

6.18 Teorema (Kolmogorov e Khinchin). Sia $\mathrm{E}\,\xi_n = 0$, $n \geq 1$. Allora se $\sum \mathrm{E}\,\xi_n^2 < \infty$, la serie $\sum \xi_n$ converge con probabilità 1.

La dimostrazione dipende in modo essenziale da un'importante disuguaglianza, detta disuguaglianza massimale, la quale essenzialmente afferma che per somme di variabili indipendenti, se $\max_{1 \leq k \leq n} |s_k|$ è grande allora con grande probabilità anche $|s_n|$ è grande.

6.19 Disuguaglianza massimale (Kolmogorov). Siano $\xi_1, \ldots, \xi_n$ v.a. indipendenti con $\mathrm{E}\,\xi_i = 0$ e $\mathrm{E}\,\xi_i^2 < \infty$, $1 \leq i \leq n$. Allora per ogni $\alpha > 0$ si ha

$$\mathrm{P}\left(\left\{\max_{1 \leq k \leq n} |s_k| \geq \alpha\right\}\right) \leq \frac{\operatorname{Var} s_n}{\alpha^2}.$$

Dimostrazione. Decomponiamo l'evento $A = \{\max_{1 \leq k \leq n} |s_k| \geq \alpha\}$ come unione disgiunta degli eventi A_k in cui $|s_k| \geq \alpha$ ma $|s_i| < \alpha$ per $i < k$. Allora si ha

$$\mathrm{E}\, s_n^2 \geq \mathrm{E}\, \left(s_n^2 \cdot 1_A\right) \geq \sum_{k=1}^{n} \mathrm{E}\, \left(s_n^2 \cdot 1_{A_k}\right).$$

D'altra parte si ha che

$$s_n^2 = s_k^2 + 2(s_n - s_k)s_k + (s_n - s_k)^2 \geq s_k^2 + 2s_k(s_n - s_k).$$

Inoltre, usando l'indipendenza delle ξ_i e il fatto che $\mathrm{E}\, \xi_i = 0$ segue che

$$\mathrm{E}\, \left((s_n - s_k)s_k \cdot 1_{A_k}\right) = \mathrm{E}\, (s_n - s_k) \cdot \mathrm{E}\, (s_k \cdot 1_{A_k}) = 0.$$

Pertanto

$$\mathrm{E}\, s_n^2 \geq \sum_{k=1}^{n} \mathrm{E}\, \left(s_k^2 \cdot 1_{A_k}\right) \geq \sum_{k=1}^{n} \alpha^2\, \mathrm{P}\, (A_k) = \alpha^2\, \mathrm{P}\, (A). \quad \text{q.e.d.}$$

Dimostrazione di (6.18). Da (6.19) (e (3.20)) si ha

$$\mathrm{P}\, \left(\left\{\max_{1 \leq k \leq N} |s_{n+k} - s_n| > \epsilon\right\}\right) \leq \frac{1}{\epsilon^2} \sum_{k=1}^{N} \mathrm{Var}\, \xi_{n+k}.$$

D'altra parte, gli eventi di cui stiamo calcolando la probabilità nel termine a sinistra formano una sequenza monotòna in N e quindi, prendendo il limite $N \to \infty$, otteniamo

$$\mathrm{P}\, \left(\left\{\sup_{k \geq 1} |s_{n+k} - s_n| > \epsilon\right\}\right) \leq \frac{1}{\epsilon^2} \sum_{k=1}^{\infty} \mathrm{Var}\, \xi_{n+k}.$$

Ma allora, siccome $\sum_n \mathrm{Var}\, \xi_n$ converge,

$$\lim_n \mathrm{P}\, \left(\left\{\sup_{k \geq 1} |s_{n+k} - s_n| > \epsilon\right\}\right) = 0,$$

per ogni $\epsilon > 0$. Sia ora $E(n, \epsilon)$ l'insieme su cui $\sup_{j,k \geq n} |s_j - s_k| > 2\epsilon$ e $E(\epsilon) = \cap_n E(n, \epsilon)$. Si ha $E(n, \epsilon) \downarrow E(\epsilon)$ e dunque, per quanto abbiamo appena trovato, $P(E(\epsilon)) = 0$. D'altra parte l'insieme $\cup_\epsilon E(\epsilon)$ (dove l'unione viene presa ad esempio sui valori razionali positivi di ϵ), contiene l'insieme su cui $\{s_n\}$ non è di Cauchy, e per quanto visto questo insieme ha probabilità zero. q.e.d.

6.20 Esempio. Nel contesto dell'esempio (5.3) con $N = 2$ poniamo $r_n(\omega) = 2\omega_n - 1$, cosicchè $r_n \in \{-1, +1\}$. Sia inoltre $\{a_n\}$ una successione assegnata e poniamo $\xi_n(\omega) = a_n \cdot r_n(\omega)$. Allora si ha $\mathrm{Var}\,\xi_n = a_n^2$ e dunque $\sum_n a_n^2 < \infty$ implica che $\sum_n \xi_n(\omega)$ converge con probabilità 1. Il caso particolare $a_n = n^{-1}$ mostra che se i segni di $\sum_n \pm n^{-1}$ sono scelti lanciando una moneta (oppure osservando lo sviluppo binario di un numero normale), allora la serie converge con probabilità 1. La serie armonica alternata è dunque, in questo senso, 'tipica'.

Un'altra applicazione non banale della legge zero-uno di Kolmogorov riguarda le passeggiate aleatorie. Sia $\xi_1, \xi_2, \ldots$ una sequenza di v.a.i.i.d. bernoulliane, ossia che possono assumere due valori soltanto, diciamo -1 e 1, con $\mathrm{P}(\{\xi_i = 1\}) = p$, $\mathrm{P}(\{\xi_i = -1\}) = q$, $p+q = 1$, $i \geq 1$, e sia $s_n = \xi_1 + \cdots + \xi_n$ la passeggiata aleatoria corrispondente. E' abbastanza intuitivo che nel caso 'simmetrico' ($p = 1/2$) una passeggiata 'tipica' attraverserà lo zero infinite volte, mentre quando $p \neq 1/2$ se ne andrà all'infinito (cfr. (2.27)). In effetti si ha il seguente risultato:

6.21 Teorema.

a) se $p \neq 1/2$ allora $\mathrm{P}(\{s_n = 0 \ \text{ i. s.}\}) = 0$;

b) se $p = 1/2$ allora $\mathrm{P}(\{s_n = 0 \ \text{ i. s.}\}) = 1$.

Dimostrazione Abbiamo già osservato che l'evento $B = \{s_n = 0 \ \text{ i. s.}\}$ non è un evento asintotico, e dunque non è a priori ovvio che la sua probabilità debba essere 0 o 1. Ora, se poniamo $B_{2n} = \{s_{2n} = 0\}$ allora si ha $\mathrm{P}(B_{2n}) = \binom{2n}{n}(pq)^n \sim \frac{(4pq)^n}{\sqrt{\pi n}}$ e dunque $\sum_n \mathrm{P}(B_{2n}) < \infty$ se $p \neq 1/2$ (perchè, come già osservato, $4p(1-p) < 1$ se $0 < p < 1$, $p \neq 1/2$). La proprietà (a) segue ora da Borel-Cantelli I. Per mostrare (b), sarà sufficiente mostrare che l'evento

$$A = \left\{ \overline{\lim} \frac{s_n}{\sqrt{n}} = \infty, \ \underline{\lim} \frac{s_n}{\sqrt{n}} = -\infty \right\}$$

ha probabilità 1 (perchè $A \subseteq B$). Poniamo

$$A_c = \left\{ \overline{\lim} \frac{s_n}{\sqrt{n}} > c, \right\} \cap \left\{ \underline{\lim} \frac{s_n}{\sqrt{n}} < -c \right\} = A'_c \cap A''_c,$$

cosicchè $A_c \downarrow A$ quando $c \to \infty$. Osserviamo che A, A_c, A'_c, A''_c sono tutti elementi di $\mathcal{F}_\infty = \cap_{n\geq 1} \mathcal{F}(\xi_n, \xi_{n+1}, \ldots)$. Per il teorema (6.16) sarà dunque sufficiente mostrare che $\mathrm{P}(A'_c) > 0$ e $\mathrm{P}(A''_c) > 0$. Osserviamo ora che se ξ_n è una sequenza di v.a. allora si ha $\{\overline{\lim}\, \xi_n > c\} \supseteq \overline{\lim}\{\xi_n > c\}$ (infatti, se $\omega \in \overline{\lim}\{\xi_n > c\}$ allora $\exists n_i(\omega) \uparrow \infty$ tale che $\xi_{n_i}(\omega) > c$, $\forall i$, e dunque $\overline{\lim}\, \xi_n(\omega) > c$). Pertanto

$$\mathrm{P}\left(\left\{ \underline{\lim} \frac{s_n}{\sqrt{n}} < -c \right\} \right) = \mathrm{P}\left(\left\{ \overline{\lim} \frac{s_n}{\sqrt{n}} > c \right\} \right) \geq \overline{\lim}\, \mathrm{P}\left(\left\{ \frac{s_n}{\sqrt{n}} > c \right\} \right) > 0,$$

dove l'ultima disuguaglianza segue dal teorema di De Moivre-Laplace. q.e.d.

Consideriamo ora più in generale una sequenza $\xi_1, \xi_2, \ldots$ di variabili aleatorie su $(\Omega, \mathcal{F}, \mathrm{P})$, indipendenti e identicamente distribuite e poniamo, al solito, $s_n = \xi_1 + \cdots + \xi_n$. Allora gli eventi $\{\omega : \lim_n s_n(\omega)/n = l\}$, $l \in \mathbb{R}$, si trovano in $\mathcal{F}_\infty$ e dunque hanno probabilità 0 o 1. Osserviamo che s_n/n *non* converge al limite l se e solo se si può trovare un ϵ tale che per nessun intero m la quantità $|s_n/n - l|$ sia minore di ϵ quando $n \geq m$. In altre parole, ciò accade se, per qualche ϵ, si abbia $|s_n/n - l| > \epsilon$ per infiniti valori di n. In altre parole si ha,

$$\overline{\{ \lim_n s_n/n = m \}} = \cup_{\epsilon > 0} \{ |\, s_n/n - m \,| > \epsilon \text{ i. s.} \}. \tag{6.22}$$

Mostriamo ora la

6.23 Legge forte dei grandi numeri I (Cantelli). Supponiamo che le v.a.i.i.d. ξ_i abbiano i primi quattro momenti finiti, cioè $\mathrm{E}\, \xi_i^k < \infty$, $k = 1, \ldots, 4$. Sia $m = \mathrm{E}\, \xi_i$. Allora

$$\mathrm{P}\left(\{\omega : \lim_n \frac{s_n(\omega)}{n} = l\}\right) = \begin{cases} 1, & \text{se } l = m; \\ 0, & \text{altrimenti.} \end{cases} \tag{6.24}$$

Dimostrazione. Senza perdita di generalità possiamo assumere $m = 0$ (se $m \neq 0$ possiamo sempre considerare $\xi_i - m$). Dalla (6.22) si evince che l'evento

$\{ \lim_n s_n/n = 0 \}$ ha probabilità 1 se e solo se $\mathrm{P}(\{ |s_n/n| > \epsilon \text{ i. s.}\}) = 0\ \forall\epsilon$. Per Borel-Cantelli I è perciò sufficiente mostrare che $\sum_n \mathrm{P}(\{ |s_n/n| > \epsilon\}) < \infty$. Infine, usando la disuguaglianza di Markov (1.57) con $k = 4$, sarà sufficiente mostrare che $\sum_n \mathrm{E}(s_n/n)^4 < \infty$. D'altra parte, ripetendo alla lettera l'argomento combinatorio usato nell'Esempio (5.3) otteniamo $\mathrm{E}\, s_n^4 = n\mathrm{E}\,\xi_i^4 + 3n(n-1)\mathrm{E}\,\xi_i^2 \le Kn^2$, e dunque l'asserto. q.e.d.

6.25 Osservazione. In accordo con quanto discusso nell'Esempio (5.3), il teorema (6.23) si può riformulare dicendo che la sequenza di variabili aleatorie s_n/n converge P-quasi ovunque al numero m, scritto talvolta

$$\frac{s_n}{n} \to m \qquad (\mathrm{P}\text{-q.o.})$$

Nella forma presentata in (6.23), dovuta a Cantelli, le assunzioni richieste sono piuttosto forti, ma come vedremo tra poco, questo ultime possono essere ridotte alla sola ipotesi che $\mathrm{E}\,|\xi_n| < \infty$ usando tecniche più raffinate (Kolmogorov).

