Università di Roma - Tor Vergata

Corso di Laurea in Matematica

Appunti del corso di

Calcolo delle Probabilità

II modulo, Anno 2001-2002

Paolo Baldi
Dipartimento di Matematica
baldi@mat.uniroma2.it

maggio 2002

Indice

1. Cenni di teoria della misura 1

1.1 Spazi misurabili, funzioni misurabili . . . 1
1.2 Misure . . . 4
1.3 Integrazione . . . 6
1.4 Esempi . . . 11
1.5 Misure prodotto . . . 13
Esercizi . . . 15

2. Variabili aleatorie 17

2.1 Probabilità e misura . . . 17
2.2 Indipendenza . . . 18
2.3 Disuguaglianze di convessità, momenti, covarianza . . . 22
2.4 Funzioni caratteristiche, trasformata di Laplace . . . 28
2.5 Leggi normali multivariate . . . 40
2.6 Statistica dei modelli gaussiani . . . 45
2.7 Leggi condizionali . . . 48
Esercizi . . . 52

3. Convergenza e approssimazione 61

3.1 Il Lemma di Borel-Cantelli . . . 61
3.2 La convergenza quasi certa . . . 63
3.3 Le leggi forti dei grandi numeri . . . 66
3.4 Convergenza in legge . . . 68
3.5 Il teorema limite centrale, il test del χ^2 . . . 77
3.6 Il lemma di Slutski . . . 83

Esercizi 85

4. Problemi 95

4.1 Problemi al capitolo 1 95
4.2 Problemi al capitolo 2 96
4.3 Problemi al capitolo 3 97
4.4 Soluzioni 98

Indice analitico 109

1
Cenni di teoria della misura

1.1 Spazi misurabili, funzioni misurabili

Siano E un insieme e $\mathcal{E} \subset \mathcal{P}(E)$ ($\mathcal{P}(E)$=la famiglia di tutti i sottoinsiemi di E). Si dice che $\mathcal{E}$ è un'algebra (risp. una σ-algebra) se $E \in \mathcal{E}$, se $\mathcal{E}$ è stabile per passaggio al complementare e per unioni e intersezioni finite (risp. numerabili). La coppia $(E, \mathcal{E})$, $\mathcal{E}$ σ-algebra su E, si chiama uno *spazio misurabile*. Osserviamo che $\mathcal{P}(E)$ è una σ-algebra e che l'intersezione di una famiglia qualunque di σ-algebre è una σ-algebra. Dunque, data una classe di insiemi $\mathcal{C} \subset \mathcal{P}(E)$, si può considerare la più piccola σ-algebra contenente $\mathcal{C}$: è l'intersezione di tutte le σ-algebre contenenti $\mathcal{C}$. Questa σ-algebra si indica $\sigma(\mathcal{C})$ e si chiama la *σ-algebra generata* da $\mathcal{C}$.

Supponiamo $E = \mathbb{R}^d$ e sia $\mathcal{O}$ la classe degli aperti di E. La σ-algebra $\sigma(\mathcal{O})$ si chiama la *σ-algebra di Borel* di $\mathbb{R}^d$ e si indica $\mathcal{B}(\mathbb{R}^d)$. È facile vedere che essa è anche generata dai chiusi, dalle palle, dai plurirettangoli e anche dai plurirettangoli a coordinate razionali (quest'ultima famiglia ha il vantaggio di essere numerabile).

Tutti queste affermazioni si dimostrano ripetendo un tipico ragionamento di teoria della misura. Ad esempio, indichiamo con $\mathcal{C}$ la famiglia delle palle aperte di centro x e raggio r, al variare di $x \in \mathbb{R}^d$ e di $r > 0$. Mostriamo che $\sigma(\mathcal{C}) = \mathcal{B}(\mathbb{R}^d)$. Intanto $\sigma(\mathcal{C}) \subset \mathcal{B}(\mathbb{R}^d)$, perché $\mathcal{B}(\mathbb{R}^d)$ è una σ-algebra contenente $\mathcal{C}$, mentre $\sigma(\mathcal{C})$ è la più piccola σ-algebra contenente $\mathcal{C}$. D'altra parte $\sigma(\mathcal{C})$ contiene gli aperti di $\mathbb{R}^d$, poiché ogni aperto di $\mathbb{R}^d$ si può scrivere come riunione numerabile di palle aperte. Dunque $\sigma(\mathcal{C})$ contiene $\mathcal{B}(\mathbb{R}^d)$, che è la più piccola σ-algebra contenente gli aperti.

Più in generale, se E è uno spazio topologico, la σ-algebra di Borel $\mathcal{B}(E)$ è la più piccola σ-algebra contenente gli aperti (ovvero la più piccola σ-algebra contenente i chiusi). Se $d = 1$, si possono quindi considerare le σ-algebre $\mathcal{B}(\mathbb{R}^+) = \{A \in \mathcal{B}(\mathbb{R}), A \subset \mathbb{R}^+\}$, $\mathcal{B}(\overline{\mathbb{R}}) = \sigma(\mathcal{B}(\mathbb{R}), \{+\infty\}, \{-\infty\})$ e $\mathcal{B}(\overline{\mathbb{R}^+}) = \sigma(\mathcal{B}(\mathbb{R}^+), \{+\infty\})$.

Si chiama *classe monotona* una famiglia $\mathcal{M}$ di parti di E tale che

- $E \in \mathcal{M}$,

- se $A, B \in \mathcal{M}$ e $A \subset B$, allora $B \setminus A \in \mathcal{M}$.
- $\mathcal{M}$ sia stabile per limite crescente: se $(A_n)_n \subset \mathcal{M}$ è una successione crescente d'insiemi, allora $A = \bigcup A_n \in \mathcal{M}$.

Il risultato seguente, chiamato il *teorema delle classi monotone* , sarà di uso costante nel seguito.

Teorema 1.1 *Sia $\mathcal{C} \subset \mathcal{P}(E)$ una famiglia d'insiemi stabile per intersezioni finite e sia $\mathcal{M}$ una classe monotona contenente $\mathcal{C}$. Allora $\mathcal{M}$ contiene $\sigma(\mathcal{C})$.*

Ad esempio, gli intervalli di $\mathbb{R}$ (o anche gli intervalli di $\mathbb{R}$ della forma $]a, b]$, cioè aperti a sinistra e chiusi a destra) cosituiscono una famiglia stabile per l'intersezione finita. Dunque la più piccola classe monotona contenente gli intervalli contiene anche la σ-algebra boreliana.

Siano $(E_1, \mathcal{E}_1)$ e $(E_2, \mathcal{E}_2)$ due spazi misurabili. Un'applicazione di E_1 in E_2 si dice *misurabile* se, per ogni $A \in \mathcal{E}_2$, $f^{-1}(A) \in \mathcal{E}_1$. È immediato verificare che l'applicazione composta di due applicazioni misurabili è misurabile.

È facile vedere che perche f sia misurabile basta che sia $f^{-1}(A) \in \mathcal{E}_1$ per ogni $A \in \mathcal{C}$, dove $\mathcal{C}$ è una classe d'insiemi tale che $\sigma(\mathcal{C}) = \mathcal{E}_2$ (in esercizio: basta verificare che la classe degli insiemi $A \subset E_2$ tali che $f^{-1}(A) \in \mathcal{E}_1$ è una σ-algebra). Questo è un criterio di misurabilità molto utile, perché spesso si conoscono esplicitamente gli insiemi di una classe $\mathcal{C}$ che genera $\mathcal{E}_2$, ma non quelli di $\mathcal{E}_2$. Nel caso che $\mathcal{E}_2$ sia la σ-algebra di uno spazio topologico E_2, per verificare la misurabilità di f basterà dunque verificare che $f^{-1}(A) \in \mathcal{E}_1$ per ogni insieme A aperto (risp. chiuso).

In particolare, se f è continua da $\mathbb{R}^d$ in $\mathbb{R}^m$, o più in generale da uno spazio topologico E in uno spazio topologico F, f è misurabile per le σ-algebre boreliane.

Quando lo spazio d'arrivo è $\mathbb{R}$, $\overline{\mathbb{R}}$, $\overline{\mathbb{R}^+}$, $\mathbb{R}^d$, $\mathbb{C}$, sottintenderemo sempre che esso è munito della sua σ-algebra boreliana.

Sia $(E, \mathcal{E})$ un spazio misurabile. Perché un'applicazione numerica (cioè a valori $\overline{\mathbb{R}}$) sia misurabile basta che, per ogni $a \in \mathbb{R}$, si abbia $\{f > a\} = \{x,\ f(x) > a\} = f^{-1}(]a, +\infty[) \in \mathcal{E}$ (in esercizio: basta mostrare che gli insiemi della forma $]a, +\infty[$ generano la σ-algebra di Borel). Si possono anche considerare gli insiemi della forma $\{f < a\}$, $\{f \le a\}$, $\{f \ge a\}$. Da questi criteri si ricava facilmente che, se f, g, f_n sono funzioni numeriche misurabili, lo stesso vale per $-f$, $\sup(f, g)$, $\inf(f, g)$, $f^+ = \sup(f, 0)$, $f^- = \sup(-f, 0)$, $\sup f_n$, $\inf f_n$. Ricordiamo che

$$(1.1) \qquad \overline{\lim_{n\to\infty}} f_n(x) = \lim_{n\to\infty} \downarrow \sup_{k\ge n} f_k(x), \qquad \underline{\lim_{n\to\infty}} f_n(x) = \lim_{n\to\infty} \uparrow \inf_{k\ge n} f_k(x),$$

dove queste quantità sono a valori $\overline{\mathbb{R}}$ e $f = \lim_{n\to\infty} f_n$ se e solo se $\overline{\lim} f_n = \underline{\lim}_{n\to\infty} f_n = f$. Ne segue che, se le funzioni f_n sono misurabili, anche $\overline{\lim}_{n\to\infty} f_n$, $\underline{\lim}_{n\to\infty} f_n$, $\lim_{n\to\infty} f_n$ (se il limite esiste) sono funzioni misurabili.

Siano f_1, f_2 applicazioni reali misurabili definite sullo spazio misurabile $(E, \mathcal{E})$. Allora l'applicazione $f = (f_1, f_2)$ è misurabile a valori in $(\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2))$. Infatti, se A_1, A_2

sono intervalli aperti, allora $f^{-1}(A_1 \times A_2) = f_1^{-1}(A_1) \cap f_2^{-1}(A_2) \in \mathcal{E}$. Poiché, come abbiamo visto prima, i rettangoli della forma $A_1 \times A_2$ generano $\mathcal{B}(\mathbb{R}^2)$, f è dunque misurabile.

Poiché l'applicazione $(x, y) \to x + y$ è continua da $\mathbb{R}^2$ in $\mathbb{R}$, essa è anche misurabile. Ne segue che l'applicazione $f_1 + f_2$ è anch'essa misurabile come composizione di applicazioni misurabili. Allo stesso modo si dimostra che sono misurabili le applicazioni $f_1 f_2$ e $\frac{f_1}{f_2}$ (se definita). Risultati simili valgono se f_1 e f_2 sono applicazioni numeriche.

Questi esempi suggeriscono che, di solito, per dimostrare la misurabilità di un'applicazione f non si cercherà mai (o quasi) di verificare la definizione. Si cercherà piuttosto di applicare il criterio studiando $f^{-1}(A)$ per A in una classe d'insiemi che genera la σ-algebra dello spazio d'arrivo, oppure cercando di dimostrare che f è somma, prodotto, limite, ... di funzioni misurabili. Un po' in questo spirito è la prossima Proposizione 1.3. Ricordiamo che, scrivendo $f_n \uparrow f$ (risp. $f_n \downarrow f$), indichiamo che, $f_n(x)$ cresce (risp. decresce) a $f(x)$ per ogni $x \in E$. Se $A \subset B$, si chiama *funzione indicatrice di A* e si scrive 1_A, la funzione che vale 1 su A e 0 su A^c. Si ha

$$1_{A^c} = 1 - 1_A,\ 1_{\cap A_n} = \Pi 1_{A_n} = \inf 1_{A_n}, \quad 1_{\cup A_n} = \sup 1_{A_n}.$$

Un'applicazione f di $(E, \mathcal{E})$ in $\mathbb{R}$ si dice *elementare* se la si può scrivere nella forma $f = \sum_{k=1}^n a_k 1_{A_k}$, $A_k \in \mathcal{B}$. Indicheremo
$b\mathcal{E}$ l'insieme delle funzioni reali misurabili limitate,
$\mathcal{E}^+$ l'insieme delle funzioni misurabili positive, cioè a valori $\overline{\mathbb{R}^+}$ (possono quindi anche prendere il valore $+\infty$),

Il risultato seguente è fondamentale, perché permette di approssimare le funzioni misurabili positive con funzioni elementari. Ce ne serviremo spesso.

Proposizione 1.2 *Ogni $f \in \mathcal{E}^+$ è limite di une successione crescente di funzioni di e$\mathcal{E}^+$.*

Dimostrazione. Basta considerare

$$f_n(x) = \sum_{k=0}^{n2^n - 1} \frac{k}{2^n} 1_{\{\frac{k}{2^n} \le f(x) < \frac{k+1}{2^n}\}} + n 1_{\{f(x) > n\}}. \tag{1.2}$$

È chiaro infatti che la successione $(f_n)_n$ è crescente e che $f(x) - \frac{1}{2^n} \le f_n(x) \le f(x)$ se $f(x) \le n$.

Sia f una applicazione di E in un spazio misurabile $(A, \mathcal{A})$. Si nota $\sigma(f)$ e si chiama *σ-algebra generata da f* la più piccola σ-algebra su E che renda f misurabile, cioè tale che $f : (E, \sigma(f)) \to (A, \mathcal{A})$ sia misurabile. È facile vedere che $\sigma(f) = \{f^{-1}(A),\ A \in \mathcal{A}\}$.