Osserviamo intanto che dalla disuguaglianza di Chebyshev si ottiene la

6.26 Legge debole dei grandi numeri. Se $\mathrm{Var}\,\xi_i = \sigma^2$ allora, per ogni $\epsilon > 0$,

$$\mathrm{P}\left(\left\{\omega : \left|\frac{s_n(\omega)}{n} - m\right| > \epsilon\right\}\right) \le \frac{\sigma^2}{n\epsilon^2} \to 0.$$

6.27 Osservazione. Abbiamo già osservato che per ogni sequenza di eventi $\{A_n\}$ si ha $\mathrm{P}(A_n) \le \mathrm{P}(\cup_{k\ge n} A_k) \to \mathrm{P}(\{ A_n \text{ i. s.}\})$. In particolare, se risulta $\mathrm{P}(\{ A_n \text{ i. s.}\}) = 0$, allora $\lim_n \mathrm{P}(A_n) = 0$ (che può essere vista come l'implicazione inversa di Borel-Cantelli I). Ora, dalla (6.22) segue che $\mathrm{P}(\{ \lim_n s_n/n = m\}) = 1$ se e solo se, per ogni $\epsilon > 0$,

$$\mathrm{P}(\{ |s_n/n - m| > \epsilon \text{ i. s.}\}) = 0, \qquad (6.28)$$

e quindi, per quanto detto,

$$\lim_{n\to\infty} \mathrm{P}(\{ |s_n/n - m| > \epsilon\}) = 0.$$

In altre parole la (6.23) implica la (6.26). Non è difficile rendersi conto che il viceversa è falso. Un semplice controesempio è il seguente: sia (ξ_n) una sequenza di v.a. indipendenti bernoulliane tali che

$$\mathrm{P}\,(\xi_n = 1) = p_n, \qquad \mathrm{P}\,(\xi_n = 0) = 1 - p_n,$$

con (p_n) una sequenza assegnata di numeri non negativi. Allora si ha $\mathrm{E}\,\xi_n = p_n$ e la prima disuguaglianza di Chebyshev implica che $\mathrm{P}\ (\xi_n \geq \epsilon) \leq p_n/\epsilon$, che tende a zero se $p_n \to 0$. Inversamente, se $\mathrm{P}\ (\xi_n \geq \epsilon) \to 0$ allora

$$p_n = \mathrm{E}\,\xi_n = \mathrm{E}\,(\xi_n;\ \xi_n \geq \epsilon) + \mathrm{E}\,(\xi_n;\ \xi_n < \epsilon) < \mathrm{P}\ (\xi_n \geq \epsilon) + \epsilon \leq 2\epsilon$$

se n è grande abbastanza. Abbiamo pertanto che ξ_n converge a 0 in probabilità se e solo se $p_n \to 0$ quando $n \to \infty$. D'altra parte, se anche $p_n \to 0$, ma $\sum_n p_n = \infty$, allora per Borel-Cantelli II si ha che $\mathrm{P}\,\{\xi_n = 1 \text{ i. s.}\}) = 1$. Da ciò si deduce che condizione necessaria affinchè $\xi_n \to 0$ (P-q.o.) è che $\sum_n p_n < \infty$. Il controesempio si ottiene ora prendendo $p_n = 1/n$.

Vediamo ora una versione ancora più forte della (6.23).

6.29 Legge forte dei grandi numeri II (Kolmogorov). Sia $\xi_1, \ldots, \xi_n$ una sequenza di v.a. indipendenti. Supponiamo che le ξ_i abbiano i primi due momenti finiti e che esista inoltre una sequenza $b_n \uparrow \infty$ di numeri positivi tale che

$$\sum_n \frac{\mathrm{Var}\,\xi_n}{b_n^2} < \infty.$$

Allora si ha

$$\frac{s_n - \mathrm{E}\,s_n}{b_n} \to 0 \qquad (\mathrm{P}\text{-q.o.})$$

La dimostrazione di questo teorema richiede due risultati preliminari. Il primo di questi afferma che se una successione $\{x_n\}$ converge ad un limite x allora anche qualunque sua media alla Cesàro converge allo stesso limite.

6.30 Lemma (Toeplitz). Sia $\{a_n\}$ una sequenza di numeri non negativi tale che $\sum_{i=1}^{n} a_i > 0$ per $n \geq 1$ e $\sum_{i=1}^{n} a_i \uparrow \infty$ per $n \to \infty$. Allora, per ogni sequenza di numeri $\{x_n\}$ convergente a un limite x si ha

$$\frac{\sum_{i=1}^{n} a_i\, x_i}{\sum_{i=1}^{n} a_i} \to x \qquad n \to \infty.$$

In particolare, se $a_n \equiv 1$ si ha

$$\frac{1}{n} \sum_{i=1}^{n} x_i \to x \qquad n \to \infty.$$

Dimostrazione. Dato $\epsilon > 0$ scegliamo $n_0 = n_0(\epsilon)$ in modo che $|x_n - x| \leq \epsilon/2$ per $n \geq n_0$. Poniamo ora $b_n \equiv \sum_{i=1}^{n} a_i$ e scegliamo inoltre $n_1 > n_0$ in modo che

$$\frac{\sum_{i=1}^{n_0} |x_i - x|}{b_{n_1}} \leq \frac{\epsilon}{2}.$$

Allora, per $n > n_1$ si ha

$$\begin{aligned}
\left| \frac{1}{b_n} \sum_{i=1}^{n} a_i\, x_i - x \right| &\leq \frac{1}{b_n} \sum_{i=1}^{n} a_i\, |x_i - x| \\
&= \frac{1}{b_n} \sum_{i=1}^{n_0} a_i\, |x_i - x| + \frac{1}{b_n} \sum_{i=n_0+1}^{n} a_i\, |x_i - x| \\
&\leq \frac{1}{b_{n_1}} \sum_{i=1}^{n_0} a_i\, |x_i - x| + \frac{1}{b_n} \sum_{i=n_0+1}^{n} a_i\, |x_i - x| \\
&\leq \frac{\epsilon}{2} \left(1 + \frac{b_n - b_{n_0}}{b_n} \right) \leq \epsilon. \quad \text{q.e.d.}
\end{aligned}$$

6.31 Lemma (Kronecker). Sia $\{b_n\}$ una sequenza di numeri positivi e crescenti tale che $b_n \uparrow \infty$ per $n \to \infty$. Sia inoltre $\{x_n\}$ una sequenza di numeri tale che $\sum x_n$ è convergente. Allora

$$\frac{1}{b_n} \sum_{i=1}^{n} b_i\, x_i \to 0, \qquad n \to \infty.$$

In particolare, se $b_n \equiv n$ e $x_n = y_n/n$ si ha

$$\frac{1}{n} \sum_{i=1}^{n} y_i \to 0, \qquad n \to \infty.$$

Dimostrazione. Poniamo $b_0 = 0$, $s_0 = 0$ e $s_n = \sum_{i=1}^n x_i$. Sommando per parti si ha

$$\sum_{i=1}^n b_i x_i = \sum_{i=1}^n b_i(s_i - s_{i-1}) = b_n\, s_n - b_0\, s_0 - \sum_{i=1}^n s_{i-1}(b_i - b_{i-1})$$

e dunque

$$\frac{1}{b_n}\sum_{i=1}^n b_i\, x_i = s_n - \frac{1}{b_n}\sum_{i=1}^n s_{i-1}\, a_i$$

con $b_n = \sum_{i=1}^n a_i$. La tesi segue ora dal lemma precedente, essendo $s_n \to x$. q.e.d.

Dimostrazione di (6.29). Scriviamo

$$\frac{s_n - \mathrm{E}\, s_n}{b_n} = \frac{1}{b_n}\sum_{i=1}^n b_i \left(\frac{\xi_i - \mathrm{E}\,\xi_i}{b_i}\right).$$

Allora per il lemma di Kronecker, una condizione sufficiente per avere (6.29) è che la serie $\sum \left(\frac{\xi_n - \mathrm{E}\,\xi_n}{b_n}\right)$ converga (P-q.o.). Ma, per il teorema (6.18), ciò è garantito dalla condizione $\sum_n \frac{\mathrm{Var}\,\xi_n}{b_n^2} < \infty$. q.e.d.

6.32 Esempio. Siano $\xi_1, \xi_2 \ldots$ v.a.i.i.d. bernoulliane con $\mathrm{P}(\{\xi_i = 1\}) = \mathrm{P}(\{\xi_i = -1\}) = 1/2$. Dalla dimostrazione del teorema 6.21 si ha, con probabilità 1,

$$\overline{\lim}\, \frac{s_n}{\sqrt{n}} = \infty, \quad \underline{\lim}\, \frac{s_n}{\sqrt{n}} = -\infty.$$

Inoltre, essendo $\sum (n \log^2 n)^{-1} < \infty$ si ha, per (6.29),

$$\frac{s_n}{\sqrt{n}\log n} \to 0,$$

con probabilità 1. Più precisamente si può mostrare la *legge del logaritmo iterato*:

$$\mathrm{P}\left\{\overline{\lim}\, \frac{s_n}{\sqrt{2n\log\log n}} = 1\right\} = \mathrm{P}\left\{\underline{\lim}\, \frac{s_n}{\sqrt{2n\log\log n}} = -1\right\} = 1,$$

che equivale a dire che, per ogni $\epsilon > 0$,

$$\mathrm{P}\left\{|s_n| \ge (1-\epsilon)\sqrt{2n\log\log n} \text{ i. s.}\right\} = 1,$$

$$\mathrm{P}\left\{|s_n| \le (1+\epsilon)\sqrt{2n\log\log n} \text{ i. s.}\right\} = 0.$$

Nel caso in cui le v.a. $\xi_1, \xi_2 \dots$ siano non solo indipendenti ma anche identicamente distribuite (come nell'esempio precedente) possiamo ottenere una legge forte dei grandi numeri sotto la solo ipotesi che le ξ_i stiano in L^1.

6.33 Legge forte dei grandi numeri III (Kolmogorov). Siano $\xi_1, \dots, \xi_n$ v.a.i.i.d. con $\mathrm{E}\,|\xi_1| < \infty$. Allora si ha

$$\frac{s_n}{n} \to m \qquad (\mathrm{P}\text{-q.o.})$$

con $m = \mathrm{E}\,\xi_1$.

La legge dei grandi numeri fa parte di una famiglia di risultati che va sotto il nome di Teoremi Ergodici, i quali forniscono condizioni generali sotto le quali medie aritmetiche di variabili aleatorie ammettono un limite. Tali condizioni si riducono generalmente alla stanzionarietà, ovvero al fatto che la distribuzione congiunta di una qualunque collezione di k v.a. $\xi_{n+1}, \dots, \xi_{n+k}$ non dipende da n (si noti che tale condizione non è verificata nel controesempio discusso in 6.27). Mentre le conseguenze sono che le medie aritmetiche $\frac{1}{n}\sum_{k=1}^{n} \xi_k(\omega)$ sono, 'tipicamente', vicine all'aspettazione e dunque asintoticamente non dipendono da ω, cioè non sono più aleatorie. In altre parole: lunghe serie di variabili aleatorie mostrano con grande probabilità regolarità deterministiche. Come già accennato altrove, tali regolarità possono essere talvolta utilizzate per la costruzione di algoritmi per il calcolo di quantità di vario tipo.

6.34 Esempio. Il metodo Monte Carlo. Sia $f(x)$ una funzione continua definita su $[0,1]$ e a valori in $[0,1]$. Sia $\xi_1, \eta_1, \xi_2, \eta_2, \dots$ una sequenza di (coppie di) v.a. indipendenti e uniformemente distribuite su $[0,1]$. Poniamo

$$\zeta_i = \begin{cases} 1, & \text{se } f(\xi_i) > \eta_i, \\ 0, & \text{se } f(\xi_i) < \eta_i. \end{cases}$$

Per ogni $i \geq 1$ la distribuzione di probabilità congiunta di ξ_i ed η_i è la distribuzione uniforme sul quadrato $[0,1] \times [0,1]$. Evidentemente si ha

$$\mathrm{E}\,\zeta_1 = \mathrm{P}\,\left(f(\xi_1) > \eta_1\right) = \int_0^1 f(x)dx,$$

e, per la legge forte dei grandi numeri,

$$\frac{1}{n}\sum_{i=1}^{n}\zeta_i \to \int_0^1 f(x)dx \qquad (\mathrm{P}\text{-q.o.})$$

Sarà perciò possibile approssimare un integrale $\int_0^1 f(x)dx$ generando numericamente una sequenza (ξ_i, η_i), $i \geq 1$, come sopra e calcolando ζ_i e le sue medie aritmetiche. L'interesse pratico di questo metodo risiede nel fatto che è facilmente estendibile al calcolo di integrali multipli di dimensione anche elevata, per i quali il calcolo numerico delle somme di Riemann corrispondenti comporterebbe una quantità enorme di operazioni.