Proposizione 1.3 *Sia $h : E \to \mathbb{R}$ (risp. $E \to \overline{\mathbb{R}^+}$). Allora h è $\sigma(f)$-misurabile se e solo se esiste $g \in f\mathcal{A}$ (risp. $g \in \mathcal{A}^+$) tale che $h = g \circ f$.*

Dimostrazione. Evidentemente se $h = g \circ f$, h è $\sigma(f)$-misurabile come composizione di applicazioni misurabili. Viceversa supponiamo dapprima che h sia $\sigma(f)$-misurabile positiva ed elementare. Si ha allora $h = \sum_{k=1}^{n} a_k 1_{B_k}$ con $B_k \in \sigma(f)$ e dunque $B_k = f^{-1}(A_k)$ per qualche $A_k \in \mathcal{A}$. Dato che $1_{B_k} = 1_{A_k} \circ f$, $h = g \circ f$ con $g = \sum_{k=1}^{n} a_k 1_{A_k}$. Se lasciamo cadere l'ipotesi che h sia elementare e consideriamo $h \in \sigma(f)^+$, si ha $h = \lim_{n\to\infty} \uparrow h_n$ con $\underline{h_n} \in$ e $[\sigma(f)]^+$ e dunque $h_n = g_n \circ f$, $g_n \in \mathcal{A}^+$. Se ne deduce che $h = g \circ f$ con $g = \overline{\lim}_{n\to\infty} g_n \in \mathcal{A}^+$. Se $h \in f[\sigma(f)]$, $h = h^+ - h^-$ e $h^+ = g_1 \circ f$, $h^- = g_2 \circ f$ con $g_i \in \mathcal{A}^+$. Si ha allora $h = g \circ f$ con $g = g_1 1_{\{g_1 < +\infty\}} - g_2 1_{\{g_2 < +\infty\}} \in f\mathcal{A}$.

Più in generale se $(f_i,\ i \in I)$ è una famiglia di applicazioni di E a valori rispettivamente negli spazi misurabili $(A_i, \mathcal{A}_i)$, si indica $\sigma(f_i,\ i \in I)$ e si chiama σ-algebra generata dalle f_i, la più piccola σ-algebra su E che renda misurabili tutte le f_i. Si ha dunque

$$\sigma(f_i,\ i \in I) = \sigma(f_i^{-1}(A_i),\ A_i \in \mathcal{A}_i,\ i \in I).$$

1.2 Misure

Sia $(E, \mathcal{E})$ uno spazio misurabile.

Definizione 1.4 *Si chiama misura su $(E, \mathcal{E})$ un'applicazione μ da $\mathcal{E}$ in $\overline{\mathbb{R}^+}$ tale che*

i) $\mu(\emptyset) = 0$,

ii) per ogni successione $(A_n)_n \subset \mathcal{E}$ d'insiemi a due a due disgiunti, $\mu(\bigcup_{n\ge1} A_n) = \sum_{n=1}^{\infty} \mu(A_n)$.

La tripla $(E, \mathcal{E}, \mu)$ si chiama uno spazio di misura.

Le proprietà seguenti sono immediate.

i) Se $A, B \in \mathcal{E}$, $A \subset B$, allora $\mu(A) \le \mu(B)$.

ii) Se $(A_n)_n \subset \mathcal{E}$, $\mu(\bigcup_{n\ge1} A_n) \le \sum_{n=1}^{\infty} \mu(A_n)$.

iii) Se $A_n)_n \subset \mathcal{E}$ e se $A_n \uparrow A$ (i.e. $1_{A_n} \uparrow 1_A$), $\mu(A_n) \uparrow \mu(A)$.

iv) Se $A_n)_n \subset \mathcal{E}$, se $A_n \downarrow A$ (i.e. $1_{A_n} \downarrow 1_A$) e se, per qualche n_0, allora $\mu(A_{n_0}) < +\infty$, $\mu(A_n) \downarrow \mu(A)$.

Se $\mu(E) < +\infty$, la misura μ si dice *finita*. Se $E = \bigcup_n E_n$ con $E_n \in \mathcal{E}$ e $\mu(E_n) < +\infty$, μ si dice *σ-finita*. Se $\mu(E) = 1$, μ si chiama una (misura di) *probabilità*.

Osservazione 1.5 La proprietà ii) della Definizione 1.4 si chiama *σ-additività*. Se nella Definizione 1.4 si suppone che $\mathcal{E}$ sia solo un'algebra, la definizione conserva un significato aggiungendo in ii) la condizione $\bigcup_n A_n \in \mathcal{E}$. Si ha allora la nozione di *misura su un'algebra*.

Proposizione 1.6 *Siano μ e ν due misure su $(E, \mathcal{E})$ e $\mathcal{C} \subset \mathcal{E}$ una classe d'insiemi stabile per intersezioni finite. Si suppone che, per ogni $A \in \mathcal{C}$, $\mu(A) = \nu(A) < +\infty$ e che $E = \lim_{n\to\infty} \uparrow E_n$ con $E_n \in \mathcal{C}$. Allora $\mu(A) = \nu(A)$ per ogni $A \in \sigma(\mathcal{C})$.*

Dimostrazione. Supponiamo dapprima $\mu(E) = \nu(E) < +\infty$. Sia $\mathcal{M} = \{A \in \mathcal{E},\ \mu(A) = \nu(A)\}$. Si verifica immediatamente che $\mathcal{M}$ è una classe monotona e le ipotesi del Teorema 1.1 sono verificate e dunque $\sigma(\mathcal{C}) \subset \mathcal{M}$. Il caso generale se tratta applicando questo risultato alle misure $\mu_n(A) = \mu(A \cap E_n)$ e $\nu_n(A) = \nu(A \cap E_n)$.

Osservazione 1.7 Se $\mu(E) = \nu(E) < +\infty$, l'enunciato della Proposizione 1.6 si semplifica: se μ e ν coincidono su una classe $\mathcal{C}$ stabile per intersezioni finite e che genera $\mathcal{E}$, esse sono uguali su $\mathcal{E}$.

Un problema interessante della teoria della misura è quello di costruire delle misure che soddisfino a particolari proprietà. Ad esempio che prendano dei valori assegnati su certe classi d'insiemi. Lo strumento chiave è il teorema seguente.

Teorema 1.8 *(Carathéodory) Sia μ una misura su una algebra $\mathcal{A}$. Allora μ si prolunga ad una misura su $\sigma(\mathcal{A})$. Per di più, se μ è σ-finita, questo prolungamento è unico.*

Una *misura di Borel* su uno spazio topologico E è una misura su $(E, \mathcal{B}(E))$ tale che $\mu(K) < +\infty$ per ogni compatto K.

Ci interessiamo ora alle misure di Borel su $\mathcal{B}(\mathbb{R})$. Osserviamo, per prima cosa, che $\mathcal{C} = \{\,]a, b], -\infty < a < b < +\infty\}$ è una classe stabile per intersezioni finite e che $\sigma(\mathcal{C}) = \mathcal{B}(\mathbb{R})$. Segue allora dalla Proposizione 1.6 che una misura su $\mathcal{B}(\mathbb{R})$, finita sugli intervalli limitati, è determinata dai valori di $\mu(]a, b])$ $a, b \in \mathbb{R}, a < b$. Poi, data una tale misura, se si pone

$$F(0) = 0;\ F(x) = \mu(]0, x]), x > 0;\ F(x) = -\mu(]x, 0]), x < 0,$$

$F(x)$ è una funzione continua a destra e crescente e si ha

$$\mu(]a, b]) = F(b) - F(a). \tag{1.3}$$

Viceversa si è condotti al problema seguente. Sia F un'applicazione di $\mathbb{R}$ in $\mathbb{R}$ continua a destra e crescente, esisterà una misura μ su $\mathcal{B}(\mathbb{R})$ tale che $\mu(]a, b]) = F(b) - F(a)$?

Cerchiamo di applicare il Teorema 1.8 di Carathéodory. Consideriamo la famiglia d'insiemi $\mathcal{C}$ formata dagli intervalli semiaperti $]a, b], a < b$. È facile descrivere l'algebra $\mathcal{A}$ generata da $\mathcal{C}$,

$$\mathcal{A} = \{A = \textstyle\bigcup_{k=1}^n]a_k, b_k], -\infty \le a_1 < b_1 < a_2 < \ldots < b_{n-1} < a_n < b_n \le +\infty\}$$

con la convenzione che, se $b_n = +\infty$, $]a_n, b_n] =]a_n, +\infty[$. Si vede subito che $\mathcal{A}$ è un'algebra contenente $\mathcal{C}$. Si definisce μ su $\mathcal{A}$ mediante $\mu(A) = \sum_{k=1}^n F(b_k) - F(a_k)$, dove $F(+\infty) = \lim_{x\to+\infty} F(x)$, $F(-\infty) = \lim_{x\to-\infty} F(x)$. È facile dimostrare che μ è additiva su $\mathcal{A}$; un po' più delicato è dimostrare che μ è σ-additiva su $\mathcal{A}$ e tralasceremo la dimostrazione di questo punto. Poiché $\sigma(\mathcal{A}) = \mathcal{B}(\mathbb{R})$, per il Teorema 1.8 di Carathéodory abbiamo dunque dimostrato:

Teorema 1.9 *Sia F un'applicazione di $\mathbb{R}$ in $\mathbb{R}$ continua a destra e crescente. Esiste una ed una sola misura μ su $\mathcal{B}(\mathbb{R})$ tale che, per ogni $a < b$, $\mu(]a, b]) = F(b) - F(a)$.*

Se si sceglie $F(x) = x$, si ottiene l'esistenza e l'unicità di una misura λ sur $\mathcal{B}(\mathbb{R})$ tale che, per ogni intervallo I, si abbia $\lambda(I) = |I|$. Questa misura si chiama la *misura di Lebesgue* su $\mathbb{R}$.

Sia $(E, \mathcal{E}, \mu)$ uno spazio di misura. Un sottoinsieme A di E si dice *trascurabile*(μ-trascurabile se c'è pericolo di ambiguità) se $A \subset B$ con $B \in \mathcal{E}$ e $\mu(B) = 0$. In particolare un insieme trascurabile può non essere misurabile. Si dice che una proprietà è vera quasi ovunque (q.o.) se è vera al di fuori di un insieme trascurabile. Per esempio $f = g$ q.o. significa che $\{x \in E, f(x) \neq g(x)\}$ è trascurabile. Se μ è una probabilità, si dice quasi certamente (q.c.) invece di quasi ovunque. Si indica $\mathcal{N}$ la classe degli insiemi trascurabili. Osserviamo che se $A_n \in \mathcal{N}$, $\bigcup_n A_n \in \mathcal{N}$. Se $\mathcal{N} \subset \mathcal{E}$, lo spazio di misura $(E, \mathcal{E}, \mu)$ si dice completo.

1.3 Integrazione

Sia $(E, \mathcal{E}, \mu)$ uno spazio di misura.

Costruiamo per prima cosa l'integrale di f rispetto a μ quando $f \in \mathcal{E}^+$. Se f è elementare positiva, ciò è molto facile; f è della forma $f = \sum_{k=1}^n a_k 1_{A_k}$, con $A_k \in \mathcal{E}$ e $\alpha_k \geq 0$ e si pone

$$\int f \, d\mu = \sum_{k=1}^{n} a_k \mu(A_k).$$

Da considerazioni elementari si vede che questo numero (che può essere $= +\infty$) non dipende dalla rappresentazione di f (i numeri a_k e gl'insiemi A_k non sono unici). Inoltre, se f, g sono elementari positive e $a, b \in \mathbb{R}^+$, si ha

- $\int (af + bg) \, d\mu = a \int f \, d\mu + b \int g \, d\mu$,
- se $f \leq g$,$\int f \, d\mu \leq \int g \, d\mu$. Si ha anche il risultato più tecnico seguente che è la chiave di volta della costruzione.

Lemma 1.10 *Se $(f_n)_n$, $(g_n)_n$ sono successioni crescenti di funzioni elementari $\mathcal{E}$-misurabili positive e se $\lim_{n\to\infty} \uparrow f_n = \lim_{n\to\infty} \uparrow g_n$, allora $\lim_{n\to\infty} \uparrow \int f_n \, d\mu = \lim_{n\to\infty} \uparrow \int g_n \, d\mu$.*

Sia f una funzione $\mathcal{E}$-misurabile positiva. Esiste (Proposizione 1.2) una successione $(f_n)_n$ di funzioni $\mathcal{E}$-misurabili positive elementari tale che $f_n \uparrow f$; allora $\int f_n \, d\mu \uparrow$ è una successione crescente e si pone $\int f \, d\mu = \lim_{n\to\infty} \uparrow \int f_n \, d\mu$. Il punto importante è che, grazie al Lemma 1.10, questo limite non dipende dalla particolare successione f_n prescelta. Passando al limite, si ottiene immediatamente che, per $f, g \in \mathcal{E}^+$ e $a, b \in \mathbb{R}^+$ si ha

- $\int (af + bg) \, d\mu = a \int f \, d\mu + b \int g \, d\mu$;
- se $f \leq g$, $\int f \, d\mu \leq \int g \, d\mu$.

Per definire l'integrale di una funzione numerica $\mathscr{E}$-misurabile, basta scrivere $f = f^+ - f^-$ dove $f^+ = f \vee 0$ e $f^- = -f \wedge 0$ (le *parti positiva e negativa* di f). Si può allora porre

$$\int f\,d\mu = \int f^+\,d\mu - \int f^-\,d\mu$$

a condizione che una almeno della quantità $\int f^+\,d\mu$ e $\int f^-\,d\mu$ siano finite. f si dice

- semi-integrabile inferiormente (s.i.i.) se $\int f^-\,d\mu < +\infty$. In questo caso l'integrale di f è ben definito (ma può prendere il valore $+\infty$).
- semi-integrabile superiormente (s.i.s.) se $\int f^+\,d\mu < +\infty$. In questo caso l'integrale di f è ben definito (ma può prendere il valore $-\infty$).
- integrabile se f^+ e f^- hanno entrambe integrale finito.