Dimostrazione di (6.33). Osserviamo innanzitutto che se ξ è una v.a. non negativa, l'espressione (5.43) insieme al fatto che $\mathrm{P}(\xi_k \geq x)$ è monotòna decrescente per $x \geq 0$, implicano che per ogni $\alpha > 0$,

$$\sum_{n\geq 1}\mathrm{P}(\xi \geq \alpha\, n) \leq \frac{\mathrm{E}\,\xi}{\alpha} \leq 1 + \sum_{n\geq 1}\mathrm{P}(\xi \geq \alpha\, n). \tag{6.35}$$

La (6.35) con $\alpha = 1$ insieme alla stazionarietà dà

$$\mathrm{E}\,|\xi_1| < \infty \Leftrightarrow \sum \mathrm{P}(|\xi_1| \geq n) < \infty \Leftrightarrow \sum \mathrm{P}(|\xi_n| \geq n) < \infty.$$

Per Borel-Cantelli I risulta pertanto $\mathrm{P}(|\xi_n| \geq n \text{ i. s.}) = 0$. Dunque, ad eccezione di un insieme finito di valori di n, si ha $|\xi_n| < n$. Poniamo

$$\xi_n^{(n)} = \begin{cases} \xi_n, & |\xi_n| < n, \\ 0, & |\xi_n| \geq n, \end{cases}$$

e $s_n^{(n)} = \xi_1^{(1)} + \cdots + \xi_n^{(n)}$. Supponiamo che $m = \mathrm{E}\,\xi_i = 0$ (altrimenti possiamo considerare le variabili $\xi_i - m$. Allora avremo che $s_n/n \to 0$ $(\mathrm{P}\text{-q.o.})$ se e solo se $s_n^{(n)}/n \to 0$ $(\mathrm{P}\text{-q.o.})$. Ora, in generale avremo $\mathrm{E}\,\xi_n^{(n)} \neq 0$, ma

$$\mathrm{E}\,\xi_n^{(n)} = \mathrm{E}\left(\xi_n \cdot 1_{\{|\xi_n|<n\}}\right) = \mathrm{E}\left(\xi_1 \cdot 1_{\{|\xi_1|<n\}}\right) \to \mathrm{E}\,\xi_1 = 0., \quad n \to \infty.$$

Per il lemma 6.30 si ha allora

$$\frac{1}{n}\sum_{k=1}^{n}\mathrm{E}\,\xi_k^{(k)} \to 0, \quad n \to \infty,$$

e dunque $s_n/n \to 0$ (P -q.o.) se e solo se

$$\frac{\eta_1 + \cdots + \eta_n}{n} \to 0, \quad n \to \infty, \quad (\mathrm{P}\text{-q.o.})$$

dove $\eta_k = \xi_k^{(k)} - \mathrm{E}\,\xi_k^{(k)}$. Ma per il lemma 6.31 questa condizione sarà verificata se $\sum(\eta_n/n)$ converge (P -q.o.), e per il teorema 6.18 ciò accadrà se la serie $\sum(\mathrm{E}\,\eta_n^2/n^2)$ è convergente. D'altra parte, si ha

$$\begin{aligned}
\sum \frac{\mathrm{E}\,\eta_n^2}{n^2} \le \sum \frac{\mathrm{E}\,(\xi_n^{(n)})^2}{n^2} &= \sum \frac{\mathrm{E}\,\left[\xi_n \cdot 1_{\{|\xi_n|<n\}}\right]^2}{n^2} \\
&= \sum \frac{\mathrm{E}\,[\,\xi_1^2 \cdot 1_{\{|\xi_1|<n\}}]}{n^2} \\
&= \sum \frac{1}{n^2} \sum_{k=1}^{n} \mathrm{E}\,\left[\xi_1^2 \cdot 1_{\{k-1 \le |\xi_1| < k\}}\right] \\
&= \sum_{k=1}^{\infty} \mathrm{E}\,\left[\xi_1^2 \cdot 1_{\{k-1 \le |\xi_1| < k\}}\right] \sum_{n \ge k} \frac{1}{n^2} \\
&\le 2 \sum_{k=1}^{\infty} \frac{1}{k} \mathrm{E}\,\left[\xi_1^2 \cdot 1_{\{k-1 \le |\xi_1| < k\}}\right] \\
&\le 2 \sum_{k=1}^{\infty} \mathrm{E}\,\left[|\xi_1| \cdot 1_{\{k-1 \le |\xi_1| < k\}}\right] \\
&= 2\,\mathrm{E}\,|\xi_1| < \infty. \qquad\qquad \text{q.e.d.}
\end{aligned}$$

6.36 Osservazione. Supponiamo di sapere che per una data sequenza $\xi_1, \xi_2 \ldots$ di v.a. i.i.d. le medie artimetiche $(\xi_1 + \cdots + \xi_n)/n$ convergano (P -q.o.) ad un limite finito m. Allora si ha che

$$\frac{\xi_n}{n} = \frac{s_n}{n} - \left(\frac{n-1}{n}\right) \frac{s_{n-1}}{n-1} \to 0 \quad (\mathrm{P}\text{-q.o.}) \tag{6.37}$$

e quindi $\mathrm{P}\,(|\xi_n| > n \text{ i. s.}) = 0$. Borel-Cantelli II implica allora che $\sum \mathrm{P}\,(|\xi_n| > n) < \infty$ e dalla (6.35) si ha $\mathrm{E}\,|\xi_1| < \infty$. Per il teorema appena dimostrato $m = \mathrm{E}\,\xi_1$. Vediamo dunque che per v.a.i.i.d la condizione di avere il primo momento assoluto finito sia necessaria e sufficiente per la convergenza (P -q.o.) delle medie aritmetiche a un limite finito.

6.38 Controesempio. Siano $\xi_1, \xi_2 \dots$ v.a.i.i.d. con $\mathrm{E}\,|\xi_1| = \infty$. Da (5.43) si ha che la funzione $\mathrm{P}\,(|\xi_1| \geq x)$ non appartiene a $L^1(\Omega, \mathcal{F}, \mathrm{P}\,)$. Inoltre dato $\alpha > 0$ si ha

$$1 + \sum_{n \geq 1} \mathrm{P}\,(|\xi_n| \geq \alpha n) = \sum_{n \geq 0} \mathrm{P}\,(|\xi_1| \geq \alpha n).$$

Da (6.35) e $\mathrm{E}\,|\xi_1| = \infty$ si ha che la serie nel termine a sinistra è divergente. Applicando Borel-Cantelli II (si ricordi che le ξ_n sono indipendenti) otteniamo dunque che $\mathrm{P}\,(\{\,|\xi_n| \geq \alpha\, n \text{ i. s.}\}) = 1$ e quindi, essendo α arbitrario, $\overline{\lim}\,|\xi_n|/n = \infty$ con probabilità 1. Insieme alla (6.37) ciò implica che $\mathrm{P}\,\left(\overline{\lim}\,|s_n|/n = \infty\right) = 1$. A sua volta si può mostrare che ciò è equivalente a

$$\overline{\lim_n} \left| \frac{s_n}{n} - a_n \right| = \infty \quad (\mathrm{P}\text{-q.o.}),$$

per ogni sequenza di costanti $\{a_n\}$.

6.39 "Statistiche anomale". Supponiamo che per una sequenza di v.a.i.i.d. $\xi_1, \xi_2 \dots$ si abbia $\mathrm{E}\,|\xi_1|^a < \infty$ per qualche $0 < a < 1$. Mostriamo che in questo caso si ha

$$\frac{s_n}{n^{1/a}} \to 0, \quad n \to \infty, \quad (\mathrm{P}\text{-q.o.}). \tag{6.40}$$

Osserviamo innanzitutto che $s_n/n^{1/a} \to 0$ se e solo se $|s_n|^a/n \to 0$. Inoltre, essendo $0 < a < 1$ si ha

$$|\xi_1 + \cdots + \xi_n|^a < |\xi_1|^a + \cdots + |\xi_n|^a.$$

D'altra parte, se poniamo $\eta_i = |\xi_i|^a$ si ha $\mathrm{E}\,|\eta_i| < \infty$ e dunque per (6.33) si ha che

$$\frac{\eta_1 + \cdots + \eta_n}{n} \to \mathrm{E}\,\eta_1 < \infty, \quad (\mathrm{P}\text{-q.o.}).$$

Ciò implica che l'insieme degli eventi elementari su cui la successione $|s_n|^a/n$ è uniformemente limitata ha probabilità 1. D'altra parte usando la (6.35) con $\xi = |\xi_1|^a$ e l'indipendenza si ha

$$\mathrm{E}\,|\xi_1|^a < \infty \Leftrightarrow \sum \mathrm{P}\,(|\xi_1|^a \geq\, n) < \infty \Leftrightarrow \sum \mathrm{P}\,(|\xi_n| \geq\, n^{1/a}) < \infty.$$

Borel-Cantelli I dà allora

$$\mathrm{P}\left(|\xi_n| \geq n^{1/a} \text{ i. s.}\right) = 0.$$

Dunque, ad eccezione di un insieme finito di valori di n, si ha $|\xi_n| < n^{1/a}$. Possiamo ora procedere in modo analogo a quanto fatto nella dimostrazione di (6.33). Poniamo

$$\xi_n^{(n)} = \begin{cases} \xi_n, & |\xi_n| < n^{1/a}, \\ 0, & |\xi_n| \geq n^{1/a}, \end{cases}$$

e $s_n^{(n)} = \xi_1^{(1)} + \cdots + \xi_n^{(n)}$. Per quanto visto si ha che $s_n/n^{1/a} \to 0$ (P-q.o.) se e solo se $s_n^{(n)}/n^{1/a} \to 0$ (P-q.o.). Per (6.31) ciò accade se $\sum \xi_n^{(n)}/n^{1/a}$ converge (P-q.o.) e quindi, per (6.18), se la serie $\sum \mathrm{E}\,(\xi_n^{(n)})^2/n^{2/a}$ è convergente. D'altra parte, si ha

$$\begin{aligned}
\sum \frac{\mathrm{E}\,(\xi_n^{(n)})^2}{n^{2/a}} &= \sum \frac{\mathrm{E}\,\left[\xi_1^2 \cdot 1_{\{|\xi_1|<n^{1/a}\}}\right]^2}{n^{2/a}} \\
&= \sum \frac{1}{n^{2/a}} \sum_{k=1}^{n} \mathrm{E}\,\left[\xi_1^2 \cdot 1_{\{(k-1)^{1/a} \leq |\xi_1| < k^{1/a}\}}\right] \\
&= \sum_{k=1}^{\infty} \mathrm{E}\,\left[\xi_1^2 \cdot 1_{\{(k-1)^{1/a} \leq |\xi_1| < k^{1/a}\}}\right] \sum_{n \geq k} \frac{1}{n^{2/a}} \\
&\leq \left(\frac{2a}{2-a}\right) \sum_{k=1}^{\infty} \frac{1}{k^{(2-a)/a}} \mathrm{E}\,\left[\xi_1^2 \cdot 1_{\{(k-1)^{1/a} \leq |\xi_1| < k^{1/a}\}}\right] \\
&\leq \left(\frac{2a}{2-a}\right) \sum_{k=1}^{\infty} \mathrm{E}\,\left[|\xi_1|^a \cdot 1_{\{(k-1)^{1/a} \leq |\xi_1| < k^{1/a}\}}\right] \\
&= \left(\frac{2a}{2-a}\right) \mathrm{E}\,|\xi_1|^a < \infty.
\end{aligned}$$

E ciò conclude la dimostrazione di (6.40). In modo simile si può mostrare che se $\mathrm{E}\,|\xi_1|^b < \infty$ per qualche $1 \leq b < 2$, allora

$$\frac{s_n - n\mathrm{E}\,\xi_1}{n^{1/b}} \to 0, \quad n \to \infty, \quad (\mathrm{P}\text{-q.o.}). \tag{6.41}$$

7. CONVERGENZA DEBOLE E FUNZIONI CARATTERISTICHE. IL TEOREMA DEL LIMITE CENTRALE.

Sia $\xi_1, \xi_2, \dots$ una sequenza di v.a.i.i.d. bernoulliane con $\mathrm{P}(\{\xi_i = 1\}) = p$, $\mathrm{P}(\{\xi_i = 0\}) = q$, $p + q = 1$, e $s_n = \xi_1 + \cdots + \xi_n$. La legge dei grandi numeri (6.26) asserisce che la successione s_n/n converge in probabilità al numero p, cioè

$$\mathrm{P}\left(\left|\frac{s_n}{n} - p\right| > \epsilon\right) \to 0. \tag{7.1}$$

7.2 Lemma. Sia ξ_n una successione di v.a. convergente alla v.a. ξ in probabilità. Sia inoltre $f \in C(I\!R)$ una funzione continua e limitata. Allora

$$\mathrm{E}\, f(\xi_n) \to \mathrm{E}\, f(\xi), \qquad n \to \infty.$$

Dimostrazione. Sia $c > 0$ tale che $|f(x)| \le c$. Dato $\epsilon > 0$ scegliamo δ in modo che $|f(x) - f(y)| \le \epsilon$ quando $|x - y| \le \delta$. Allora si ha

$$\begin{aligned}
\mathrm{E}\,(|f(\xi_n) - f(\xi)|) &= \mathrm{E}\,(|f(\xi_n) - f(\xi)|\,;\ |\xi_n - \xi| \le \delta) \\
&\quad + \mathrm{E}\,(|f(\xi_n) - f(\xi)|\,;\ |\xi_n - \xi| > \delta) \\
&\le \epsilon + 2c\,\mathrm{P}\,(|\xi_n - \xi| > \delta).
\end{aligned}$$

Ora, $\mathrm{P}\,(|\xi_n - \xi| > \delta) \to 0$ e dunque $\mathrm{E}\,(|f(\xi_n) - f(\xi)|) \le 2\epsilon$ per n abbastanza grande; e la tesi segue dall'arbitrarietà di ϵ. q.e.d.

Poniamo ora

$$F_n(t) = \mathrm{P}\left(\frac{s_n}{n} \le t\right), \qquad F(t) = \begin{cases} 1, & t \ge p, \\ 0, & t < p. \end{cases}$$

dove $F(t)$ è la funzione distribuzione della v.a. degenere $\xi \equiv p$. Siano poi P_n e P le misure di probabilità su $(I\!R, \mathcal{B}(I\!R))$ corrispondenti alle distribuzioni F_n e F. Osserviamo che si ha

$$\mathrm{E}\, f\left(\frac{s_n}{n}\right) = \int_{I\!R} f(x)\,\mathrm{P}_n(dx), \qquad \mathrm{E}\, f(p) = \int_{I\!R} f(x)\,\mathrm{P}(dx).$$

Dal lemma 7.2 si evince pertanto che la (7.1) implica

$$\int_{I\!R} f(x)\,\mathrm{P}_n(dx) \to \int_{I\!R} f(x)\,\mathrm{P}(dx), \quad \forall f \in C(I\!R), \tag{7.3}$$

ovvero

$$\int_{\mathbb{R}} f(x)\, dF_n(x) \to \int_{\mathbb{R}} f(x)\, dF(x), \quad \forall f \in C(\mathbb{R}). \tag{7.4}$$

In analisi la (7.3) (o la (7.4)) si chiama *convergenza debole* di P_n a P e si indica con $\mathrm{P}_n \Rightarrow \mathrm{P}$ (o anche $F_n \Rightarrow F$). Osserviamo inoltre che si può scrivere

$$F_n(t) = \mathrm{E}\,(1_{\{s_n/n \leq t\}}), \quad F(t) = \mathrm{E}\,(1_{\{t \geq p\}}),$$

e dunque, per il lemma 7.2, si ha che

$$F_n \Rightarrow F \quad \Longrightarrow \quad F_n(t) \to F(t), \quad n \to \infty, \tag{7.5}$$

per tutti i $t \in \mathbb{R}$ ad eccezione del solo punto $t = p$, dove $1_{\{t \geq p\}}$ è discontinua. Vediamo dunque che la convergenza debole $F_n \Rightarrow F$ *non* implica la convergenza puntuale $F_n(t) \to F(t)$, $\forall t \in \mathbb{R}$, ma solo per $t \in P_C(F)$ dove $P_C(F) \subseteq \mathbb{R}$ è l'insieme dei punti di continuità di F, detta anche convergenza *in generale*. Vedremo tra poco (Teorema 7.7) che vale anche il viceversa e dunque che convergenza debole e convergenza in generale sono equivalenti. Se poi, per ogni $n \geq 1$, F_n è la funzione distribuzione di una v.a. ξ_n, e F quella di ξ, allora un modo equivalente di dire che $F_n \Rightarrow F$ è che ξ_n converge *in distribuzione* a ξ, denotato con $\xi_n \xrightarrow{d} \xi$.
Così, se poniamo

$$F_n(t) = \mathrm{P}\left(\frac{s_n - np}{\sqrt{npq}} \leq t\right), \qquad F(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{x^2/2} dx,$$

allora il teorema di De Moivre-Laplace (2.27) asserisce che $F_n(t) \to F(t)$ per ogni $t \in \mathbb{R}$ (in questo caso $P_C(F) \equiv \mathbb{R}$), e per quanto detto ciò costituisce un'affermazione anche sulla convergenza debole di F_n a F, o anche sulla convergenza in distribuzione di $\xi_n = (s_n - np)/\sqrt{npq}$ a una v.a. gaussiana normalizzata. Dimostriamo ora alcuni risultati generali di notevole interesse. Supponiamo innanzitutto di avere uno spazio metrico completo e separabile $(E, \mathcal{E}, \rho)$ con metrica $\rho = \rho(x, y)$ e σ-algebra $\mathcal{E}$ generata dai sottoinsiemi aperti, e una sequenza di misure di probabilità $\mathrm{P}, \mathrm{P}_1, \mathrm{P}_2 \ldots$ su $(E, \mathcal{E}, \rho)$.

7.6 Teorema. Le seguenti proprietà sono equivalenti:

a) $\mathrm{P}_n \Rightarrow \mathrm{P}$;

b) $\overline{\lim}_n \mathrm{P}_n(A) \leq \mathrm{P}(A)$, per ogni A chiuso;

c) $\underline{\lim}_n \mathrm{P}_n(A) \geq \mathrm{P}(A)$, per ogni A aperto.

Dimostrazione. a) $\Rightarrow$ b). Sia A chiuso, $f(x) = 1_A(x)$ e $f_\epsilon(x) = g(\epsilon^{-1}\rho(x, A))$ con $\rho(x, A) = \inf\{\rho(x, y) : y \in A\}$ e g la funzione definita da

$$g(t) = \begin{cases} 1, & t \leq 0, \\ 1 - t, & 0 \leq t \leq 1, \\ 0, & t \geq 1. \end{cases}$$

Poniamo inoltre $A_\epsilon = \{x : \rho(x, A) < \epsilon\}$, cosicché $A_\epsilon \downarrow A$ quando $\epsilon \downarrow 0$. La funzione $f_\epsilon(x)$ è limitata, continua e soddisfa

$$\mathrm{P}_n(A) = \int_E 1_A(x) \mathrm{P}_n(dx) \leq \int_E f_\epsilon(x) \mathrm{P}_n(dx)$$

e dunque

$$\overline{\lim_n} \mathrm{P}_n(A) \leq \overline{\lim_n} \int_E f_\epsilon(x) \mathrm{P}_n(dx) = \int_E f_\epsilon(x) \mathrm{P}(dx) \leq \mathrm{P}(A_\epsilon) \downarrow \mathrm{P}(A), \quad \epsilon \downarrow 0,$$

che dà l'implicazione cercata. L'equivalenza b) $\Leftrightarrow$ c) segue subito prendendo i complementari degli insiemi considerati. In particolare, se prendiamo un sottoinsieme A tale che $\mathrm{P}(\partial A) = 0$ ed indichiamo con $A^o = A \setminus \partial A$ la sua parte interna e con $\mathrm{cl}(A)$ la sua chiusura, allora b) e c) danno

$$\overline{\lim_n} \mathrm{P}_n(A) \leq \overline{\lim_n} \mathrm{P}_n(\mathrm{cl}(A)) \leq \mathrm{P}(\mathrm{cl}(A)) = \mathrm{P}(A),$$
$$\underline{\lim_n} \mathrm{P}_n(A) \geq \underline{\lim_n} \mathrm{P}_n(A^o) \geq \mathrm{P}(A^o) = \mathrm{P}(A),$$

e dunque per mostrare l'implicazione b) $\Leftrightarrow$ c) $\Rightarrow$ a), basterà mostrare che a) segue dal fatto che $\mathrm{P}_n(A) \to \mathrm{P}(A)$ per ogni A tale che $\mathrm{P}(\partial A) = 0$. A questo scopo sia $f(x)$ una funzione continua tale che $|f(x)| \leq M$. Sia D l'insieme, al più numerabile, dato da

$$D = \{t \in \mathbb{R} : \mathrm{P}\{x : f(x) = t\} \neq 0\},$$

e consideriamo una decomposizione $\mathcal{T}_k$ dell'intervallo $[-M, M]$ del tipo

$$-M = a_0 < a_1 < \cdots < a_k = M, \qquad k \geq 1,$$

con $a_i \notin D$, $i = 0, \ldots, k$. Sia inoltre $B_i = \{x : a_i \leq f(x) \leq a_{i+1}\}$. La continuità di f implica che l'insieme $f^{-1}(a_i, a_{i+1})$ è aperto. Perciò $\partial B_i \subseteq f^{-1}\{a_i\} \cup f^{-1}\{a_{i+1}\}$. Ricordando che $a_i, a_{i+1} \notin D$ si ha $\mathrm{P}\,(\partial B_i) = 0$. Ma allora per la supposta proprietà

$$\sum_{i=0}^{k} a_i \mathrm{P}_n(B_i) \to \sum_{i=0}^{k} a_i \mathrm{P}\,(B_i).$$

D'altra parte,

$$\begin{aligned}\left| \int_E f(x) \mathrm{P}_n(dx) - \int_E f(x) \mathrm{P}\,(dx) \right| &\leq \left| \int_E f(x) \mathrm{P}_n(dx) - \sum_{i=0}^{k} a_i \mathrm{P}_n(B_i) \right| + \\ &+ \left| \sum_{i=0}^{k} a_i \mathrm{P}_n(B_i) - \sum_{i=0}^{k} a_i \mathrm{P}\,(B_i) \right| + \left| \sum_{i=0}^{k} a_i \mathrm{P}\,(B_i) - \int_E f(x) \mathrm{P}\,(dx) \right| \\ &\leq 2 \max_{0 \leq i < k} (a_{i+1} - a_i) + \left| \sum_{i=0}^{k} a_i \mathrm{P}_n(B_i) - \sum_{i=0}^{k} a_i \mathrm{P}\,(B_i) \right|,\end{aligned}$$

da cui segue la tesi per quanto visto sopra e per l'arbitrarietà della partizione $\mathcal{T}_k$. q.e.d.

Nel caso particolare in cui $E = I\!\!R$ e $\mathcal{E} = \mathcal{B}(I\!\!R)$ è la σ-algebra generata dalla metrica euclidea $\rho(x, y) = |x - y|$ si ha il seguente,

7.7 Teorema. La convergenza debole $\mathrm{P}_n \Rightarrow \mathrm{P}$ è equivalente alla convergenza $F_n \to F$ in generale (cioè sull'insieme dei punti di continuità) delle rispettive funzioni distribuzione.

Dimostrazione. Dalla dimostrazione del teorema precedente segue che $\mathrm{P}_n \Rightarrow \mathrm{P}$ è equivalente al fatto che $\mathrm{P}_n(A) \to \mathrm{P}\,(A)$ per ogni A tale che $\partial A = \emptyset$. Allora se $\mathrm{P}_n \Rightarrow \mathrm{P}$ si avrà in particolare $\mathrm{P}_n(-\infty, t] \to \mathrm{P}\,(-\infty, t]$ per ogni $t \in I\!\!R$ tale che $\mathrm{P}\,\{t\} = 0$, ovvero $F_n \to F$ in generale (abbiamo già osservato che $F(t)$ è continua in t^* se e solo se $\mathrm{P}\,\{t^*\} = 0$).