È chiaro che una funzione positiva è sempre s.i.i. (dato che $f^- = 0$) ed una negativa è sempre s.i.s. È facile vedere che f è integrabile se e solo se $\int |f|\,d\mu < +\infty$. In ogni caso, se f è semi-integrabile, si ha

$$\Big|\int f\,d\mu\Big| \le \int |f|\,d\mu \tag{1.4}$$

L'integrale si definisce immediatamente anche per le funzioni a valori complessi. Se $f : E \to \mathbb{C}$, e scriviamo $f = f_1 + if_2$, allora si vede subito che se $\int |f|\,d\mu < +\infty$ (ora $|\ |$ indica il modulo complesso), allora sia f_1 che f_2 sono integrabili. Si pone allora

$$\int f\,d\mu = \int f_1\,d\mu + i \int f\,d\mu$$

Si vede anche (è un po' meno evidente) che la (1.4) continua a valere, intendendo con $|\ |$ il modulo complesso.

Indicheremo con $\mathscr{L}^1$ e $\mathscr{L}^1_{\mathbb{C}}$ le funzioni integrabili a valori reali e complessi rispettivamente. Scriveremo $\mathscr{L}^1(\mu)$ o $\mathscr{L}^1_{\mathbb{C}}(\mu)$ quando sarà necessario precisare la misura rispetto alla quale si integra.

Proprietà (in esercizio).

i) Se f è $\mathscr{E}$-misurabile positiva e se $\int f\,d\mu < +\infty$, allora $f < +\infty$ q.o.

ii) Se f è $\mathscr{E}$-misurabile positiva e se $\int f\,d\mu = 0$, $f = 0$ q.o.

iii) Se f è misurabile positiva (risp. integrabile) e $A \in \mathscr{E}$, allora $f 1_A$ è anch'essa positiva (risp. integrabile). Si può allora definire

$$\int_A f\,d\mu := \int f 1_A\,d\mu$$

Mostrare che se f è positiva (risp. integrabile) e se, per ogni $A \in \mathscr{B}$, $\int_A f\,d\mu \ge 0$, allora $f \ge 0$ q.o.

Restano da enunciare risultati di passaggio al limite. Il primo, da cui seguono facilmente gli altri, si chiama teorema di convergenza monotona o *teorema di Beppo-Levi* .

Teorema 1.11 *Sia* $(f_n)_n \subset \mathscr{E}^+$ *una successione crescente, allora*

$$\lim_{n\to\infty} \uparrow \int f_n \, d\mu = \int \lim_{n\to\infty} \uparrow f_n \, d\mu.$$

Corollario 1.12 *Sia* $(g_n)_n$ *una successione di funzioni* $\mathscr{E}$*-misurabili positive, allora*

$$\sum_n \int g_n \, d\mu = \int \sum_n g_n \, d\mu.$$

Proposizione 1.13 *(Lemma di Fatou) (i) Sia* $(f_n)_n$ *una successione di funzioni* $\mathscr{E}$*-misurabili positive, allora*

$$\int \underline{\lim}_{n\to\infty} f_n \, d\mu \le \underline{\lim}_{n\to\infty} \int f_n \, d\mu.$$

Il Lemma di Fatou implica il celebre *teorema di Lebesgue* .

Teorema 1.14 *Siano* f_n *funzioni integrabili tali che* $f_n \to f$ *q.o., con* $|f_n| \le g \in \mathscr{L}^1$*, allora*

$$\lim_{n\to\infty} \int f_n \, d\mu = \int f \, d\mu.$$

Questo teorema ha una versione ‘‘continua’’ molto utile.

Corollario 1.15 *Siano* $(f_t, t \in U)$ *una famiglia di elementi di* $\mathscr{L}^1_{\mathbb{C}}$ *e* U *un aperto di* $\mathbb{R}^d$*. Si suppone che* $\lim_{t\to t_0} f_t = f$ *q.o. e che, per ogni* $t \in U$*,* $|f_t| \le g \in \mathscr{L}^1$*, allora* $\lim_{t\to t_0} \int f_t \, d\mu = \int f \, d\mu$*.*

Dimostrazione. Basta osservare che $\lim_{t\to t_0} \int f_t \, d\mu = \int f \, d\mu$ se e solo se, per ogni successione $(t_n)_n$ convergente a t_0, $\lim_{t_n\to t_0} \int f_{t_n} \, d\mu = \int f \, d\mu$ e poi applicare il Teorema 1.14.

Diamo un esempio di applicazione di questo corollario.

Proposizione 1.16 *(Teorema di derivazione sotto il segno) Siano* $(E, \mathscr{E}, \mu)$ *uno spazio di misura,* I *un intervallo aperto e* $(f(t,x),\ t \in I)$ *una famiglia di funzioni integrabili a valori in* $\mathbb{C}$*. Poniamo, per ogni* $t \in I$*,* $\phi(t) = \int f(t,x) \, d\mu(x)$*. Si suppone che esista un insieme* $A \in \mathscr{E}$ *tale che* $\mu(A^c) = 0$ *e che, per ogni* $x \in A$*, si abbia che*

- $t \to f(t,x)$ *sia derivabile su* I*;*
- *esista una funzione* $g \in \mathscr{L}^1(\mu)$ *tale che per* $t \in I$*,* $|\frac{\partial f}{\partial t}(t,x)| \le g(x)$*.*

Allora ϕ *è derivabile su* I *e* $\phi'(t) = \int \frac{\partial f}{\partial t}(t,x) \, d\mu(x)$*.*

Dimostrazione. Si ha

$$\frac{1}{h}(\phi(t+h)-\phi(t)) = \int_A \frac{1}{h}(f(t+h,x)-f(t,x))\,d\mu(x).$$

e, naturalmente, per ogni $x \in A$,

$$\frac{1}{h}(f(t+h,x)-f(t,x)) \underset{h\to 0}{\to} \frac{\partial f}{\partial t}(t,x).$$

Grazie alla formula degli incrementi finiti, si ha, per h è abbastanza piccolo e per $x \in A$,

$$\left|\frac{1}{h}(f(t+h,x)-f(t,x))\right| = \left|\frac{\partial f}{\partial t}(\theta,x)\right| \le g(x)$$

dove $t \le \theta \le t+h$. Si può dunque applicare il Teorema di Lebesgue nella versione del Corollario 1.15 e si ottiene

$$\int_A \frac{1}{h}(f(t+h,x)-f(t,x))\,d\mu(x) \underset{h\to 0}{\to} \int_A \frac{\partial f}{\partial t}(t,x)\,d\mu(x) = \int \frac{\partial f}{\partial t}(t,x)\,d\mu(x).$$

Se f è una funzione boreliana, si pone, per $1 \le p < +\infty$,

$$\|f\|_p = \left[\int |f|^p\,d\mu\right]^{\frac{1}{p}}$$

e, per $p = +\infty$,

$$\|f\|_\infty = \inf(M, \mu(|f| > M) = 0).$$

Queste due quantità possono naturalmente essere $= +\infty$. Poniamo, per $1 \le p \le +\infty$,

$$\mathscr{L}^p = \{f, \|f\|_p < +\infty\}.$$

Si hanno due disuguaglianze fondamentali. Per f, g boreliane si ha

$$\|f+g\|_p \le \|f\|_p + \|g\|_p,\ 1 \le p \le +\infty \tag{1.5}$$

che si chiama la disuguaglianza di Minkowski e

$$\|fg\|_1 \le \|f\|_p\,\|g\|_q,\quad 1 \le p \le +\infty,\ \frac{1}{p}+\frac{1}{q} = 1 \tag{1.6}$$

che si chiama la disuguaglianza di Hölder. Osserviamo che per $p = q = 2$, (1.6) implica la disuguaglianza di Schwartz

$$\left[\int |fg|\,d\mu\right]^2 \le \int f^2\,d\mu \int g^2\,d\mu.$$

Grazie alla disuguaglianza di Minkowski, gli insiemi $\mathcal{L}^p$ sono degli spazi vettoriali e $\| \ \|_p$ è una seminorma. Non è una norma perché può succedere che una funzione $f \neq 0$ sia tale che $\|f\|_p = 0$ (succede se e solo se $f = 0$ q.o.). Se però definiamo una relazione di equivalenza su $\mathcal{L}^p$ ponendo $f \sim g$ se $f = g$ q.o. e poi poniamo $L^p = \mathcal{L}^p / \sim$, allora L^p risulta essere uno spazio normato. Infatti, poiché $f = g$ q.o. implica $\int |f|^p \, d\mu = \int |g|^p \, d\mu$, se f e g sono in $\mathcal{L}^p$, si può definire senza ambiguità, per $f \in L^p$, $\|f\|_p$. È utile ricordare che L^p è uno spazio che non è formato da funzioni, ma da classi di equivalenza di funzioni; questa distinzione è però raramente importante e *con abuso di linguaggio* nel seguito confonderemo una funzione f e la sua classe di equivalenza.

Si dimostra anzi che L^p è uno spazio di Banach e che L^2 è uno spazio di Hilbert per il prodotto scalare

$$\langle f, g\rangle = \int fg \, d\mu.$$

Si può anche considerare il caso delle funzioni a valori complessi. Si definisce allo stesso modo $L^p_{\mathbb{C}} = L^p_{\mathbb{C}}(E, \mathscr{E}, \mu)$. Occorre osservare che $L^2_{\mathbb{C}}$ è associato al prodotto scalare

$$\langle f, g\rangle = \int f\bar{g} \, d\mu.$$

Proposizione 1.17 *Per* $1 \le p < +\infty$*,* $\mathscr{E}^0 = \{f; f = \sum_{k=1}^n a_k 1_{A_k}, A_k \in \mathscr{E}, \mu(A_k) < +\infty\}$ *è denso in* $L^p(E, \mathscr{E}, \mu)$*.*

Dimostrazione. Basta considerare il caso $f \ge 0$. Allora esiste (Proposizione 1.2) una successione $(f_n)_n \subset \mathscr{E}^0$ tale che $f_n \uparrow f$. Poiché $f < +\infty$ q.o., $|f - f_n|^p \to 0$ q.o. e, dato che $f_n^p \le f^p \in \mathcal{L}^1$, si ha $|f - f_n|^p \le f^p \in \mathcal{L}^1$ Si può dunque applicare il teorema di Lebesgue e si ha $\int |f - f_n|^p \, d\mu \to 0$.

Sia μ una misura su $(E, \mathscr{E})$. Ad essa si può associare un'applicazione $I : \mathscr{E}^+ \to \overline{\mathbb{R}^+}$ ponendo $I(f) = \int f \, d\mu$, $f \in \mathscr{E}^+$. L'applicazione I a le proprietà seguenti:

i) se $f, g \in \mathscr{E}^+$ e $a, b \in \overline{\mathbb{R}^+}$, $I(af + bg) = I(af) + I(bg)$;

ii) se $f_n \in \mathscr{E}^+$ e se $f_n \uparrow f$, $I(f_n) \uparrow I(f)$.

Proposizione 1.18 *Siano* $(E, \mathscr{E})$ *un spazio misurabile e* I *un'applicazione di* $\mathscr{E}^+$ *in* $\overline{\mathbb{R}^+}$ *tale che valgano le i) e ii) precedenti.*

Allora $\mu(A) = I(1_A)$*,* $A \in \mathscr{E}$*, definisce una misura su* $\mathscr{E}$ *e si ha, per ogni* $f \in \mathscr{E}^+$*,* $I(f) = \int f \, d\mu$*.*

Dimostrazione. Siano $A_n \in \mathscr{E}$ degli insiemi a due a due disgiunti la cui unione sia uguale ad A; allora $1_A = \sum_n 1_{A_n} = \lim_{n\to\infty} \uparrow \sum_{k=1}^n 1_{A_k}$ e

$$\mu(A) = I(1_A) = I\Big(\lim_{n\to\infty} \uparrow \sum_{k=1}^n 1_{A_k}\Big) = \lim_{n\to\infty} \uparrow I\Big(\sum_{k=1}^n 1_{A_k}\Big) = \lim_{n\to\infty} \uparrow \sum_{k=1}^n I(1_{A_k}) =$$

$$= \sum_{k=1}^{\infty} \mu(A_k).$$

Ciò mostra che μ è una misura. Si ha allora, per ogni funzione elementare positiva f, $I(f) = \int f\,d\mu$. Si conclude facilmente usando la Proposizione 1.2.

Lemma 1.19 *Siano μ_1 e μ_2 misure di Borel su $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. Supponiamo che, per ogni $f \in C_K$ (funzioni continue a supporto compatto), $\int f\,d\mu_1 = \int f\,d\mu_2$. Allora $\mu_1 = \mu_2$.*

Dimostrazione. Indichiamo con $\mathcal{C}$ la classe degli aperti limitati. Per ogni $U \in \mathcal{C}$, esiste una successione $(f_n)_n \subset C_K$ tale che $1_U = \lim_{n\to\infty} \uparrow f_n$. Dunque, per il Teorema di Beppo Levi $\mu_1(U) = \mu_2(U) < +\infty$. Poiché $\mathcal{C}$ è stabile per intersezioni finite, genera $\mathcal{B}(\mathbb{R}^d)$ e $\mathbb{R}^d = \lim \uparrow U_n$, $U_n \in \mathcal{C}$, si conclude grazie alla Proposizione 1.6.

1.4 Esempi

Vediamo ora degli esempi di misure e alcune tecniche con le quali si possono costruire nuove misure a partire da misure date.

- (Masse di Dirac). Se $x \in E$ è fissato, consideriamo la misura su $\mathcal{P}(E)$ definita da

$$\mu(A) = 1_A(x)$$

cioè la misura di A vale 1 o 0 a seconda che $x \in A$ oppure no. La verifica che si tratta di una misura è immediata; questa misura si indica δ_x e si chiama *massa di Dirac* concentrata in x. Vale la formula

$$\int f\,d\delta_x = f(x)$$

che si dimostra facilmente allo stesso modo della prossima Proposizione 1.21

- (Insiemi numerabili) Se E è un insieme numerabile, si può costruire una misura su $(E, \mathcal{P}(E))$ in modo molto semplice. Se ad ogni $x \in E$ associamo un numero $\mu_x \in \mathbb{R}^+$, poniamo, per $A \subset E$, $\mu(A) = \sum_{x \in A} \mu_x$. Le proprietà di sommabilità delle serie a termini positivi implicano che μ così definita è una misura.
- (Misura immagine) Siano $(E, \mathcal{E})$ e $(F, \mathcal{F})$ spazi misurabili, $f\colon E \to F$ un'applicazione misurabile e μ una misura su $(E, \mathcal{E})$; si può definire una misura ν su $(F, \mathcal{F})$ ponendo

$$\nu(A) := \mu(f^{-1}(A)) \qquad A \in \mathcal{F} \tag{1.7}$$

Anche qui la verifica che ν è una misura è immediata. ν si chiama la *misura immagine* di μ tramite f e si indica $f(\mu)$ oppure $\mu \circ f^{-1}$. Nel resto di questo paragrafo supporremo che μ sia finita (cioè che $\mu(E) < +\infty$).

Proposizione 1.21 *Una funzione misurabile* $g: F \to \mathbb{R}$ *è* ν*-integrabile se e solo se* $g \circ f$ *è* μ*-integrabile. In questo caso si ha*

$$\int g\, d\nu = \int g \circ f\, d\mu \tag{1.8}$$

Dimostrazione. Basta mostrare (1.8) quando g è positiva. Se $g = 1_A$ la relazione (1.8) coincide con la definizione (1.7) perché $1_A \circ f = 1_{f^{-1}(A)}$. (1.8) è quindi vera se g è combinazione lineare d'indicatrici di insiemi di $\mathscr{F}$. Basta ora osservare che ogni funzione misurabile positiva è inviluppo superiore di funzioni di questo tipo (Proposizione 1.2) e applicare il teorema di Beppo Levi.

• (Misure definite da una densità) Sia μ una misura σ-finita su $(E, \mathscr{E})$. Diremo che una funzione misurabile positiva $f: E \to \mathbb{R}$ è una *densità* se esiste una successione $(A_n)_n \subset \mathscr{E}$ tale che $\bigcup_n A_n = E$, $\mu(A_n) < +\infty$ e che $f 1_{A_n}$ sia integrabile per ogni n. D'ora in avanti useremo la notazione

$$\int_A f\, d\mu \stackrel{\text{def}}{=} \int f 1_A\, d\mu$$

Proposizione e Definizione 1.22 *Si chiama misura di densità* f *rispetto a* μ *(e si scrive* $\nu = f\, d\mu$*) la misura* ν σ*-finita su* $(E, \mathscr{E})$ *definita da*

$$\nu(A) = \int_A f\, d\mu \tag{1.9}$$

Una funzione misurabile $g: E \to \mathbb{R}$ *è integrabile rispetto a* ν *se e solo se* gf *è integrabile rispetto a* μ *e in questo caso si ha*

$$\int g\, d\nu = \int g\, f\, d\mu \tag{1.10}$$

Dimostrazione. Che ν definita da (1.9) sia una misura segue facilmente dal fatto che essa passa al limite sulle successioni crescenti per il teorema di Beppo Levi. ν è σ-finita perché se $(A_n)_n$ è una successione di insiemi di $\mathscr{E}$ tale che $\bigcup_n A_n = E$ e $f 1_{A_n}$ sia integrabile per ogni n, allora

$$\nu(A_n) = \int f\, 1_{A_n}\, d\mu < +\infty$$

La (1.10) infine si dimostra allo stesso modo della Proposizione 1.21, verificandola cioè prima per le funzioni g della forma 1_A, quindi per linearità per le funzioni semplici e poi per tutte le funzioni positive, approssimandole con funzioni semplici (Proposizione 1.2) e usando il teorema di Beppo Levi.

Consideriamo due misure μ e ν σ-finite sullo spazio misurabile $(E, \mathscr{E})$. Diremo che ν è *assolutamente continua* rispetto a μ, e scriveremo $\nu \ll \mu$, se e solo se ogni insieme $A \in \mathscr{E}$ μ-trascurabile (tale cioè che $\mu(A) = 0$) è anche ν-trascurabile. Se ν ha densità f rispetto a μ allora è chiaro che $\nu \ll \mu$: infatti se A è μ-trascurabile allora la funzione $f 1_A$ è anch'essa trascurabile, poiché è diversa da 0 solo su A. Un risultato notevole e non ovvio è che vale anche il viceversa.

Teorema 1.23 (Radon-Nikodym) *Se μ e ν sono σ-finite e $\nu \ll \mu$ allora ν ha densità rispetto a μ.*

Osserviamo che, a voler essere precisi, non è corretto parlare di ''*la* densità di ν rispetto a μ'': se f è una densità, tale è anche ogni funzione g μ-equivalente a f. Si può dimostrare anzi che se due funzioni f e g sono entrambe densità di una stessa misura ν rispetto a μ se e solo se f e g sono μ-equivalenti.

1.5 Misure prodotto

Siano $(E_1, \mathscr{E}_1)$ $(E_2, \mathscr{E}_2)$ due spazi misurabili. Si definisce una σ-algebra su $E_1 \times E_2$, chiamata σ-algebra prodotto di $\mathscr{E}_1$ e $\mathscr{E}_2$ e indicata $\mathscr{E}_1 \otimes \mathscr{E}_2$, mediante

$$\mathscr{E}_1 \otimes \mathscr{E}_2 = \sigma(A_1 \times A_2; A_1 \in \mathscr{B}_1,\ A_2 \in \mathscr{E}_2).$$

Siano μ_1 e μ_2 misure σ-finite su $(E_1, \mathscr{E}_1)$ e $(E_2, \mathscr{E}_2)$ rispettivamente. Per ogni insieme della forma $A = A_1 \times A_2$ poniamo

$$\mu(A) = \mu_1(A_1)\mu_2(A_2) \tag{1.12}$$

Si può prolungare μ ad una misura su tutta la σ-algebra $\mathscr{E} = \mathscr{E}_1 \otimes \mathscr{E}_2$?

In ogni caso, se questo prolungamento esiste, è necessariamente unico, grazie alla Proposizione 1.6, perché la classe dei rettangoli $A_1 \times A_2$ è stabile per intersezioni finite. Per mostrare l'esistenza in questo caso il teorema di Carathéodory non è molto pratico da utilizzare. Si può procedere in maniera più semplice nel modo seguente. Sia $f : E_1 \times E_2 \to \overline{\mathbb{R}^+}$ è una funzione $\mathscr{E}_1 \otimes \mathscr{E}_2$-misurabile.

1) Si dimostra che, per ogni $x_1 \in E_1$, $x_2 \in E_2$ le funzioni $f(x_1, \cdot)$ e $f(\cdot, x_2)$ sono rispettivamente $\mathscr{E}_2$- e $\mathscr{E}_1$-misurabili.

2) Si dimostra che, per ogni $x_1 \in E_1$, $x_2 \in E_2$ le funzioni

$$x_1 \to \int f(x_1, x_2)\, d\mu_2(x_2), \qquad x_2 \to \int f(x_1, x_2)\, d\mu_1(x_1)$$

sono rispettivamente $\mathscr{E}_1$- e $\mathscr{E}_2$-misurabili.

3) Se si pone

$$I(f) = \int d\mu_2(x_2) \int f(x_1, x_2)\, d\mu_1(x_1)$$

allora il funzionale I soddisfa alle ipotesi i) e ii) che precedono la Proposizione 1.18 (si usa due volte il Teorema di Beppo Levi).

Ne segue che, posto $\mu(A) = I(1_A)$, è una misura su $\mathscr{E}_1 \otimes \mathscr{E}_2$. Poiché è chiaro che μ soddisfa alla (1.12) sui rettangoli, si tratta del prolungamento che cercavamo. La misura μ si chiama la *misura prodotto* di μ_1 e μ_2 e si scrive $\mu = \mu_1 \otimes \mu_2$.

La dimostrazione dei punti 1) e 2) precedenti è senza sorprese: si tratta di proprietà che sono vere se f è l'indicatrice di un rettangolo. Si passa al caso in cui f è la funzione indicatrice di un insieme di $\mathscr{E}_1 \otimes \mathscr{E}_2$ con il Teorema delle classi monotone 1.1 e poi a tutte le funzioni $\mathscr{E}_1 \otimes \mathscr{E}_2$-misurabili positive con la Proposizione 1.2 ed il Teorema di Beppo Levi.

In pratica per integrare rispetto alla misura prodotto si usa il teorema seguente, che è molto importante.

Teorema 1.25 *(Fubini) Sia f una funzione reale, numerica o complessa $\mathscr{E}_1 \otimes \mathscr{E}_2$-misurabile. Allora si ha*

$$\int d\mu_2(x_2) \int |f(x_1, x_2)|\, d\mu_1(x_1) = \int d\mu_1(x_1) \int |f(x_1, x_2)|\, d\mu_2(x_2).$$

Inoltre f è integrabile rispetto alla misura prodotto $\mu = \mu_1 \otimes \mu_2$ se e solo se la quantità precedente è finita. In questo caso vale la formula

$$\int f\, d\mu = \int d\mu_2(x_2) \int f(x_1, x_2)\, d\mu_1(x_1) = \int d\mu_1(x_1) \int f(x_1, x_2)\, d\mu_2(x_2).$$

Ciò si può estendere senza troppa fatica al caso di n spazi misurabili. Ci sono alcune verifiche un po' noiose da fare del tipo $\mu_1 \otimes (\mu_2 \otimes \mu_3) = (\mu_1 \otimes \mu_2) \otimes \mu_3$. Per di più nelle formule d'integrazione le variabili si possono integrare in tutti gli ordini possibili. Grosso modo, il grande principio per applicare il teorema di Fubini quando si vuole integrare una funzione rispetto alla misura prodotto è: se f è positiva, tutto è permesso (si può cioè integrare rispetto alle variabili in qualunque ordine); se f è di segno qualunque o complessa, tutto è permesso solo se la funzione è integrabile; occorre cioè considerare prima $|f|$ e mostrare che $|f|$ è integrabile.

Consideriamo $(\mathbb{R}, \mathscr{B}(\mathbb{R}), \lambda)$, λ misura di Lebesgue. Intanto è abbastanza facile verificare che $\mathscr{B}(\mathbb{R}) \otimes \mathscr{B}(\mathbb{R}) \otimes \ldots \otimes \mathscr{B}(\mathbb{R}) = \mathscr{B}(\mathbb{R}^d)$ (in esercizio). Si definisce allora $\lambda_d = \lambda \otimes \lambda \otimes \ldots \otimes \lambda$. Si può applicare la Proposizione 1.6 a

$$\mathscr{C} = \bigl\{A,\ A = \textstyle\prod_{i=1}^d\,]a_i, b_i[,\ -\infty < a_i < b_i < +\infty\bigr\}.$$

Si ottiene che λ_d è l'unica misura su $\mathscr{B}(\mathbb{R}^d)$ tale che, per ogni $-\infty < a_i < b_i < +\infty$,

$$\lambda_d\bigl(\textstyle\prod_{i=1}^d\,]a_i, b_i[\bigr) = \displaystyle\prod_{i=1}^d (b_i - a_i).$$

Indichiamo con λ_d la misura di Lebesgue di $\mathbb{R}^d$.

Esercizi

E1.1 Sia $g \in L^p(\mu)$, $1 \le p < +\infty$ e poniamo $g_n = g \wedge n \vee (-n)$. Mostrare che $g_n \to g$ in L^p per $n \to +\infty$.

2
Variabili aleatorie

2.1 Probabilità e misura

Uno spazio di probabilità è una tripla $(\Omega, \mathcal{A}, \mathrm{P})$ dove $(\Omega, \mathcal{A})$ è uno spazio misurabile e P una probabilità su $(\Omega, \mathcal{A})$. Vedremo che altri oggetti della teoria della misura intervengono nel Calcolo delle Probabilità e che, in questo caso, vengono ribattezzati ricevendo un nuovo nome che tiene conto del ruolo che essi giocano in relazione ai fenomeni aleatori. Ad esempio si chiamano *eventi* gli insiemi della σ-algebra $\mathcal{A}$.

Si chiama (v.a.) un'applicazione misurabile di $(\Omega, \mathcal{A}, \mathrm{P})$ a valori in $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ oppure $(\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}}))$. Naturalmente si possono considerare anche v.a. a valori in uno spazio misurabile qualunque $(E, \mathcal{E})$, ma quando lo spazio di arrivo non sarà precisato sottintenderemo sempre che si tratta di v.a. a valori reali o numerici. Parleremo di v.a. m-dimensionali per le v.a. a valori in $(\mathbb{R}^m, \mathcal{B}(\mathbb{R}^m))$. Le v.a. si denotano tradizionalmente con lettere maiuscole $(X, Y, Z, \ldots)$.