Viceversa, supponiamo che $F_n \to F$ in generale. Per il teorema precedente, per mostrare che $\mathrm{P}_n \Rightarrow \mathrm{P}$ sarà sufficiente far vedere che $\underline{\lim}_n \mathrm{P}_n(A) \geq \mathrm{P}(A)$ per ogni aperto A. Ora, se A è aperto, si può trovare una collezione numerabile di intervalli aperti e disgiunti $I_1, I_2, \ldots$ tali che $A = \cup_k I_k$. Dato $\epsilon > 0$ selezioniamo da ciascun $I_k = (a_k, b_k)$ un sottointervallo $I'_k = (a'_k, b'_k]$ in modo tale che $a'_k, b'_k \in P_C(F)$ e $\mathrm{P}(I_k) \leq \mathrm{P}(I'_k) + \epsilon \cdot 2^{-k}$ (tali intervalli certamente esistono poichè $F(t)$ ha al più un insieme numerabile di discontinuità). Per il lemma di Fatou[1],

$$\underline{\lim}_n \mathrm{P}_n(A) = \underline{\lim}_n \sum_k \mathrm{P}_n(I_k) \geq \sum_k \underline{\lim}_n \mathrm{P}_n(I_k) \geq \sum_k \underline{\lim}_n \mathrm{P}_n(I'_k).$$

D'altra parte

$$\mathrm{P}_n(I'_k) = F_n(b'_k) - F_n(a'_k) \to F(b'_k) - F(a'_k) = \mathrm{P}(I'_k).$$

Pertanto

$$\underline{\lim}_n \mathrm{P}_n(A) \geq \sum_k \mathrm{P}(I'_k) \geq \sum_k \left(\mathrm{P}(I_k) - \frac{\epsilon}{2^k}\right) = \mathrm{P}(A) - \epsilon,$$

e l'asserto segue per l'arbitrarietà di $\epsilon > 0$. q.e.d.

In generale, data una sequenza di misure di probabilità $\{\mathrm{P}_n\}$, prima ancora di considerare la questione della sua convergenza (debole) a qualche misura di probabilità P, si dovrà evidentemente stabilire se la sequenza in esame converge a una misura qualsiasi, o perlomeno ammette una sottosequenza convergente. Non è difficile costruire esempi in cui nessuna di queste condizioni è soddisfatta.

[1] (*Lemma di Fatou*) Se ξ_n è una sequenza di v.a. non negative si ha $\underline{\lim}\, \mathrm{E}\,\xi_n \geq \mathrm{E}\, \underline{\lim}_n \xi_n$. *Dimostrazione.* Poniamo $\eta_n = \inf_{k \geq n} \xi_k$, allora $0 \leq g_n \uparrow g = \underline{\lim}_n \xi_n$ e per il teorema di convergenza dominata (vedi più sotto) $\mathrm{E}\,\xi_n \geq \mathrm{E}\,g_n \to \mathrm{E}\,g$. q.e.d.

(*Teorema di convergenza dominata*) Se $|\xi_n| \leq \eta$ P-q.o. con $\eta \in L^1(\Omega, \mathcal{F}, \mathrm{P})$ e se $\xi_n \to \xi$ con probabilità 1, allora $\mathrm{E}\,\xi_n \to \mathrm{E}\,\xi$. *Dimostrazione.* Esercizio (ad esempio come nella dimostrazione del Lemma 7.2, con $f \equiv 1$).

Sia ad esempio P_n, $n \geq 1$, concentrata in $\{n\}$, cioè $\mathrm{P}_n\{n\} = 1$. Dal momento che $\lim_n \mathrm{P}_n(a, b] = 0$ per ogni $a < b$, una eventuale misura limite dovrebbe essere identicamente nulla, contraddicendo il fatto che $1 = \mathrm{P}_n(\mathbb{R}) \not\to 0$. Osserviamo che in questo caso la corrispondente sequenza $\{F_n\}$ di funzioni distribuzione:

$$F_n(t) = \begin{cases} 1, & t \geq n, \\ 0, & t < n, \end{cases}$$

converge per ogni $t \in \mathbb{R}$: $\lim_n F_n(t) \to F(t)$, tuttavia $F(t)$ non è una funzione distribuzione nel senso della definizione 5.28 (perchè $1 \neq \lim_{t\to\infty} F(t) = 0$). Questo esempio mostra che lo spazio delle funzioni distribuzione non è compatto. Inoltre, affinchè una sequenza di funzioni distribuzione converga ad un limite dato anch'esso da una funzione distribuzione occorre introdurre delle ipotesi che impediscano il fenomeno della 'fuga di massa all'infinito'.

Definizione 7.8. Una famiglia di misure di probabilità $\mathcal{P} = \{\mathrm{P}_\alpha\}$ su uno spazio metrico $(E, \mathcal{E}, \rho)$ si dice *relativamente compatta* se ogni sequenza estratta da $\mathcal{P}$ contiene una sottosequenza che converge debolmente a una misura di probabilità.

Osserviamo che in questa definizione la misura limite deve essere una misura *di probabilità*, sebbene non necessariamente un elemento della famiglia $\mathcal{P}$ (da cui l'avverbio 'relativamente'). Consideriamo il caso particolare $E = \mathbb{R}$. Allora, data una famiglia $\mathcal{P} = \{\mathrm{P}_\alpha\}$ su $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, possiamo applicare un classico teorema di selezione di Helly che garantisce, per ogni sequenza $\{\mathrm{P}_n\}$ estratta da $\mathcal{P}$, l'esistenza di una sottosequenza $\{\mathrm{P}_{n_k}\}$ e di una funzione F tali che per le corrispondenti funzioni distribuzionne valga $F_{n_k}(t) \to F(t)$, $\forall t \in P_C(F)$. Tuttavia, come l'esempio discusso più sopra illusra chiaramente, la funzione F può non essere una funzione distribuzione o, che è lo stesso, la misura limite può non essere una misura di probabilità. D'altra parte, se $\mathrm{P}(\mathbb{R}) = 1$ allora per il teorema 7.7 si ha $\mathrm{P}_{n_k} \Rightarrow \mathrm{P}$. Dunque, la famiglia $\mathcal{P}$ sarà relativamente compatta se riusciamo ad assicurare che *ogni* eventuale misura limite P sia normalizzata: $\mathrm{P}(\mathbb{R}) = 1$. Ma $\mathrm{P}(\mathbb{R}) = 1$ se per ogni $\epsilon > 0$ esistono a e b tali che $\mathrm{P}(a, b] \geq 1 - \epsilon$. Supponiamo allora che per ogni $\epsilon > 0$ vi siano a e b tali che per per ogni sequenza

$\{\mathrm{P}_n\}$ estratta da $\mathcal{P}$ si abbia $\mathrm{P}_n(a,b] \geq 1-\epsilon$, $n \geq 1$. Ora, fissata una sequenza $\{\mathrm{P}_n\}$, questa condizione continua a valere se a viene fatto decrescere e b viene fatto crescere e possiamo dunque sceglierli in modo da essere punti di continuità di F. Ma allora da $F_{n_k}(t) \to F(t)$ e dal teorema 7.7 segue che $\mathrm{P}(a,b] \geq 1-\epsilon$. Ne segue che la famiglia $\mathcal{P}$ è relativamente compatta se per ogni $\epsilon > 0$ esistono a e b tali che $\mathrm{P}_\alpha(a,b] \geq 1-\epsilon$ per ogni α. Questa condizione impedisce in questo caso la fuga di massa a cui abbiamo accennato più sopra. D'altra parte, se tale condizione non è soddisfatta, allora esiste un $\epsilon > 0$ tale che, quali che siano a e b, $\mathrm{P}_\alpha(a,b] < 1-\epsilon$ per ogni qualche α. Potremmo dunque scegliere $\mathrm{P}_n \in \mathcal{P}$ in modo tale che $\mathrm{P}_n(-n,n] < 1-\epsilon$. Se ora esistesse una sottosequenza $\{\mathrm{P}_{n_k}\}$ debolmente convergente a una misura di probabilità Q, si avrebbe, per ogni $x \in \mathbb{R}$,

$$Q(-x,x) \leq \overline{\lim}\, \mathrm{P}_{n_k}(-x,x) \leq \overline{\lim}\, \mathrm{P}_{n_k}(-n_k,n_k] < 1-\epsilon,$$

che è impossibile. La condizione suddetta è dunque necessaria e sufficiente. Ora, siccome un intervallo $(a,b]$ ha chiusura compatta ed ogni insieme compatto può essere contenuto in un tale intervallo, la condizione può essere riformulata come segue: una famiglia $\{\mathrm{P}_\alpha\}$ di misure di probabilità su $(\mathbb{R},\mathcal{B}(\mathbb{R}))$ è relativamente compatta se e solo se per ogni $\epsilon > 0$ si può trovare un insieme compatto in modo che $\mathrm{P}_\alpha(K) \geq 1-\epsilon$ per ogni α. Posta in questo modo la condizione ha un senso in qualunque spazio metrico.

Definizione 7.9. Una famiglia $\mathcal{P} = \{\mathrm{P}_\alpha\}$ di misure di probabilità su uno spazio metrico $(E,\mathcal{E},\rho)$ si dice *serrata* (*tight*) se, per ogni $\epsilon > 0$, si può trovare un insieme compatto $K \subseteq E$ tale che

$$\sup_\alpha \mathrm{P}_\alpha(E \setminus K) < \epsilon.$$

Quanto visto sopra nel caso $E = \mathbb{R}$ è un caso particolare di un importante risultato generale.

7.10 Teorema (Prokhorov). Una famiglia $\mathcal{P} = \{\mathrm{P}_\alpha\}$ di misure di probabilità su uno spazio metrico completo e separabile $(E,\mathcal{E},\rho)$ è relativamente compatta se e solo se è serrata.

Introduciamo ora un nuovo concetto che, tra le altre cose, risulta di grande utilità per studiare il limite debole di certe sequenze di misure.

7.11 Funzioni caratteristiche. Sia ξ una v.a. su uno spazio di probabilità $(\Omega, \mathcal{F}, \mathrm{P})$. La sua funzione caratteristica $\varphi_\xi(\lambda)$ è la funzione a valori complessi della variabile reale λ definita da

$$\varphi_\xi(\lambda) = \mathrm{E}\, e^{i\xi\lambda}. \tag{7.12}$$

In accordo con 5.41, $\varphi_\xi(\lambda)$ è anche *la trasformata di Fourier-Stieltjes di* $F_\xi(x)$:

$$\varphi_\xi(\lambda) = \int_{\mathbb{R}} e^{i\lambda x}\, dF_\xi(x). \tag{7.13}$$

Così, se ξ è una v.a. semplice con distribuzione $p_k = \mathrm{P}(\xi = x_k)$, $k = 1, 2, \ldots$, allora si ha

$$\varphi_\xi(\lambda) = \sum_k p_k e^{i\lambda x_k}. \tag{7.14}$$

Se invece ξ è distribuita con densità di probabilità $p_\xi(x)$ allora

$$\varphi_\xi(\lambda) = \int_{\mathbb{R}} e^{i\lambda x}\, p_\xi(x)\, dx, \tag{7.15}$$

e dunque $\varphi_\xi(\lambda)$ è *la trasformata di Fourier* di $p_\xi(x)$. Osserviamo che la funzione caratteristica di una v.a. dipende solo dalla sua funzione distribuzione, e che dunque due v.a. distribuite allo stesso modo hanno la stessa funzione caratteristica. Vedremo tra poco che vale anche il viceversa: se F e G sono due funzioni distribuzione con la stessa funzione caratteristica allora $F(x) = G(x)$.
Alcune semplici proprietà sono le seguenti. Se $\eta = a\xi + b$ allora

$$\varphi_\eta(\lambda) = \mathrm{E}\, e^{i\lambda(a\xi+b)} = e^{i\lambda b}\mathrm{E}\, e^{i\lambda a\xi} = e^{i\lambda b}\varphi_\xi(a\lambda). \tag{7.16}$$

Se poi $\xi_1, \xi_2, \ldots, \xi_n$ sono v.a. indipendenti e $s_n = \xi_1 + \cdots + \xi_n$, allora si ha, usando il lemma 3.20,

$$\varphi_{s_n}(\lambda) = \mathrm{E}\, e^{i\lambda(\xi_1+\cdots+\xi_n)} = \mathrm{E}\, e^{i\lambda\xi_1} \cdots \mathrm{E}\, e^{i\lambda\xi_n} = \prod_{i=1}^n \varphi_{\xi_i}(\lambda). \tag{7.17}$$

Questa proprietà è essenziale nella dimostrazione di svariati teoremi limite per somme di variabili indipendenti con il metodo delle funzioni caratteristiche. A questo proposito osserviamo che per diretta estensione di (5.36) al caso di n v.a. indipendenti si ha che

$$F_{s_n}(t) = F_{\xi_1} * \cdots * F_{\xi_n}(t) \tag{7.18}$$

dove $*$ indica il prodotto di convoluzione. Dal confronto tra (7.13), (7.17) e (7.18) segue una nota proprietà delle trasformate di Fourier.