Si chiama *legge* o *distribuzione* della v.a. X la misura immagine di P tramite X, cioè la probabilità μ su $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ definita da

$$\mu(A) = \mathrm{P}(X^{-1}(A)) = \mathrm{P}(\omega; X(\omega) \in A)$$

dove naturalmente A è un boreliano di $\mathbb{R}$. Spesso useremo la scrittura $\mathrm{P}(X \in A)$ al posto di $\mathrm{P}(\omega; X(\omega) \in A)$. Si chiama *funzione di ripartizione* di X la funzione di ripartizione di μ, cioè la funzione F, non decrescente e continua a destra, definita da

$$F(x) = \mu(]-\infty, x]) = \mathrm{P}(X \le x) \ .$$

Se X è semi-integrabile (superiormente o inferiormente) rispetto a P, si chiama *speranza matematica* (o *attesa* o *media*) di X, e si indica con $\mathrm{E}(X)$, l'integrale $\int X\, d\mathrm{P}$. Se $X = (X_1, \ldots, X_m)$ è una v.a. m-dimensionale porremo

$$\mathrm{E}(X) = (\mathrm{E}(X_1), \ldots, \mathrm{E}(X_m)) \ .$$

Una v.a. X si dice *centrata* se $\mathrm{E}(X) = 0$. La Proposizione 1.21, d'integrazione rispetto a una legge immagine, permette di affermare che se $f : \mathbb{R} \to \mathbb{R}$ è una funzione boreliana, allora la v.a. $f(X)$ è integrabile se e solo se

$$\int |f(x)|\, d\mu(x) < +\infty$$

ed in questo caso

$$\mathrm{E}[f(X)] = \int f(x)\, d\mu(x)\ . \tag{2.1}$$

Naturalmente la relazione precedente vale anche se la v.a. $f(X)$ è solo semi-integrabile (il che accade sempre se f è positiva, ad esempio). In particolare se X è integrabile oppure positiva

$$\mathrm{E}(X) = \int_{\mathbb{R}} x\, \mu(dx)\ . \tag{2.2}$$

Questa relazione è quella che si usa in pratica per il calcolo della speranza matematica di una v.a.; inoltre, da un punto di vista concettuale, essa mostra che la speranza matematica dipende solo dalla legge di X: v.a. diverse (eventualmente definite su spazi di probabilità diversi) ma aventi la stessa legge hanno la stessa speranza matematica.

Anzi, se la (2.1) vale per ogni funzione f misurabile limitata (risp. positiva), allora necessariamente μ è la legge di X. Nel caso di v.a. reali è sufficiente che (2.1) sia vera per ogni funzione $f \in C_K$. Questa osservazione fornisce un metodo per determinare la legge di una v.a. X, come si vedrà negli esercizi.

Data una legge di probabilità μ su $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$, o comunque su uno spazio misurabile $(E, \mathscr{E})$, è sempre possibile costruire uno spazio di probabilità $(\Omega, \mathscr{A}, \mathrm{P})$ su cui è definita una v.a. X avente legge μ. Basterà ad esempio porre $\Omega = E$, $\mathscr{A} = \mathscr{E}$, $\mathrm{P} = \mu$ e $X(x) = x$.

Nel seguito commetteremo spesso un piccolo abuso: considereremo delle v.a. senza preoccuparci troppo di precisare lo spazio di probabilità su cui sono definite. La giustificazione di questo modo di procedere è che per poter fare i calcoli spesso la sola cosa necessaria è conoscere la legge di una v.a. e comunque la costruzione esplicita di uno spazio di probabilità su cui le variabili aleatorie sono definite è sempre possibile e spesso ovvia.

Sempre più spesso come modello di un fenomeno aleatorio considereremo uno spazio di probabilità $(\Omega, \mathscr{A}, \mathrm{P})$, di cui ignoreremo la natura, sul quale sono definite delle variabili aleatorie $X_1, \dots, X_n$ con date leggi di probabilità.

2.2 Indipendenza

In questo paragrafo $(\Omega, \mathscr{A}, \mathrm{P})$ è uno spazio di probabilità e tutte le σ-algebre che si considerano sono delle sotto-σ-algebre di $\mathscr{A}$.

Definizione 2.1 *Le σ -algebre $\mathcal{B}_i$, $i = 1, \dots, n$ si dicono indipendenti se*

$$\mathrm{P}\Big(\bigcap_{i=1}^{n} A_i\Big) = \prod_{i=1}^{n} \mathrm{P}(A_i)$$

per ogni $A_i \in \mathcal{B}_i$. Le σ -algebre di una famiglia qualunque $(\mathcal{B}_i, i \in I)$ si dicono indipendenti se ogni sotto-famiglia finita è composta da σ -algebre indipendenti.

Questa definizione ha una conseguenza evidente ma importante.

Lemma 2.2 *Se le σ -algebre $(\mathcal{B}_i, i \in I)$ sono indipendenti e se, per ogni $i \in I$, $\mathcal{B}'_i$ è una sotto-σ -algebra di $\mathcal{B}_i$, le σ -algebre $(\mathcal{B}'_i, i \in I)$ sono indipendenti.*

Proposizione 2.3 *Siano $\mathcal{C}_k \subset \mathcal{P}(\Omega)$, $k = 1, \dots, n$ delle classi stabili par intersezioni finite, contenenti Ω e tali che $\mathcal{B}_k = \sigma(\mathcal{C}_k)$, $k = 1, \dots, n$. Supponiamo che, per ogni $A_k \in \mathcal{C}_k$,*

$$\mathrm{P}\Big(\bigcap_{k=1}^{n} A_k\Big) = \prod_{k=1}^{n} \mathrm{P}(A_k)\ . \tag{2.3}$$

Allora le σ -algebre $\mathcal{B}_k$, $k = 1, \dots, n$, sono indipendenti.

Dimostrazione. Consideriamo, per $k = 1, \dots, n$, la proprietà

$$\mathrm{P}(\textstyle\bigcap_{i=1}^{n} A_i) = \displaystyle\prod_{i=1}^{n} \mathrm{P}(A_i), \text{ dove } A_i \in \sigma(\mathcal{C}_i), i = 1, \dots, k-1, \text{ e } A_i \in \mathcal{C}_i, i = k, \dots, n\ .$$

Par ipotesi essa è vera per $k = 1$; osserviamo che la tesi non è altro che affermare che questa proprietà è vera per $k = n + 1$. Supponiamola vera per $k = r$. Fissiamo degli eventi $A_i \in \mathcal{B}_i, i = 1, \dots, r - 1$ e $A_i \in \mathcal{C}_i$, $i = r + 1, \dots, n$ e consideriamo sulla σ-algebra $\mathcal{B}_r$ le due misure

$$\mu_1(B) = \mathrm{P}(A_1 \cap \dots \cap A_{r-1} \cap B \cap A_{r+1} \cap \dots \cap A_n)$$
$$\mu_2(B) = \mathrm{P}(A_1) \dots \mathrm{P}(A_{r-1})\mathrm{P}(B)\mathrm{P}(A_{r+1}) \dots \mathrm{P}(A_n)\ .$$

Queste due misure coincidono su $\mathcal{C}_i$, grazie all'ipotesi di ricorrenza. Per la Proposizione 1.6 esse coincidono su $\mathcal{B}_r$, e dunque la proprietà precedente è vera per $k = r + 1$. Per ricorrenza essa è vera anche per $k = n + 1$, cioè la tesi.

Vale anche una proprietà ''d'indipendenza per pacchetti"

Proposizione 2.4 *Siano $(\mathcal{B}_i, i \in I)$ delle σ -algebre indipendenti e $(I_j, j \in J)$ une partizione di I. Allora le σ -algebre $(\sigma(\mathcal{B}_i, i \in I_j), j \in J)$ sono indipendenti.*

Dimostrazione. Basta considerare il caso J finito. Sia

$$\mathcal{C}_j = \{B,\ B = A_1 \cap A_2 \cap \ldots \cap A_n,\ A_k \in \bigcup_{i \in I_j} \mathcal{B}_i\}\ .$$

Per l'indipendenza delle $\mathcal{B}_i$, si ha, per ogni scelta di $B_j \in \mathcal{C}_j$, $\mathrm{P}(\bigcap_j B_j) = \prod_j \mathrm{P}(B_j)$. Ma le $\mathcal{C}_j$ sono stabili per intersezioni finite, $\Omega \in \mathcal{C}_j$ e $\sigma(\mathcal{C}_j) = \sigma(\mathcal{B}_i,\ i \in I_j)$. Basta quindi applicare la Proposizione 2.3.

Definizione 2.5 *Le v.a. $(X_i)_{i \in \mathcal{I}}$ a valori negli spazi misurabili $(E_i, \mathcal{E}_i)$ si dicono indipendenti se le σ-algebre $(\sigma(X_i))_{i \in \mathcal{I}}$ sono indipendenti.*

Gli eventi $(A_i)_{i \in \mathcal{I}}$ si dicono indipendenti se le σ-algebre $(\sigma(A_i))_{i \in \mathcal{I}}$ sono indipendenti.

Si ha immediatamente,

Lemma 2.6 *Se le σ-algebre $(\mathcal{B}_i)_{i \in \mathcal{I}}$ sono indipendenti e se, per ogni $i \in \mathcal{I}$, X_i è una v.a. $\mathcal{B}_i$-misurabile, le v.a. $(X_i)_{i \in \mathcal{I}}$ sono indipendenti.*

Dimostrazione. Basta osservare che $\sigma(X_i) \subset \mathcal{B}_i$ e applicare il Lemma 2.2.

Teorema 2.7 *Siano X_i delle v.a. a valori $(E_i, \mathcal{E}_i)$ $i = 1, \ldots, n$. Indichiamo con μ la legge di $(X_1, \ldots, X_n)$, a valori nello spazio prodotto degli $(E_i, \mathcal{E}_i)$, e con $\mu_{X_i}, i = 1, \ldots, n$ le leggi delle $X_i, i = 1, \ldots, n$ rispettivamente. Le affermazioni seguenti sono equivalenti.*

i) Le v.a. $X_1, \ldots, X_n$ sono indipendenti.

ii) Per ogni $\Gamma_i \in \mathcal{E}_i$, $\mathrm{P}(X_1 \in \Gamma_1, \ldots, X_n \in \Gamma_n) = \mathrm{P}(X_1 \in \Gamma_1) \ldots \mathrm{P}(X_n \in \Gamma_n)$.

iii) Per ogni $\Gamma_i \in \mathcal{D}_i$, $\mathrm{P}(X_1 \in \Gamma_1, \ldots, X_n \in \Gamma_n) = \mathrm{P}(X_1 \in \Gamma_1) \ldots \mathrm{P}(X_n \in \Gamma_n)$ dove, per ogni i, $\mathcal{D}_i$ è une classe stabile per intersezioni finite, contenenti E_i e tale che $\sigma(\mathcal{D}_i) = \mathcal{E}_i$.

iv) Per ogni $f_i \in \mathcal{E}_i^+$ (risp. ogni $f_i \in b\mathcal{E}_i$),

$$\mathrm{E}(f_1(X_1) \ldots f_n(X_n)) = \mathrm{E}(f_1(X_1)) \ldots \mathrm{E}(f_n(X_n))\ .$$

v) $\mu = \mu_{X_1} \otimes \ldots \otimes \mu_{X_n}$.

Dimostrazione. (i)⇔(ii). È la definizione. (ii)⇒(v). Si ha, per ogni $\Gamma_i \in \mathcal{E}_i$,

$$\begin{aligned}\mu(\Gamma_1 \times \ldots \times \Gamma_n) &= \mathrm{P}(X_1 \in \Gamma_1, \ldots, X_n \in \Gamma_n) = \mathrm{P}(X_1 \in \Gamma_1) \ldots \mathrm{P}(X_n \in \Gamma_n) = \\ &= \mu_{X_1}(\Gamma_1) \ldots \mu_{X_n}(\Gamma_n)\end{aligned}$$

Dunque μ coincide con la misura prodotto $\mu_{X_1} \otimes \ldots \otimes \mu_{X_n}$ sui rettangoli. Quindi $\mu = \mu_{X_1} \otimes \ldots \otimes \mu_{X_n}$. (v)⇒(iv). È il Teorema 1.25, di Fubini. Infatti

$$\mathrm{E}(f_1(X_1) \ldots f_n(X_n)) = \int f(X_1) \ldots f(X_n)\, d\mathrm{P} =$$

$$= \int f(x_1) \ldots f(x_n)\, d\mu(x_1, \ldots, x_n) = \int f(x_1)\, d\mu_{X_1}(x_1) \ldots \int f(x_n)\, d\mu_{X_n}(x_n) =$$

$$= \mathrm{E}(f_1(X_1)) \ldots \mathrm{E}(f_n(X_n))$$

(iv)⇒(ii). Basta prendere $f_i = 1_{\Gamma_i}$:

$$\mathrm{E}(1_{\Gamma_1}(X_1)\dots 1_{\Gamma_n}(X_n)) = \mathrm{E}(1_{X_1\in\Gamma_1}\dots 1_{X_n\in\Gamma_n}) = \mathrm{E}(1_{\{X_1\in\Gamma_1\}\cap\dots\cap\{X_n\in\Gamma_n\}}) = \\ = \mathrm{P}(X_1\in\Gamma_1,\dots,X_n\in\Gamma_n)$$

e allo stesso modo $\mathrm{E}(1_{\Gamma_1}(X_1))\dots\mathrm{E}(1_{\Gamma_n}(X_n)) = \mathrm{P}(X_1\in\Gamma_1)\dots\mathrm{P}(X_n\in\Gamma_n)$.
(iii)⇔(ii). Si applica la Proposizione 2.3 ponendo $\mathcal{B}_i = \sigma(X_i)$ e $\mathcal{C}_i = \{X_i^{-1}(\Gamma),\ \Gamma\in\mathcal{D}_i\}$.

Corollario 2.8 *Siano $X_1,\dots,X_n$ delle v.a. reali, le affermazioni seguenti sono equivalenti*

i) Le v.a. $X_1,\dots,X_n$ sono indipendenti.
ii) Per ogni $a_i, b_i\in\mathbb{R}$,

$$\mathrm{P}(a_1 < X_1 < b_1,\dots,a_n < X_n < b_n) = \mathrm{P}(a_1 < X_1 < b_1)\dots\mathrm{P}(a_n < X_n < b_n)\ .$$

iii) Per ogni f_i continua a supporto compatto

$$\mathrm{E}(f_1(X_1)\dots f_n(X_n)) = \mathrm{E}(f_1(X_1))\dots\mathrm{E}(f_n(X_n))\ .$$

Dimostrazione. Basta osservare che (iii)⇒(ii) poiché $1_{]a,b[} = \lim\uparrow f_m$ con $f_m\in C_K$ e applicare il Teorema 2.7.