(7.19) Esempio. Sia $0 < p < 1$ e ξ una v.a. bernoulliana con $\mathrm{P}(\xi = 1) = p$, $\mathrm{P}(\xi = 0) = q = 1 - p$. Allora si vede subito che

$$\varphi_\xi(\lambda) = pe^{i\lambda} + q.$$

Se poi $\xi_1, \ldots, \xi_n$ sono v.a.i.i.d. come ξ allora ponendo $t_n = (s_n - np)/\sqrt{npq}$ si ha

$$\varphi_{t_n}(\lambda) = e^{-i\lambda\sqrt{np/q}} \left[p\, e^{i\lambda/\sqrt{npq}} + q\right]^n = \left[p\, e^{i\lambda\sqrt{q/np}} + q\, e^{-i\lambda\sqrt{p/nq}}\right]^n.$$

Sostituendo gli sviluppi

$$pe^{i\lambda\sqrt{q/np}} = p + i\lambda\sqrt{\frac{pq}{n}} - \frac{\lambda^2 q}{2n} + o\left(\frac{1}{n}\right), \quad qe^{-i\lambda\sqrt{p/nq}} = q - i\lambda\sqrt{\frac{pq}{n}} - \frac{\lambda^2 p}{2n} + o\left(\frac{1}{n}\right),$$

si vede che

$$\varphi_{t_n}(\lambda) \to e^{-\lambda^2/2}, \quad n \to \infty.$$

D'altra parte, come ora vedremo, $e^{-\lambda^2/2}$ è la funzione caratteristica di una variabile distribuita normalmente.

(7.20) Esempio. Sia ξ una v.a. gaussiana, cioè distribuita con densità $p_\xi(x) = \frac{1}{\sqrt{2\pi}\sigma}\, e^{-(x-m)^2/2\sigma^2}$ con $|m| < \infty$ e $\sigma^2 > 0$. Allora si ha

$$\varphi_\xi(\lambda) = e^{i\lambda m} e^{-\lambda^2\sigma^2/2}.$$

Sia infatti $\eta = (\xi - m)/\sigma$ la corrispondente v.a. normalizzata. Per la (7.16) si ha

$$\varphi_\xi(\lambda) = e^{i\lambda m}\varphi_\eta(\sigma\lambda).$$

Sarà dunque sufficiente mostrare che $\varphi_\eta(\lambda) = e^{-\lambda^2/2}$. Si ha

$$\begin{aligned}
\varphi_\eta(\lambda) &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{i\lambda x}e^{-x^2/2}dx = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\sum_{n=0}^{\infty}\frac{(i\lambda x)^n}{n!}e^{-x^2/2}dx \\
&= \sum_{n=0}^{\infty}\frac{(i\lambda)^n}{n!}\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} x^n e^{-x^2/2}dx = \sum_{n=0}^{\infty}\frac{(i\lambda)^{2n}}{(2n)!}(2n-1)!! \\
&= \sum_{n=0}^{\infty}\frac{(i\lambda)^{2n}}{2^n n!} = \sum_{n=0}^{\infty}\frac{(-\lambda^2)^n}{n!} = e^{-\lambda^2/2},
\end{aligned}$$

dove abbiamo usato il fatto che

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} x^{2n+1}e^{-x^2/2}dx \equiv \mathrm{E}\,\eta^{2n+1} = 0,$$

mentre

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} x^{2n}e^{-x^2/2}dx \equiv \mathrm{E}\,\eta^{2n} = (2n-1)(2n-3)\cdots 1 \equiv (2n-1)!!$$

che si ottiene facilmente integrando per parti. Infine

$$\frac{(2n-1)!!}{(2n)!} = \frac{1}{(2n)!!} = \frac{1}{2^n\,n!}.$$

(7.21) Esempio. Sia ξ una v.a. poissoniana con parametro ρ:

$$\mathrm{P}\,(\xi = k) = \frac{e^{-\rho}\rho^k}{k!}.$$

Allora si ha

$$\varphi_\xi(\lambda) = \sum_{k=0}^{\infty} e^{i\lambda k}\frac{e^{-\rho}\rho^k}{k!} = e^{-\rho}\sum_{k=0}^{\infty}\frac{(\rho\, e^{i\lambda})^k}{k!} = \exp\rho(e^{i\lambda}-1).$$

(7.22) Esempio. Sia ξ una v.a. distribuita uniformemente sull'intervallo $[-c,c]$: per $-c < a < b < c$,

$$\mathrm{P}\,(a < \xi < b) = \frac{b-a}{2c}$$

Allora si ha

$$\varphi_\xi(\lambda) = \frac{1}{2c}\int_{-c}^{c} e^{i\lambda x}dx = \frac{\sin c\lambda}{c\lambda}.$$

(7.23) Ulteriori proprietà delle funzioni caratteristiche.

1) $|\varphi_\xi(\lambda)| \le \varphi_\xi(0) = 1$. Se poi $|\varphi_\xi(\lambda_0)| = 1$ per qualche $\lambda_0 \neq 1$ allora ξ è concentrata su $\{a+nh\}_{n\in\mathbb{N}}$ per qualche $a \in \mathbb{R}$ e $h = 2\pi/\lambda_0$.

(*Dimostrazione.*) La prima proprietà è evidente. Per la seconda osserviamo che nelle sotto le ipotesi fatte si può trovare un numero a tale che $\varphi_\xi(\lambda_0) = e^{i\lambda_0 a}$. Pertanto

$$\int_{\mathbb{R}} e^{i\lambda_0 x}dF_\xi(x) = e^{i\lambda_0 a} \Longrightarrow \int_{\mathbb{R}} e^{i\lambda_0(x-a)}dF_\xi(x) = 1$$

che equivale a

$$\int_{\mathbb{R}} [1-\cos(\lambda_0(x-a))]\, dF_\xi(x) = 0$$

Essendo $1-\cos(\lambda_0(\xi-a)) \ge 0$, per la proprietà (5.44)-IX si ha $\cos(\lambda_0(\xi-a)) = 1$ (P-q.o.), che a sua volta è equivalente a

$$\sum_{n\in\mathbb{N}} \mathrm{P}\{\xi = a+nh\} = 1, \qquad \text{q.e.d.}$$

2) $\varphi_\xi(\lambda)$ è uniformemente continua per $\lambda \in \mathbb{R}$.

(*Dimostrazione.*) Ciò segue dalla disuguaglianza $|\varphi_\xi(\lambda+h)-\varphi_\xi(\lambda)| = |\mathrm{E}\, e^{i\lambda\xi}(e^{ih\xi}-1)| \le \mathrm{E}\,|e^{ih\xi}-1|$ e dal teorema di convergenza dominata, per cui $\mathrm{E}\,|e^{ih\xi}-1| \to 0$ quando $h \to 0$. q.e.d.

3) Se ξ ha un momento assoluto di ordine $n \ge 1$, cioè se $\mathrm{E}\,|\xi|^n < \infty$, allora $\varphi_\xi(\lambda)$ ha n derivate continue e $\varphi_\xi^{(n)}(0) = i^n\,\mathrm{E}\,\xi^n$. Se poi $\mathrm{E}\,|\xi|^n < \infty$ per ogni $n \ge 1$ e

$$\overline{\lim_n} \frac{(\mathrm{E}\,|\xi|^n)^{\frac{1}{n}}}{n} = \frac{1}{e\cdot R} < \infty, \tag{7.23}$$

allora

$$\varphi_\xi(\lambda) = \sum_{n=0}^{\infty} \frac{(i\lambda)^n}{n!}\mathrm{E}\,\xi^n, \qquad \forall |\lambda| < R. \tag{7.24}$$

(*Dimostrazione.*) Prendiamo innanzitutto $n = 1$ e consideriamo il rapporto incrementale

$$\frac{\varphi_\xi(\lambda + h) - \varphi_\xi(\lambda)}{h} = \mathrm{E}\, e^{i\lambda\xi} \left(\frac{e^{ih\xi} - 1}{h} \right).$$

Essendo $|(e^{ihx} - 1)/h| \leq x$ e $\mathrm{E}\,|\xi| < \infty$, dal teorema di convergenza dominata segue che il limite $\lim_{h \to 0} \mathrm{E}\, e^{i\lambda\xi}(e^{ih\xi} - 1)/h$ esiste ed è uguale a

$$\lim_{h \to 0} \mathrm{E}\, e^{i\lambda\xi} \left(\frac{e^{ih\xi} - 1}{h} \right) = i\, \mathrm{E}\,(\xi\, e^{i\lambda\xi}) = i \int_{I\!R} x\, e^{i\lambda x} dF(x).$$

Perciò $\varphi'_\xi(\lambda)$ esiste ed è uguale all'(ultima)espressione scritta qui sopra. A questo punto possiamo facilmente procedere per induzione in n per ottenere la prima parte della proprietà 3. Per ottenere la (7.24) assumiamo innanzitutto che $\mathrm{E}\,|\xi|^n < \infty$ ed osserviamo che essendo

$$e^{ix} = \cos x + i \sin x = \sum_{k=0}^{n-1} \frac{(ix)^k}{k!} + \frac{(ix)^n}{n!} \left(\cos(\vartheta_1 x) + i \sin(\vartheta_2 x) \right)$$

con $|\vartheta_1|, |\vartheta_2| \leq 1$, si ha

$$e^{i\lambda\xi} = \sum_{k=0}^{n-1} \frac{(i\lambda\xi)^k}{k!} + \frac{(i\lambda\xi)^n}{n!} \left(\cos(\vartheta_1 \lambda\xi) + i \sin(\vartheta_2 \lambda\xi) \right)$$

con $\vartheta_i = \vartheta_i(\omega)$, $i = 1, 2$, e quindi

$$\mathrm{E}\, e^{i\lambda\xi} = \sum_{k=0}^{n} \frac{(i\lambda)^k}{k!} \mathrm{E}\, \xi^k + R_n(\lambda), \tag{7.25}$$

dove $R_n(\lambda) = ((i\lambda)^n / n!) \mathrm{E}\, [\xi^n (\cos(\vartheta_1 \lambda\xi) + i \sin(\vartheta_2 \lambda\xi) - 1)]$, cosicchè $|R_n(\lambda)| \leq 3\lambda^n \mathrm{E}\,|\xi|^n / n!$ Vale osservare da $\mathrm{E}\,|\xi|^n < \infty$ e dal teorema di convergenza dominata si ha che $R_n(\lambda) \to 0$ quando $\lambda \to 0$. Supponiamo ora che $\mathrm{E}\,|\xi|^n < \infty$ per ogni $n \geq 1$. Allora, dato $0 < \lambda < R$, dalla formula di Stirling si ha

$$\overline{\lim_n} \frac{(\mathrm{E}\,|\xi|^n)^{\frac{1}{n}}}{n} < \frac{1}{e \cdot \lambda} \Rightarrow \overline{\lim_n} \frac{(\lambda^n \mathrm{E}\,|\xi|^n)^{\frac{1}{n}}}{n} < \frac{1}{e} \Rightarrow \overline{\lim_n} \left(\frac{\lambda^n \mathrm{E}\,|\xi|^n}{n!} \right)^{\frac{1}{n}} < 1$$

e la tesi segue dalla (7.25), dalla stima del resto R_n e dal test di Cauchy per la convergenza delle serie. q.e.d.

4) (Formula d'inversione) Per ogni intervallo (a,b) si ha

$$(7.26)\quad \lim_{c\to\infty}\frac{1}{2\pi}\int_{-c}^{c}\frac{e^{-i\lambda a}-e^{-i\lambda b}}{i\lambda}\varphi_\xi(\lambda)\,d\lambda = \mathrm{P}_\xi\{(a,b)\}+\frac{1}{2}\mathrm{P}_\xi\{a\}+\frac{1}{2}\mathrm{P}_\xi\{b\}.$$

Da cui segue in particolare che due se due v.a. hanno le stesse funzioni caratteristiche allora hanno anche le stesse distribuzioni di probabilità (teorema di unicità).

(*Dimostrazione.*) Applicando il teorema di Fubini troviamo

$$\begin{aligned}\frac{1}{2\pi}\int_{-c}^{c}\frac{e^{-i\lambda a}-e^{-i\lambda b}}{i\lambda}\varphi_\xi(\lambda)\,d\lambda &= \frac{1}{2\pi}\int_{-c}^{c}\frac{e^{-i\lambda a}-e^{-i\lambda b}}{i\lambda}\left(\int_{I\!R}e^{i\lambda x}\mathrm{P}_\xi(dx)\right)d\lambda\\ &= \frac{1}{2\pi}\int_{I\!R}\mathrm{P}_\xi(dx)\int_{-c}^{c}\frac{e^{-i\lambda a}-e^{-i\lambda b}}{i\lambda}e^{i\lambda x}d\lambda\end{aligned}.$$

Se ora nell'ultimo integrale usiamo la formula di Eulero per l'esponenziale e osserviamo che l'intervallo d'integrazione è simmetrico attorno a $\lambda=0$ troviamo, dopo alcuni semplici passaggi,

$$\frac{1}{2\pi}\int_{-c}^{c}\frac{e^{-i\lambda a}-e^{-i\lambda b}}{i\lambda}\varphi_\xi(\lambda)\,d\lambda = \int_{I\!R}\mathrm{P}_\xi(dx)\frac{1}{\pi}\int_{c(x-b)}^{c(x-a)}\frac{\sin\mu}{\mu}d\mu\,.$$

Ora, l'integrale $\frac{1}{\pi}\int_{c(x-b)}^{c(x-a)}\frac{\sin\mu}{\mu}d\mu$ è uniformemente limitato in c e dunque

$$\begin{aligned}\lim_{c\to\infty}\frac{1}{2\pi}\int_{-c}^{c}\frac{e^{-i\lambda a}-e^{-i\lambda b}}{i\lambda}\varphi_\xi(\lambda)\,d\lambda &= \int_{I\!R}\mathrm{P}_\xi(dx)\lim_{c\to\infty}\frac{1}{\pi}\int_{c(x-b)}^{c(x-a)}\frac{\sin\mu}{\mu}d\mu\\ &= \int_{I\!R}\Psi_{a,b}(x)\mathrm{P}_\xi(dx),\end{aligned}$$

dove

$$\Psi_{a,b}(x)=\begin{cases}0 & \text{se } x<a,\\ \frac{1}{2} & \text{se } x=a,\\ 1 & \text{se } a<x<b,\\ \frac{1}{2} & \text{se } x=b,\\ 0 & \text{se } x>b.\end{cases}$$

L'asserto finale segue ora da (7.26) e dal fatto che nelle suddette ipotesi le due corrispondenti funzioni distribuzione devono assumere gli stessi valori su tutti i punti di continuità, e dunque devono coincidere dappertutto. q.e.d.

(7.27) Problema. Se la funzione caratteristica $\varphi(\lambda) = \int_{\mathbb{R}} e^{i\lambda x} dF(x)$ è analitica in un intorno di $\lambda = 0$, allora esistono due costanti $C > 0$ e $\kappa > 0$ tale che $\mathrm{P}\{(t, \infty)\} = 1 - F(t) \le C\, e^{-\kappa t}$.

Mostriamo ora un risultato che insieme alle (7.17) e (7.20) fornisce uno strumento potente per studiare la distribuzione limite di somme di variabili aleatorie indipendenti.

(7.28) Teorema di continuità. Sia $\{\mathcal{F}_n\}$ una successione di funzioni distribuzione e $\{\varphi_n\}$ la corrispondente successione di funzioni caratteristiche:

$$\varphi_n(\lambda) = \int_{\mathbb{R}} e^{i\lambda x} dF_n(x).$$

Sia inoltre F una data funzione distribuzione e φ la sua funzione caratteristica. Condizione necessaria e sufficiente per avere $F_n \Rightarrow F$ è che $\varphi_n(\lambda) \to \varphi(\lambda)$ per ogni $\lambda \in \mathbb{R}$.

Dimostrazione. La necessità segue immediatamente dal fatto che $e^{i\lambda x}$ è uniformemente limitata in modulo e applicando la definizione di convergenza debole alla parte reale e alla parte immaginaria di $e^{i\lambda x}$. Per la sufficienza, osserviamo che per il teorema di Fubini si ha

$$\begin{aligned}
\frac{1}{2a}\int_{-a}^{a}(1-\varphi_n(\lambda))d\lambda &= \int_{\mathbb{R}}\left(\frac{1}{2a}\int_{-a}^{a}(1-e^{i\lambda x})d\lambda\right)\mathrm{P}_n(dx) \\
&= \int_{\mathbb{R}}\left(1-\frac{\sin ax}{ax}\right)\mathrm{P}_n(dx) \\
&\ge \int_{|x|\ge 2/a}\left(1-\frac{1}{|ax|}\right)\mathrm{P}_n(dx) \\
&\ge \mathrm{P}_n\left(x : |x| \ge 2/a\right)\cdot
\end{aligned}$$

Ora, siccome φ è continua nell'origine e $\varphi(0) = 1$, per ogni dato ϵ si può trovare un numero a in modo che $\frac{1}{2a}\int_{-a}^{a}(1-\varphi(\lambda))d\lambda < \epsilon$. D'altra parte φ_n converge a φ e dunque vi sarà un intero n_0 tale che $\frac{1}{2a}\int_{-a}^{a}(1-\varphi_n(\lambda))d\lambda < 2\epsilon$ per ogni $n \ge n_0$ e dunque, per quanto trovato più sopra $\mathrm{P}_n\left(x : |x| \ge 2/a\right) < 2\epsilon$ se $n \ge n_0$. Diminuendo a se necessario possiamo sempre fare in modo che questa disuguaglianza sia verificata anche per l'insieme finito degli n con $n < n_0$. Dunque $\{\mathrm{P}_n\}$ è serrata. Per il teorema 7.10 basterà dunque mostrare che ogni sottosuccessione $\{\mathrm{P}_{n_k}\}$ che converge debolmente, converge a P. Ora se $\mathrm{P}_{n_k} \Rightarrow \mathrm{P}'$ quando $k \to \infty$, allora per l'implicazione di necessità vista sopra P' deve avere funzione caratteristica φ, ma allora per il teorema di unicità (corollario di (7.26)) deve essere $\mathrm{P}' = \mathrm{P}$. q.e.d.

Vediamo ora alcune applicazioni immediate del teorema precedente allo studio del comportamento di somme di v.a.i.i.d., riservandoci di discutere in seguito il caso più interessante di somme v.a. indipendenti ma distribuite in modo arbitrario. Mostriamo innanzitutto come si può facilmente ottenere la legge (debole) dei grandi numeri sotto le stesse ipotesi del teorema 6.33 (e dunque senza poter usare la seconda disuguaglianza di Chebyshev).

7.29 Legge dei grandi numeri. Siano $\xi_1, \ldots, \xi_n$ v.a.i.i.d. con $\mathrm{E}\,|\xi_1| < \infty$ e $s_n = \xi_1 + \cdots + \xi_n$. Allora s_n/n converge in probabilità al numero $m = \mathrm{E}\,\xi_1$.

Dimostrazione. Da (7.16) e (7.17) si ha

$$\varphi_{s_n/n}(\lambda) = \left[\varphi_{\xi_1}\left(\frac{\lambda}{n}\right)\right]^n.$$

Ora, essendo $\mathrm{E}\,|\xi_1| < \infty$ si ha per (7.25),

$$\varphi_{\xi_1}(\lambda) = 1 + \lambda m + o(\lambda), \qquad \lambda \to 0,$$

e dunque, per ogni $\lambda \in \mathbb{R}$,

$$\varphi_{\xi_1}\left(\frac{\lambda}{n}\right) = 1 + i\,\frac{\lambda m}{n} + o\left(\frac{\lambda}{n}\right), \quad n \to \infty,$$

da cui si ottiene

$$\varphi_{s_n/n}(\lambda) = \left[1 + i\,\frac{\lambda m}{n} + o\left(\frac{\lambda}{n}\right)\right]^n \to e^{i\lambda m}.$$

Ma la funzione $\varphi(\lambda) = e^{i\lambda m}$ è continua in 0 ed è la funzione caratteristica della v.a. degenere $\xi \equiv m$. Per il teorema 7.28 ciò mostra che $s_n/n \xrightarrow{d} m$. D'altra parte non è difficile rendersi conto (esercizio) che $\xi_n \xrightarrow{d} c$ con c costante implica che $\xi_n \to c$ in probabilità (dunque, nel caso in cui la v.a. limite sia degenere è vero anche il viceversa del lemma 7.2). q.e.d.

7.30 Teorema limite di Poisson. Per ogni $n \geq 1$, siano date n v.a. indipendenti $\xi_{n1}, \dots, \xi_{nn}$ tali che

$$\mathrm{P}\left(\xi_{nk} = 1\right) = p_{nk}, \quad \mathrm{P}\left(\xi_{nk} = 0\right) = q_{nk}, \quad p_{nk} + q_{nk} = 1,$$

Supponiamo che $\max_{1 \leq k \leq n} p_{nk} \to 0$ ma $\sum_{k=1}^{n} p_{nk} \to \rho > 0$ quando $n \to \infty$. Se poniamo $s_n = \xi_{n1} + \cdots + \xi_{nn}$, allora per ogni $m = 0, 1, 2 \dots$ si ha

$$\mathrm{P}\left(s_n = m\right) \to \frac{e^{-\rho}\,\rho^m}{m!}, \qquad n \to \infty.$$

Dimostrazione. Si ha $\mathrm{E}\, e^{i\lambda\xi_{nk}} = p_{nk}e^{i\lambda} + q_{nk}$ e dunque per la (7.17) e per $n \to \infty$,

$$\varphi_{s_n}(\lambda) = \prod_{k=1}^{n}\left(p_{nk}e^{i\lambda} + q_{nk}\right) = \prod_{k=1}^{n}\left(1 + p_{nk}(e^{i\lambda} - 1)\right) \to \exp\left(\rho(e^{i\lambda} - 1)\right),$$

e l'asserto segue da 7.21 e 7.28. q.e.d.

7.31 Teorema del limite centrale per v.a.i.i.d. Siano $\xi_1, \dots, \xi_n$ v.a.i.i.d. (non degeneri) con $\mathrm{E}\,\xi_1^2 < \infty$ e $s_n = \xi_1 + \cdots + \xi_n$. Allora

$$\mathrm{P}\left(\frac{s_n - \mathrm{E}\, s_n}{\sqrt{\mathrm{Var}\, s_n}} \leq t\right) \to \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-x^2/2} dx, \qquad t \in I\!R.$$

Dimostrazione. Poniamo $\mathrm{E}\,\xi_1 = m$, $\mathrm{Var}\,\xi_1 = \sigma^2$ e $\eta_n = (s_n - \mathrm{E}\, s_n)/\sqrt{\mathrm{Var}\, s_n}$. Allora si ha

$$\varphi_{\eta_n}(\lambda) = \left[\varphi_{\xi_1 - m}\left(\frac{\lambda}{\sigma\sqrt{n}}\right)\right]^n.$$

Ora, essendo $\mathrm{E}\,\xi_1^2 < \infty$ si ha per (7.25),

$$\varphi_{\xi_1 - m}(\lambda) = 1 - \frac{\sigma^2\lambda^2}{2} + o(\lambda^2), \qquad \lambda \to 0,$$

e dunque, per ogni $\lambda \in I\!R$,

$$\varphi_{\eta_n}(\lambda) = \left[1 - \frac{\sigma^2\lambda^2}{2\sigma^2 n} + o\left(\frac{1}{n}\right)\right]^n \to e^{-\lambda^2/2}, \quad n \to \infty,$$

e l'asserto segue da 7.20 e 7.28. q.e.d.