Corollario 2.9 *Siano $X_1,\dots,X_n$ delle v.a. reali integrabili indipendenti. Si ha*

$$\mathrm{E}(X_1\dots X_n) = \mathrm{E}(X_1)\dots\mathrm{E}(X_n)\ .$$

Dimostrazione. Si ha, grazie al Teorema 2.7 (iv), $\mathrm{E}(|X_1\dots X_n|) = \mathrm{E}(|X_1|)\dots\mathrm{E}(|X_n|) < +\infty$. Dunque $X_1\dots X_n$ è integrabile e si applica il Teorema 2.7 (v) ed il teorema di Fubini.

Osservazione 2.10 Se le v.a. $X_1,\dots,X_n$ sono a valori di spazi misurabili E_i numerabili e muniti della σ-algebra di tutte le parti, esse sono indipendenti (basta sommare) se e solo se, per ogni $x_i\in E_i$,

$$\mathrm{P}(X_1 = x_1,\dots,X_n = x_n) = \mathrm{P}(X_1 = x_1)\dots\mathrm{P}(X_n = x_n)\ .$$

Osservazione 2.11 Attenzione: può succedere che sia X indipendente da Y, X indipendente da Z senza che X sia indipendente da (Y, Z). Per esempio siano X e Y due v.a. indipendenti tali che $\mathrm{P}(X = 1) = \mathrm{P}(Y = 1) = \mathrm{P}(X = -1) = \mathrm{P}(Y = -1) = \frac{1}{2}$ e poniamo $Z = XY$. Si ha ancora $\mathrm{P}(Z = 1) = \mathrm{P}(Z = -1) = \frac{1}{2}$. Si verifica facilmente che X e Z sono indipendenti: infatti $\mathrm{P}(X = 1, Z = 1) = \mathrm{P}(X = 1, Y = 1) = \frac{1}{4} = \mathrm{P}(X = 1)\mathrm{P}(Z = 1),\dots$ X non è però indipendente da $Z/Y = X$ poiché ciò implicherebbe che X è q.c. costante. La classe $\mathcal{C} = \{A,\ A = \{Y\in\Gamma_1$ dove $A = \{Z\in\Gamma_2\}\}$ non è stabile per intersezione.

Teorema 2.12 *(La legge* 0-1 *di Kolmogorov) Sia* $(X_n)_n$ *una successione di v.a. indipendenti. Poniamo* $\mathscr{B}^n = \sigma(X_k, k \ge n)$ *e* $\mathscr{B}^\infty = \bigcap_{n=1}^\infty \mathscr{B}^n$. *(la* σ*-algebra terminale). Allora, per ogni* $A \in \mathscr{B}^\infty$, *si ha* $\mathrm{P}(A) = 0$ *oppure* $\mathrm{P}(A) = 1$. *Per di più, se* X *è una v.a.* $\mathscr{B}^\infty$*-misurabile,* X *è costante q.c.*

Dimostrazione. Poniamo $\mathscr{A}_n = \sigma(X_k,\ k \le n)$, $\mathscr{A}_\infty = \sigma(X_k,\ k \ge 0)$. Per la Proposizione 2.4, $\mathscr{A}_n$ è indipendente da $\mathscr{B}^{n+1}$ e da $\mathscr{B}^\infty \subset \mathscr{B}^{n+1}$. Dunque $\mathscr{B}^\infty$ è indipendente da $\mathscr{A}_\infty$ (è la Proposizione 2.3 applicata a $\cup\mathscr{A}_n$) ma $\mathscr{B}^\infty \subset \mathscr{A}_\infty$, per cui $\mathscr{B}^\infty$ è indipendente da se stessa. Se $A \in \mathscr{B}^\infty$, si ha dunque $\mathrm{P}(A) = \mathrm{P}(A \cap A) = \mathrm{P}(A)\mathrm{P}(A)$ i.e. $\mathrm{P}(A) = 0$ oppure $\mathrm{P}(A) = 1$. Se X è une v.a. $\mathscr{B}^\infty$-misurabile, $\mathrm{P}(X \le a) = 0$ oppure $\mathrm{P}(X \le a) = 1$. Dunque se $c = \sup(a,\ \mathrm{P}(X \le a) = 0)$,

$$\mathrm{P}(X = c) = \lim \downarrow \mathrm{P}(X \le c + \epsilon) - \lim \uparrow \mathrm{P}(X \le c - \epsilon) = 1\ .$$

Vediamo delle applicazioni della legge 0-1, che verranno sviluppate più tardi. Sia $(X_n)_n$ una successione di v.a. indipendenti e poniamo $\bar{X}_n = \frac{1}{n}(X_1 + \ldots + X_n)$. Allora la v.a. $\overline{\lim}_{n\to\infty} \bar{X}_n$ è terminale. Infatti il valore del $\overline{\lim}$ non dipende da $X_1, \ldots, X_n$, qualunque sia n. Ne segue intanto che la v.a. $\overline{\lim}_{n\to\infty} \bar{X}_n$ è q.c. costante. D'altra parte lo stesso argomento vale per la v.a. $\underline{\lim}_{n\to\infty} \bar{X}_n$. Si ha dunque che

$$\{\text{la successione } (\bar{X}_n)_n \text{ converge}\} = \{\overline{\lim_{n\to\infty}}\ \bar{X}_n = \underline{\lim_{n\to\infty}}\ \bar{X}_n\}$$

ha probabilità 1. Quindi o la successione $(\bar{X}_n)_n$ converge con probabilità 1 (e in questo caso il limite è una v.a. q.c. costante) oppure non converge con probabilità 1.

Un discorso simile si può fare quando si vuole studiare la convergenza della serie $\sum_{n=1}^\infty X_n$. Anche qui è facile vedere che l'evento {la serie converge} è terminale, poiché la convergenza di una serie non dipende dai suoi primi termini. Dunque o la serie non converge con probabilità 1 oppure è convergente con probabilità 1. In questo caso però il valore della serie dipende anche dai suoi primi termini. Dunque la v.a. $\sum_{n=1}^\infty X_n$ *non* è terminale e può non essere costante.

2.3 Disuguaglianze di convessità, momenti, covarianza

Vedremo ora alcune proprietà dell'integrale di una funzione misurabile rispetto ad una misura di probabilità (cioè della speranza matematica).

Si tratta di proprietà in generale non vere per integrali rispetto ad altre misure e legate al fatto che per le misure di probabilità l'integrale assume il significato di *media* o, per v.a. a valori in $\mathbb{R}^n$, di *baricentro*.

Ricordiamo che una funzione $\phi\colon \mathbb{R}^m \to \mathbb{R} \cup \{+\infty\}$ si dice *convessa* se e solo se per ogni $0 \le \lambda \le 1$ e $x, y \in \mathbb{R}^m$ si ha

$$\phi(\lambda x + (1-\lambda)y) \le \lambda\phi(x) + (1-\lambda)\phi(y)\ . \tag{2.4}$$

Si dice *concava* se $-\phi$ è convessa. Se ϕ è concava naturalmente si ha

$$\phi(\lambda x + (1-\lambda)y) \geq \lambda\phi(x) + (1-\lambda)\phi(y) \ . \tag{2.5}$$

Una funzione $f: \mathbb{R}^m \to \mathbb{R}$ si dice *affine* se è della forma

$$f(x) = \langle \alpha, x \rangle + b$$

dove $b \in \mathbb{R}$, $\alpha \in \mathbb{R}^m$ e $\langle\ ,\ \rangle$ indica il prodotto scalare in $\mathbb{R}^m$. È immediato verificare che se f è affine e X una v.a. m-dimensionale integrabile, allora $f(X)$ è anch'essa una v.a. (reale) integrabile e

$$\mathrm{E}(f(X)) = f(\mathrm{E}(X)) \ . \tag{2.6}$$

Inoltre una funzione affine f è continua e convessa. Anzi la (2.4) è verificata con = al posto di $\leq$ per cui f è anche concava.

Useremo nel seguito il fatto che se ϕ è convessa e semi-continua inferiormente (s.c.i.) allora

$$\phi(x) = \sup_f f(x)$$

dove l'estremo superiore è preso al variare di f tra tutte le funzioni affini tali che $f \leq \phi$. Un risultato analogo vale naturalmente per le funzioni concave e s.c.s. (con inf).

Ricordiamo che se μ è una misura, una funzione f si dice semi-integrabile inferiormente (s.c.i.) rispetto a μ se e solo se essa è minorata da una funzione integrabile rispetto a μ. In questo caso l'integrale $\int f\, d\mu$ è definito (eventualmente $= +\infty$). Analogamente f si dice semi- integrabile superiormente (s.c.s.) se è maggiorata da una funzione integrabile (e in questo caso l'integrale può prendere il valore $-\infty$).

Teorema 2.13 (Disuguaglianza di Jensen) *Sia X una v.a integrabile m-dimensionale e $\phi: \mathbb{R}^m \to \mathbb{R} \cup \{+\infty\}$ una funzione convessa s.c.i. (risp. concava e s.c.s.). Allora la v.a. reale $\phi(X)$ è semi-integrabile inferiormente (risp. semi-integrabile superiormente) e*

$$\mathrm{E}(\phi(X)) \geq \phi(\mathrm{E}(X)) \qquad\qquad (\text{risp. } \mathrm{E}(\phi(X)) \leq \phi(\mathrm{E}(X)))\ .$$

Dimostrazione. Se f è una funzione affine minorante ϕ, allora $f(X)$ è integrabile e $\phi(X) \geq f(X)$; quindi $\phi(X)$ è semi-integrabile inferiormente. Inoltre

$$\mathrm{E}(\phi(X)) \geq \mathrm{E}(f(X)) = f(\mathrm{E}(X)) \ .$$

Prendendo il sup su tutte le funzioni affini f minoranti ϕ, per la (2.6) si ha

$$\mathrm{E}(\phi(X)) \geq \phi(\mathrm{E}(X))$$

cioè la tesi.

Scegliendo delle particolari funzione ϕ dalla disuguaglianza di Jensen si possono ricavare facilmente le disuguaglianze classiche che abbiamo già visto nel capitolo 1.

- *Disuguaglianza di Hölder.* Se X e Y sono v.a. reali positive e α, β numeri > 0 tali che $\alpha + \beta = 1$, allora

$$\mathrm{E}(X^\alpha Y^\beta) \le \mathrm{E}(X)^\alpha \mathrm{E}(Y)^\beta \ . \tag{2.7}$$

Basta applicare la disuguaglianza di Jensen alla funzione concava s.c.s.

$$\phi(x, y) = \begin{cases} x^\alpha y^\beta & x, y \ge 0 \\ -\infty & \text{altrimenti} \end{cases} \ .$$

- *Disuguaglianza di Schwartz.* Se p, q sono reali positivi tali che $\frac{1}{p} + \frac{1}{q} = 1$ allora

$$\mathrm{E}(|XY|) \le \mathrm{E}(|X|^p)^{1/p} \mathrm{E}(|Y|^q)^{1/q} \ . \tag{2.8}$$

Basta ora applicare la disuguaglianza di Hölder alle v.a. $|X|^p, |Y|^q$, con $\alpha = \frac{1}{p}, \beta = \frac{1}{q}$.

- *Disuguaglianza di Minkowski.* Per ogni $p \ge 1$ si ha

$$\mathrm{E}(|X + Y|^p)^{1/p} \le \mathrm{E}(|X|^p)^{1/p} + \mathrm{E}(|Y|^p)^{1/p} \ . \tag{2.9}$$

Si applica la disuguaglianza di Jensen alla funzione concava s.c.s.

$$\phi(x, y) = \begin{cases} (x^{1/p} + y^{1/p})^p & x, y \ge 0 \\ -\infty & \text{altrimenti} \end{cases}$$

ed alle v.a. $|X|^p, |Y|^p$. Infatti con queste notazioni $\phi(|X|^p, |Y|^p) = (|X| + |Y|)^p$ e dunque

$$\begin{aligned} \mathrm{E}(|X + Y|^p) &\le \mathrm{E}((|X| + |Y|)^p) = \mathrm{E}(\phi(|X|^p, |Y|^p)) \le \\ &\le \phi(\mathrm{E}(|X|^p), \mathrm{E}(|Y|^p)) = \big(\mathrm{E}(|X|^p)^{1/p} + \mathrm{E}(|Y|^p)^{1/p}\big)^p \end{aligned}$$

e basta ora elevare alla potenza $\frac{1}{p}$ ambo i membri.

La disuguaglianza di Jensen vale solo per misure di probabilità, poiché essa implica la (2.6) che è propria delle misure di massa totale $= 1$. Le disuguaglianze di Hölder, Schwartz e Minkowski sono invece vere per ogni misura σ-finita. Nel caso di una misura di probabilità esse sono però un caso particolare della disuguaglianza di Jensen.

Un'altra applicazione notevole della disuguaglianza di Jensen è la seguente. Se $p > q$ la funzione $\phi(x) = |x|^{p/q}$ è convessa. Dunque

$$\|X\|_p^p = \mathrm{E}(|X|^p) = \mathrm{E}[\phi(|X|^q)] \ge \phi(\mathrm{E}[|X|^q]) = \mathrm{E}(|X|^q)^{p/q}$$

e, prendendo la radice p-esima,

$$\|X\|_p \ge \|X\|_q \ . \tag{2.10}$$

Veramente il calcolo precedente non è proprio corretto, perché la disuguaglianza di Jensen si applica a delle v.a. integrabili e l'integrabilità di $|X|^q$ è appunto invece quello che volevamo mostrare. Però non è difficile rimediare. Se poniamo $X_n = X \vee n \wedge (-n)$, allora $|X_n| \le n$ e, ripetendo il calcolo di poco fa si ha

$$\mathrm{E}(|X|^p) \ge \mathrm{E}(|X_n|^p) \ge \mathrm{E}(|X_n|^q)^{p/q} \ .$$

Basta ora fare tendere n all'infinito e usare il teorema di Beppo Levi, dato che $|X_n| \uparrow |X|$.