7.32 Teorema del limite centrale per somme di v.a. indipendenti. Sia $\xi_1, \dots, \xi_n$ una sequenza di v.a. indipendenti con $\mathrm{E}\,\xi_i^2 < \infty$, per ogni i. Poniamo $m_i = \mathrm{E}\,\xi_i$, $\sigma_i^2 = \mathrm{Var}\,\xi_i$, $s_n = \sum_{i=1}^n \xi_i$, $d_n^2 = \sum_{i=1}^n \sigma_i^2$ e sia $F_i(t)$ la f.d. della v.a. ξ_i. Supponiamo inoltre che valga la *condizione di Lindeberg*: per ogni $\epsilon > 0$

$$\frac{1}{d_n^2} \sum_{i=1}^n \int_{\{x : |x-m_i| \geq \epsilon d_n\}} (x-m_i)^2 dF_i(x) \to 0, \quad n \to \infty. \tag{7.33}$$

Allora

$$\mathrm{P}\left(\frac{s_n - \mathrm{E}\,s_n}{d_n} \leq t \right) \to \frac{1}{\sqrt{2\pi}} \int_{-\infty}^t e^{-x^2/2} dx, \qquad t \in I\!R. \tag{7.34}$$

Dimostrazione. Supponiamo senza perdita di generalità che $m_i = \mathrm{E}\,\xi_i = 0$, $\forall i$. Posto $t_n = s_n/d_n$, per quanto visto in precedenza sarà sufficiente mostrare che

$$\varphi_{t_n}(\lambda) = \varphi_{s_n}\left(\frac{\lambda}{d_n}\right) = \prod_{i=1}^n \varphi_{\xi_i}\left(\frac{\lambda}{d_n}\right) \to e^{-\lambda^2/2}, \quad n \to \infty.$$

A tale scopo consideriamo le rappresentazioni

$$e^{iz} = 1 + iz + \frac{\vartheta_1 z^2}{2!}, \quad e^{iz} = 1 + iz - \frac{z^2}{2!} + \frac{\vartheta_2 |z|^3}{3!},$$

valide per ogni $z \in I\!R$ e $0 \leq |\vartheta_1(z)|, |\vartheta_2(z)| \leq 1$. Usando tali rappresentazioni e il fatto che $m_i = \int x dF_i(x) = 0$ otteniamo

$$\begin{aligned} \varphi_{\xi_i}(\lambda) &= \int_R e^{i\lambda x} dF_i(x) \\ &= 1 + \frac{\lambda^2}{2} \int_{|x| \geq \epsilon d_n} \vartheta_1 x^2 dF_i - \frac{\lambda^2}{2} \int_{|x| < \epsilon d_n} x^2 dF_i + \frac{|\lambda|^3}{6} \int_{|x| < \epsilon d_n} \vartheta_2 |x|^3 dF_i. \end{aligned}$$

Inoltre si ha

$$\frac{1}{2} \int_{|x| \geq \epsilon d_n} \vartheta_1 x^2 dF_i = \overline{\vartheta}_1 \int_{|x| \geq \epsilon d_n} x^2 dF_i$$

per qualche $\overline{\vartheta}_1 = \overline{\vartheta}_1(\lambda, i, n)$ con $|\overline{\vartheta}_1| \leq 1/2$; e similmente

$$\frac{1}{6} \int_{|x| < \epsilon d_n} \vartheta_2 |x|^3 dF_i = \overline{\vartheta}_2 \int_{|x| < \epsilon d_n} \epsilon d_n x^2 dF_i$$

con $|\overline{\vartheta}_2| \leq 1/6$. Poniamo anche

$$A_{in} = \frac{1}{d_n^2} \int_{|x|<\epsilon d_n} x^2 dF_i, \quad B_{in} = \frac{1}{d_n^2} \int_{|x|\geq \epsilon d_n} x^2 dF_i,$$

cosicchè $A_{in} < \epsilon^2$,

$$\sum_{i=1}^{n} (A_{in} + B_{in}) = 1,$$

e, per la condizione di Lindeberg,

$$\sum_{i=1}^{n} B_{in} \to 0, \quad n \to \infty.$$

Mettendo tutto insieme abbiamo trovato che

$$\varphi_{\xi_i}\left(\frac{\lambda}{d_n}\right) = 1 - \frac{\lambda^2 A_{in}}{2} + \lambda^2 \overline{\vartheta}_1 B_{in} + \epsilon\, |\lambda|^3\, \overline{\vartheta}_2 A_{in} =: 1 + C_{in},$$

dove, per n abbastanza grande,

$$\max_{1\leq i\leq n} |C_{in}| \leq \lambda^2 \epsilon^2 + \epsilon |\lambda|^3$$

e

$$\sum_{i=1}^{n} |C_{in}| \leq \lambda^2 + \epsilon |\lambda|^3.$$

Ricordiamo ora che per ogni $z \in I\!R$ con $|z| < 1/2$ si ha che

$$\ln(1+z) = \sum_{n\geq 1} (-1)^{n-1} \frac{z^n}{n} = z + \vartheta \cdot z^2$$

dove $\vartheta = \vartheta(z)$ soddisfa $|\vartheta(z)| \leq 1$. Dunque si ha, per ogni λ fissato, n abbastanza grande ed ϵ abbastanza piccolo,

$$\ln \varphi_{\xi_i}\left(\frac{\lambda}{d_n}\right) = \ln(1 + C_{in}) = C_{in} + \vartheta_{in} C_{in}^2, \quad |\vartheta_{in}| \leq 1,$$

e quindi

$$\ln \varphi_{t_n}(\lambda) = \sum_{i=1}^{n} \ln \varphi_{\xi_i}(\lambda) = \sum_{i=1}^{n} C_{in} + \sum_{i=1}^{n} \vartheta_{in} C_{in}^2.$$

Ora

$$\frac{\lambda^2}{2}+\sum_{i=1}^{n} C_{in} = \frac{\lambda^2}{2}\left(1-\sum_{i=1}^{n} A_{in}\right)+\lambda^2\sum_{i=1}^{n}\overline{\vartheta}_1(\lambda,i,n)B_{in}+\epsilon|\lambda|^3\sum_{i=1}^{n}\overline{\vartheta}_2(\lambda,i,n)A_{in},$$

e per quanto visto sopra, per ogni $\delta > 0$ possiamo trovare n_0 e $\epsilon > 0$ in modo che per ogni $n \geq n_0$ si ha

$$\left|\frac{\lambda^2}{2}+\sum_{i=1}^{n} C_{in}\right| \leq \frac{\delta}{2},$$

e, nello stesso tempo,

$$\left|\sum_{i=1}^{n}\vartheta_{in}C_{in}^2\right| \leq \max_{1\leq i\leq n}|C_{in}|\sum_{i=1}^{n}|C_{in}| \leq (\lambda^2\epsilon^2+\epsilon|\lambda|^3)(\lambda^2+\epsilon|\lambda|^3) \leq \frac{\delta}{2},$$

e infine

$$\left|\frac{\lambda^2}{2}+\ln\varphi_{t_n}(\lambda)\right| \leq \delta.$$

Quindi, per ogni $\lambda \in I\!R$,

$$\varphi_{t_n}(\lambda)\, e^{\lambda^2/2} \to 1, \quad n\to\infty, \qquad \text{q.e.d.}$$

7.35 Osservazione. Vediamo alcuni casi speciali in cui la condizione di Lindeberg è soddisfatta:

a) supponiamo che sia soddisfatta la *condizione di Lyapunov*: per qualche $\delta > 0$

$$\frac{1}{d_n^{2+\delta}}\sum_{i=1}^{n} \mathrm{E}\,|\xi_i-m_i|^{2+\delta} \to 0, \quad n\to\infty.$$

Sia ora dato $\epsilon > 0$. Allora si ha

$$\begin{aligned}\mathrm{E}\,|\xi_i-m_i|^{2+\delta} &= \int_{I\!R}|x-m_i|^{2+\delta}dF_i(x)\\ &\geq \int_{\{x:|x-m_i|\geq\epsilon\, d_n\}}|x-m_i|^{2+\delta}dF_i(x)\\ &\geq \epsilon^\delta\, d_n^\delta\int_{\{x:|x-m_i|\geq\epsilon\, d_n\}}(x-m_i)^2dF_i(x),\end{aligned}$$

e dunque

$$\frac{1}{d_n^2}\sum_{i=1}^n \int_{\{x:|x-m_i|\geq \epsilon d_n\}} (x-m_i)^2 dF_i(x) \leq \frac{1}{\epsilon^\delta}\frac{1}{d_n^{2+\delta}}\sum_{i=1}^n \mathrm{E}\,|\xi_i - m_i|^{2+\delta}.$$

E dunque la condizione di Lyapunov implica quella di Lindeberg.

b) Siano $\xi_1, \xi_2 \ldots$ v.a.i.i.d. con media $m = \mathrm{E}\,\xi_1$ e varianza $0 < \sigma^2 = \mathrm{Var}\,\xi_1 < \infty$. Allora

$$\frac{1}{d_n^2}\sum_{i=1}^n \int_{\{x:|x-m|\geq \epsilon d_n\}} (x-m_i)^2 dF_i(x) = \frac{n}{n\sigma^2}\int_{\{x:|x-m|\geq \epsilon\sigma\sqrt{n}\}} |x-m|^2 dF_1(x) \to 0$$

perchè $\{x : |x-m| \geq \epsilon\sigma\sqrt{n}\} \downarrow \emptyset$ quando $n \to \infty$. Pertanto la condizione di Lindeberg è soddisfatta e dunque (7.31) segue da (7.32).

c) Siano $\xi_1, \xi_2 \ldots$ v.a. indipendenti, uniformemente limitate, cioè tali che

$$|\xi_i| \leq K < \infty, \quad \forall i,$$

e che inoltre valga $d_n \to \infty$, $n \to \infty$. Allora si ha, usando la disuguaglianza di Chebyshev,

$$\begin{aligned}\int_{\{x:|x-m|\geq \epsilon d_n\}} (x-m_i)^2 dF_i(x) &= \mathrm{E}\,\left[(\xi_i - m_i)^2\, 1_{\{|\xi_i - m_i| \geq \epsilon\, d_n\}}\right] \\ &\leq (2K)^2 \mathrm{P}\,(|\xi_i - m_i| \geq \epsilon\, d_n) \leq (2K)^2 \frac{\sigma_i^2}{\epsilon^2 d_n^2},\end{aligned}$$

e quindi, ancora una volta, la condizione di Lindeberg è soddisfatta.

7.36 Osservazione. Se poniamo

$$\Phi(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^x e^{-x^2/2} dx \tag{7.37}$$

e $t_n = (s_n - \mathrm{E}\,s_n)/d_n$, allora (7.34) implica che per ogni $x \in I\!R$,

$$F_{t_n}(x) \to \Phi(x), \quad n \to \infty. \tag{7.38}$$

Dal momento che $\Phi(x)$ è continua la convergenza è in realtà uniforme:

$$\sup_{x \in I\!R} |F_{t_n}(x) - \Phi(x)| \to 0, \quad n \to \infty. \tag{7.39}$$

In particolare ciò implica che

$$\mathrm{P}\,(s_n \le x) \to \Phi\left(\frac{x - \mathrm{E}\,s_n}{d_n}\right), \quad n \to \infty. \tag{7.40}$$

Ciò si esprime talvolta dicendo che *per n abbastanza grande s_n è distribuita in modo approssimativamente gaussiano con media $\mathrm{E}\,s_n$ e varianza d_n^2*.

Una stima della velocità con cui il limite (7.39) è raggiunto si può ottenere sotto l'ipotesi che le ξ_i siano v.a. indipendenti identicamente e uniformemente distribuite con $\mathrm{E}\,|\xi_1|^3 < \infty$. In questo caso si ha la *disuguaglianza di Berry-Esseen*:

$$\sup_{x\in\mathbb{R}} |F_{t_n}(x) - \Phi(x)| \le C\,\frac{\mathrm{E}\,|\xi_1 - \mathrm{E}\,\xi_1|^3}{\sigma^3\sqrt{n}}. \tag{7.41}$$

7.42 Altre distribuzioni limite. In alcune situazioni è necessario considerare distribuzioni che decadono così lentamente che $\int_{\mathbb{R}} x^2\,dF(x) = \infty$. In questo caso si ottengono distribuzioni limite di natura diversa. Ad esempio se $\xi_1, \xi_2 \ldots$ sono v.a.i.i.d. distribuite con densità $p(x)$ tale che $p(x) = p(-x)$ e $p(x) \sim c_1/|x|^{a+1}$ per qualche costante $c_1 > 0$ e $0 < a < 2$, allora se poniamo $\eta_n = s_n/n^{1/a}$ (cfr. (6.39)), la sua funzione caratteristica $\varphi_{\eta_n}(\lambda)$ converge per ogni $\lambda \in \mathbb{R}$ ad una funzione caratteristica della forma $e^{-c_2\,|\lambda|^a}$, per qualche costante $c_2 > 0$.