In particolare, se $p \ge q$, $L^p \subset L^q$. Questa proprietà d'inclusione non vale, in generale, per gli spazi L^p di misure che non siano di probabilità. Data una v.a. X e $\alpha > 0$, si chiama *momento assoluto* di ordine α la quantità $\mathrm{E}(|X|^\alpha) = \|X\|_\alpha^\alpha$. Si chiama *momento centrato assoluto* di ordine α la quantità $\mathrm{E}(|X - \mathrm{E}(X)|^\alpha)$ (eventualmente $= +\infty$).

Si chiama *varianza* di una v.a. X il momento centrato del second'ordine, cioè

$$\mathrm{Var}(X) = \mathrm{E}[(X - \mathrm{E}(X))^2] \ . \tag{2.11}$$

Osserviamo che X ha varianza finita se e solo se $X \in L^2$: se X ha varianza finita, poiché $X = (X - \mathrm{E}(X)) + \mathrm{E}(X)$, X è in L^2 come somma di v.a. di L^2. E se $X \in L^2$, anche $X - \mathrm{E}(X) \in L^2$ per lo stesso motivo.

Ricordando che $\mathrm{E}(X)$ è una v.a. costante si ha

$$\begin{aligned} \mathrm{E}(X - \mathrm{E}(X))^2) &= \mathrm{E}(X^2 - 2X\mathrm{E}(X) + \mathrm{E}(X)^2) = \\ &= \mathrm{E}(X^2) - 2\mathrm{E}(X\mathrm{E}(X)) + \mathrm{E}(X)^2 = \mathrm{E}(X^2) - \mathrm{E}(X)^2 \end{aligned}$$

da cui si ricava una espressione alternativa per la varianza

$$\mathrm{Var}(X) = \mathrm{E}(X^2) - \mathrm{E}(X)^2 \tag{2.12}$$

che è talvolta più comoda per il calcolo. Questa relazione mostra anche che si ha sempre $\mathrm{E}(X^2) \ge \mathrm{E}(X)^2$, cosa peraltro ovvia per la disuguaglianza di Jensen.

Come per la speranza matematica anche i momenti di una v.a. X sono quantità che dipendono solo dalla legge. Infatti per il Teorema 1.21, d'integrazione rispetto a una legge immagine,

$$\begin{aligned} \mathrm{E}(|X|^\alpha) &= \int |x|^\alpha \mu(dx) \\ \mathrm{E}(|X - \mathrm{E}(X)|^\alpha) &= \int |x - \mathrm{E}(X)|^\alpha \mu(dx) \\ \mathrm{Var}(X) &= \int |x - \mathrm{E}(X)|^2 \mu(dx) = \int x^2 \mu(dx) - \Big(\int x\, \mu(dx)\Big)^2 \ . \end{aligned}$$

Ai momenti sono legate due disuguaglianze importanti.

La *disuguaglianza di Markov*:

$$\mathrm{P}(|X| > t) \le \frac{\mathrm{E}(|X|^\alpha)}{t^\alpha} \tag{2.13}$$

che è immediata perché

$$\mathrm{E}(|X|^\alpha) \ge \mathrm{E}(|X|^\alpha 1_{\{|X|>t\}}) \ge t^\alpha \mathrm{P}(|X| > t)$$

dove usiamo il fatto che $|X|^\alpha 1_{\{|X|>t\}} \ge t^\alpha 1_{\{|X|>t\}}$.

Applicata alla v.a. $X - \mathrm{E}(X)$ e $\alpha = 2$ la (2.13) dà luogo alla *disuguaglianza di Chebyshev*

$$\mathrm{P}(|X - \mathrm{E}(X)| \ge t) \le \frac{\mathrm{Var}(X)}{t^2} \ .$$

Per la varianza valgono le seguenti proprietà, la cui verifica è immediata

$$\begin{aligned} \mathrm{Var}(X + a) &= \mathrm{Var}(X) \qquad && a \in \mathbb{R} \\ \mathrm{Var}(\lambda X) &= \lambda^2 \,\mathrm{Var}(X) \qquad && \lambda \in \mathbb{R} \ . \end{aligned}$$

Cerchiamo ora una formula per la varianza della somma di due v.a.:

$$\begin{aligned} \mathrm{Var}(X + Y) &= \mathrm{E}\big((X + Y - \mathrm{E}(X) - \mathrm{E}(Y))^2\big) = \\ &= \mathrm{E}\big((X - \mathrm{E}(X))^2\big) + \mathrm{E}\big((Y - \mathrm{E}(Y))^2\big) + 2\mathrm{E}\big((X - \mathrm{E}(X))(Y - \mathrm{E}(Y))\big) \end{aligned}$$

ovvero se poniamo

$$\mathrm{Cov}(X, Y) = \mathrm{E}\big((X - \mathrm{E}(X))(Y - \mathrm{E}(Y))\big)$$

allora

$$\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\,\mathrm{Cov}(X, Y) \ .$$

La quantità $\mathrm{Cov}(X, Y)$ si chiama la *covarianza* di X e Y. Se X e Y sono indipendenti allora per il Corollario 2.9

$$\mathrm{Cov}(X, Y) = \mathrm{E}\big((X - \mathrm{E}(X))(Y - \mathrm{E}(Y))\big) = \mathrm{E}\big((X - \mathrm{E}(X))\big)\mathrm{E}\big(Y - \mathrm{E}(Y))\big) = 0$$

cioè se X e Y sono indipendenti

$$\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) \ .$$

Non è invece vero il viceversa: si conoscono esempi di v.a. che hanno covarianza nulla, senza essere indipendenti. Si chiama coefficiente di correlazione di X e Y la quantità

$$\rho_{X,Y} = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Var}(X)\,\mathrm{Var}(Y)}} \ .$$

Notiamo che per la disuguaglianza di Schwartz

$$|\operatorname{Cov}(X, Y)| \le \operatorname{Var}(X)^{1/2} \operatorname{Var}(Y)^{1/2}$$

per cui $-1 \le \rho_{X,Y} \le 1$.

Anche se si conoscono esempi di v.a. non indipendenti per cui la covarianza (e dunque il coefficiente di correlazione) è nulla, la quantità $\rho_{X,Y}$ viene comunque usata (spesso impropriamente) per dare una valutazione di ‘‘quanto sono indipendenti’’ X e Y; nel senso che valori di $\rho_{X,Y}$ vicini a 0 indicano v.a. ‘‘quasi indipendenti’’ mentre valori vicini a 1 o -1 indicano una ‘‘forte dipendenza" (vedi l'Esercizio 4.2).

Se $\rho_{X,Y} = 0$ le v.a. X e Y si dicono *non correlate*. Osserviamo che dire che X e Y sono non correlate significa semplicemente che $X - \mathrm{E}(X)$ e $Y - \mathrm{E}(Y)$ sono ortogonali in L^2.

Esempio 2.14 (Retta di regressione) Consideriamo due v.a. reali X e Y definite sullo stesso spazio di probabilità $(\Omega, \mathscr{A}, \mathrm{P})$. Cerchiamo due numeri $a, b \in \mathbb{R}$ che rendano minima la quantità $\mathrm{E}((aX + b - Y)^2)$. In un certo senso si tratta di trovare la funzione lineare-affine di X che approssima meglio Y. Supporremo nel seguito che entrambe le variabili siano di quadrato integrabile. Converrà piuttosto scrivere

$$\mathrm{E}((aX + b - Y)^2) = \mathrm{E}((a(X - \mathrm{E}(X)) + \tilde{b} - (Y - \mathrm{E}(Y)))^2)$$

dove $\tilde{b} = b + a\mathrm{E}(X) - \mathrm{E}(Y)$. Si tratta di trovare il punto di minimo della funzione

$$\begin{aligned} S(a, \tilde{b}) &= \mathrm{E}((a(X - \mathrm{E}(X)) + \tilde{b} - (Y - \mathrm{E}(Y)))^2) = \\ &= a^2 \operatorname{Var}(X) + \tilde{b}^2 + \operatorname{Var}(Y) - a \operatorname{Cov}(X, Y)\ . \end{aligned}$$

È chiaro che il minimo si raggiunge per $\tilde{b} = 0$. Poiché S è un trinomio di secondo grado in a ed il coefficiente di a^2 è positivo, il punto critico è anche di minimo. Poiché

$$\frac{\partial S}{\partial a} = 2a \operatorname{Var}(X) - 2 \operatorname{Cov}(X, Y) = 0$$

deve essere

$$a = \frac{\operatorname{Cov}(X, Y)}{\operatorname{Var}(X)}$$

e dunque

$$b = \mathrm{E}(Y) - a\mathrm{E}(X) = \mathrm{E}(Y) - \frac{\operatorname{Cov}(X, Y)}{\operatorname{Var}(X)} \mathrm{E}(X) \cdot$$

La funzione $x \to ax + b$ per i valori calcolati si chiama la *retta di regressione* di Y su X. Osserviamo che il coefficiente angolare ha lo stesso segno della covarianza. La retta di regressione è dunque una funzione crescente o decrescente a seconda che $\operatorname{Cov}(X, Y)$

sia positiva o negativa, in accordo con il significato intuitivo della covarianza che è stato illustrato precedentemente.

Qual è il numero b per cui la quantità $b \to \mathrm{E}(|Y-b|^2)$ è minima? Come caso particolare del calcolo dell'Esempio 2.14 (scegliendo $X = 0$) otteniamo $b = \mathrm{E}(Y)$.

Se $X = (X_1, \dots, X_n)$ è una v.a. m-dimensionale, si chiama *matrice di covarianza* di X la matrice $m \times m$ C_X i cui elementi sono

$$c_{ij} = \mathrm{E}\big((X_i - \mathrm{E}(X_i))(X_j - \mathrm{E}(X_j))\big) \ .$$

Si tratta di una matrice simmetrica che ha sulla diagonale le varianze delle componenti di X e fuori della diagonale le loro covarianze. Quindi se $X_1, \dots, X_n$ sono indipendenti la loro matrice di correlazione è diagonale. Il viceversa naturalmente non è vero.

La matrice di covarianza è sempre semi-definita positiva, cioè per ogni vettore $\xi \in \mathbb{R}^n$

$$\langle C_X \xi, \xi \rangle = \sum c_{ij} \xi_i \xi_j \geq 0 \ .$$

Infatti

$$\sum c_{ij} \xi_i \xi_j = \sum \mathrm{E}\big(\xi_i (X_i - \mathrm{E}(X_i)) \xi_j (X_j - \mathrm{E}(X_j))\big) = \mathrm{E}\big(\langle \xi, X - \mathrm{E}(X) \rangle^2\big) \geq 0 \ .$$

2.4 Funzioni caratteristiche, trasformata di Laplace

Sia X una v.a. m-dimensionale. Si chiama *funzione caratteristica* di X la funzione $\phi : \mathbb{R}^m \to \mathbb{C}$ definita da

$$\phi(\theta) = \mathrm{E}(\mathrm{e}^{i\langle \theta, X \rangle}) = \mathrm{E}(\cos\langle \theta, X \rangle) + i\mathrm{E}(\sin\langle \theta, X \rangle) \ . \tag{2.14}$$

La funzione caratteristica è sempre definita perché, qualunque sia $\theta \in \mathbb{R}^m$, $|\mathrm{e}^{i\langle \theta, X \rangle}| = 1$, e anzi per (2.14)

$$|\phi(\theta)| \leq 1 \qquad \text{per ogni } \theta \in \mathbb{R}^m \ .$$

Per di più è ovvio che $\phi(0) = 1$. Il Teorema 1.21 d'integrazione rispetto ad una legge immagine dà inoltre

$$\phi(\theta) = \int_{\mathbb{R}^m} \mathrm{e}^{i\langle \theta, x \rangle} \mu(dx) \tag{2.15}$$

dove μ denota la legge di X. La funzione caratteristica quindi dipende in realtà solo dalla legge di X e potremo parlare indifferentemente di funzione caratteristica di una v.a. oppure di una legge di probabilità.

Le funzioni caratteristiche godono di molte proprietà che le rendono uno strumento di calcolo particolarmente utile. Quando ci sia pericolo di ambiguità scriveremo ϕ_X oppure ϕ_μ per indicare la funzione caratteristica della v.a. X oppure della sua legge μ. Talvolta scriveremo $\hat{\mu}(\theta)$ invece di $\phi_\mu(\theta)$.

Se μ e ν sono leggi di probabilità si ha

$$\phi_{\mu*\nu}(\theta) = \phi_\mu(\theta)\phi_\nu(\theta) . \tag{2.16}$$

Infatti se X e Y sono v.a. indipendenti di legge μ e ν rispettivamente

$$\begin{aligned}\phi_{\mu*\nu}(\theta) = \phi_{X+Y}(\theta) = \mathrm{E}(\mathrm{e}^{i\langle\theta,X+Y\rangle}) = \mathrm{E}(\mathrm{e}^{i\langle\theta,X\rangle}\mathrm{e}^{i\langle\theta,Y\rangle}) =\\ = \mathrm{E}(\mathrm{e}^{i\langle\theta,X\rangle})\mathrm{E}(\mathrm{e}^{i\langle\theta,Y\rangle}) = \phi_\mu(\theta)\phi_\nu(\theta) .\end{aligned}$$

Inoltre

$$\phi_{-X}(\theta) = \mathrm{E}(\mathrm{e}^{-i\langle\theta,X\rangle}) = \mathrm{E}(\overline{\mathrm{e}^{i\langle\theta,X\rangle}}) = \overline{\phi_X(\theta)} . \tag{2.17}$$

Quindi se X è una v.a. simmetrica (cioè tale che X e $-X$ hanno la stessa legge) allora ϕ_X è una funzione a valori reali.

Se $Y = AX + b$, dove A è una matrice $p \times m$ e $b \in \mathbb{R}^p$, Y è una v.a. a valori in $\mathbb{R}^p$ e per $\theta \in \mathbb{R}^p$

$$\phi_Y(\theta) = \mathrm{E}(\mathrm{e}^{i\langle\theta,AX+b\rangle}) = \mathrm{e}^{i\langle\theta,b\rangle}\mathrm{E}(\mathrm{e}^{i\langle A^*\theta,X\rangle}) = \phi_X(A^*\theta)\mathrm{e}^{i\langle\theta,b\rangle} . \tag{2.18}$$

Esempi 2.16 Negli esempi seguenti $m = 1$ e quindi $\theta \in \mathbb{R}$. Per il calcolo useremo sempre la (2.15).

a) Binomiale $B(n, p)$: per la regola del binomio

$$\phi(\theta) = \sum_{k=0}^{n} \binom{n}{k} p^k(1-p)^{n-k}\mathrm{e}^{i\theta k} = \sum_{k=0}^{n} \binom{n}{k} (p\mathrm{e}^{i\theta})^k(1-p)^{n-k} = (1-p+p\mathrm{e}^{i\theta})^n .$$

b) Geometrica

$$\phi(\theta) = \sum_{k=0}^{\infty} p(1-p)^k\mathrm{e}^{i\theta k} = \sum_{k=0}^{\infty} p((1-p)\mathrm{e}^{i\theta})^k = \frac{p}{1-(1-p)\mathrm{e}^{i\theta}} .$$

c) Poisson

$$\phi(\theta) = \mathrm{e}^{-\lambda}\sum_{k=0}^{\infty} \frac{\lambda^k}{k!}\mathrm{e}^{i\theta k} = \mathrm{e}^{-\lambda}\sum_{k=0}^{\infty} \frac{(\lambda\mathrm{e}^{i\theta})^k}{k!} = \mathrm{e}^{-\lambda}\mathrm{e}^{\lambda\mathrm{e}^{i\theta}} = \mathrm{e}^{\lambda(\mathrm{e}^{i\theta}-1)} .$$

d) Esponenziale

$$\phi(\theta)=\lambda\int_0^{+\infty}\mathrm{e}^{-\lambda x}\mathrm{e}^{i\theta x}\,dx=\lambda\int_0^{+\infty}\mathrm{e}^{x(i\theta-\lambda)}\,dx=\frac{\lambda}{i\theta-\lambda}\mathrm{e}^{x(i\theta-\lambda)}\Big|_0^{+\infty}=$$
$$=\frac{\lambda}{i\theta-\lambda}\Big(\lim_{x\to+\infty}\mathrm{e}^{x(i\theta-\lambda)}-1\Big)\ .$$

Ma il numero complesso $\mathrm{e}^{x(i\theta-\lambda)}$ ha modulo $|\mathrm{e}^{x(i\theta-\lambda)}|=\mathrm{e}^{-\lambda x}$ che tende a 0 per $x\to+\infty$, dunque $\lim_{x\to+\infty}\mathrm{e}^{x(i\theta-\lambda)}=0$ e

$$\phi(\theta)=\frac{\lambda}{\lambda-i\theta}\ \cdot$$

(Nel calcolo precedente siamo stati un po' sbrigativi dando per scontato che l'integrazione dell'esponenziale complesso si faccia come per l'esponenziale reale. È però facile controllare che l'integrazione è corretta scomponendo in parte reale e parte immaginaria).

Vediamo ora quali sono le proprietà di regolarità delle funzioni caratteristiche. Da (2.15) si può vedere $\hat\mu$ come una funzione definita da un integrale dipendente da un parametro.

Cominciamo con le proprietà di continuità. Si ha

$$|\hat\mu(\theta)-\hat\mu(\theta_0)|=|\mathrm{E}(\mathrm{e}^{i\langle\theta,X\rangle})-\mathrm{E}(\mathrm{e}^{i\langle\theta_0,X\rangle})|\le\mathrm{E}(|\mathrm{e}^{i\langle\theta,X\rangle}-\mathrm{e}^{i\langle\theta_0,X\rangle}|)\ .$$

Se facciamo tendere $\theta\to\theta_0$, allora $|\mathrm{e}^{i\langle\theta,X\rangle}-\mathrm{e}^{i\langle\theta_0,X\rangle}|\to 0$. Poiché $|\mathrm{e}^{i\langle\theta,X\rangle}-\mathrm{e}^{i\langle\theta_0,X\rangle}|\le 2$, si può applicare il Teorema di Lebesgue, e quindi

$$\lim_{\theta\to\theta_0}|\hat\mu(\theta)-\hat\mu(\theta_0)|=0$$

che prova che $\hat\mu$ è continua. Si può anzi dimostrare che $\hat\mu$ è uniformemente continua (vedi Esercizio 2.24).

Per studiare la derivabilità, supponiamo dapprima $m=1$ (cioè che μ sia una probabilità su $\mathbb{R}$). La Proposizione 1.16 (derivabilità degli integrali dipendenti da un parametro afferma che perché $\theta\to\mathrm{E}[f(X,\theta)]$ sia derivabile basta che $\frac{\partial f}{\partial\theta}$ esista e che valga la maggiorazione.

$$\sup_{\theta\in\mathbb{R}}\Big|\frac{\partial}{\partial\theta}f(\theta,x)\Big|\le g(x)$$

dove g è una funzione tale che $g(X)$ sia integrabile. In questo caso

$$\Big|\frac{\partial}{\partial\theta}\mathrm{e}^{i\theta x}\Big|=|ix\mathrm{e}^{i\theta x}|=|x|\ .$$

Dunque se X è integrabile, per la Proposizione 1.16 μ è derivabile e si può derivare sotto il segno; cioè

$$\hat\mu'(\theta)=\mathrm{E}(iX\mathrm{e}^{i\theta X})=\int ix\mathrm{e}^{i\theta x}\,\mu(dx)\ . \tag{2.19}$$

Ripetendo lo stesso ragionamento per l'integrando $f(\theta, x) = ix\mathrm{e}^{i\theta x}$ si trova

$$\Big|\frac{\partial}{\partial\theta} ix\mathrm{e}^{i\theta x}\Big| = |ix\mathrm{e}^{i\theta x}| = |x|^2 .$$

Dunque, se X ha momento del second'ordine finito, $\hat{\mu}$ è due volte derivabile e si ha

$$\hat{\mu}''(\theta) = -\int x^2\mathrm{e}^{i\theta x}\,\mu(dx) . \tag{2.20}$$

Ripetendo questo ragionamento si vede facilmente per ricorrenza che se μ ha momento assoluto di ordine k finito, allora $\hat{\mu}$ è k volte derivabile e

$$\hat{\mu}^{(k)}(\theta) = \int (ix)^k\mathrm{e}^{i\theta x}\,\mu(dx) . \tag{2.21}$$

Più precisamente vale il risultato seguente

Proposizione 2.17 *Se μ ha momento di ordine k finito allora $\hat{\mu}$ è k volte derivabile e vale* (2.21)*. Viceversa se $\hat{\mu}$ è k volte derivabile e k è pari allora μ ha momento di ordine k finito e (quindi) vale la* (2.21)*.*

Dimostrazione. La prima affermazione è già stata provata. Poiché ϕ è due volte derivabile sappiamo che

$$\lim_{\theta\to 0}\frac{\phi(\theta)+\phi(-\theta)-2\phi(0)}{\theta^2} = \phi''(0)$$

(basta sostituire a ϕ il suo sviluppo di Taylor al second'ordine). Ma

$$\frac{2\phi(0)-\phi(\theta)-\phi(-\theta)}{\theta^2} = \int\frac{2-\mathrm{e}^{i\theta x}-\mathrm{e}^{-i\theta x}}{\theta^2}\,\mu(dx) = \int 2\,\frac{1-\cos(\theta x)}{x^2\theta^2}x^2\,\mu(dx)$$

L'integrando dell'ultimo integrale è positivo e converge a x^2 per $\theta\to 0$. Dunque per il Lemma di Fatou

$$-\phi''(0) \ge \int x^2\,\mu(dx)$$

che dimostra che μ ha momento del second'ordine finito. Inoltre, grazie alla prima parte dell'enunciato,

$$\phi''(\theta) = \int (ix)^2\mathrm{e}^{i\theta x}\,\mu(dx) .$$

La dimostrazione si completa facilmente per induzione: supponiamo di avere dimostrato che se $\hat{\mu}$ è k volte derivabile (k pari) allora μ ha momento di ordine k finito e vale

$$\phi^{(k)}(\theta) = \int (ix)^k\mathrm{e}^{i\theta x}\mu(dx) .$$

Allora, se ϕ è $k+2$ volte derivabile,

$$\lim_{\theta\to 0} \frac{\phi^{(k)}(\theta)+\phi^{(k)}(-\theta)-2\phi^{(k)}(0)}{\theta^2} = \phi^{(k+2)}(0)$$

e dunque

$$\frac{2\phi^{(k)}(0)-\phi^{(k)}(\theta)-\phi^{(k)}(-\theta)}{\theta^2} = \int \frac{2-\mathrm{e}^{i\theta x}-\mathrm{e}^{-i\theta x}}{\theta^2}\,(ix)^k\,\mu(dx) = \\ = \int 2\,\frac{1-\cos(\theta x)}{x^2\theta^2}\,i^k x^{k+2}\,\mu(dx)$$

da cui, per $\theta \to 0$ e usando come prima il lemma di Fatou, si ha

$$i^k\phi^{(k+2)}(0) \ge \int x^{k+2}\,\mu(dx)$$

(poiché k è pari, $i^k = i^{-k}$).

Osservazione 2.18 Uno sguardo alla dimostrazione precedente permette d'indebolire le ipotesi: se k è pari basta che ϕ sia derivabile k volte *nell'origine* perché esista finito il momento di ordine k di μ. In particolare se ϕ è derivabile k volte in 0 e k è pari, allora essa è derivabile k volte ovunque.

Per $\theta = 0$ (2.21) diviene

$$\hat{\mu}^{(k)}(0) = i^k \int x^k \mu(dx) \tag{2.22}$$

che permette di calcolare i momenti di μ semplicemente derivando $\hat{\mu}$ in 0. Attenzione però: si conoscono esempi in cui ϕ è derivabile senza che X abbia speranza matematica finita. Se invece ϕ è due volte derivabile, per la Proposizione 2.17 (2 è pari), X ha momento di ordine 2 finito (e dunque anche speranza matematica finita).

Ragionamenti simili (solo più complicati da esprimere) danno risultati analoghi nel caso di probabilità su $\mathbb{R}^m$. Più precisamente se $\alpha = (\alpha_1, \dots, \alpha_m)$ è un multiindice e indichiamo al solito

$$|\alpha| = \alpha_1 + \ldots + \alpha_m \qquad x^\alpha = x_1^{\alpha_1}\ldots x_m^{\alpha_m}$$
$$\frac{\partial^\alpha}{\partial\theta\alpha} = \frac{\partial^{\alpha_1}}{\partial\theta^{\alpha_1}}\cdots\frac{\partial^{\alpha_m}}{\partial\theta^{\alpha_m}}$$

allora se

$$\int |x|^{|\alpha|}\,\mu(dx) < +\infty$$

$\hat{\mu}$ è α volte derivabile e

$$\frac{\partial^\alpha}{\partial\theta^\alpha}\hat{\mu}(\theta) = \int (ix)^\alpha e^{i\langle\theta,x\rangle}\,\mu(dx)\ .$$

In particolare

$$\frac{\partial}{\partial\theta_j}\hat{\mu}(0) = i\int x_j\,\mu(dx)$$
$$\frac{\partial^2}{\partial\theta_j\partial\theta_h}\hat{\mu}(0) = -\int x_h x_j\,\mu(dx)\ .$$

Cioè il gradiente di μ all'origine è i volte la media e, se μ è centrata, lo Hessiano di $\hat{\mu}$ all'origine è uguale alla matrice di covarianza cambiata di segno.

Esempio 2.19 (Funzione caratteristica di una legge normale) Se μ è $N(0,1)$ calcoliamo

$$\hat{\mu}(\theta) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} e^{i\theta x}e^{-x^2/2}\,dx\ . \tag{2.23}$$

Questo integrale si calcola direttamente col metodo dei residui oppure nel modo seguente.
Poiché μ ha media finita possiamo applicare (2.19) e, integrando per parti,

$$\hat{\mu}'(\theta) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} ix e^{i\theta x}e^{-x^2/2}\,dx =$$
$$= \underbrace{-\frac{1}{\sqrt{2\pi}} i e^{i\theta x}e^{-x^2/2}\Big|_{-\infty}^{+\infty}}_{=0} + \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} i\cdot i\theta e^{i\theta x}e^{-x^2/2}\,dx = -\theta\hat{\mu}(\theta)\ .$$

Cioè $\hat{\mu}$ è soluzione dell'equazione differenziale lineare

$$u'(\theta) = -\theta u(\theta)$$

con la condizione iniziale $u(0) = 1$. Risolvendola si ha facilmente

$$\hat{\mu}(\theta) = e^{-\theta^2/2}\ .$$

Se invece Y è $N(m,\sigma^2)$, sappiamo che Y si può scrivere $Y = \sigma X + m$, dove X è $N(0,1)$ e dunque per la (2.18)

$$\phi_Y(\theta) = e^{-\frac{1}{2}\sigma^2\theta^2}e^{i\theta m}\ .$$

La proprietà fondamentale delle funzioni caratteristiche è la seguente

Teorema 2.20 *Siano μ e ν leggi di probabilità su $\mathbb{R}^m$ tali che*

$$\hat{\mu}(\theta) = \hat{\nu}(\theta) \qquad \text{per ogni } \theta \in \mathbb{R}\,.$$

Allora $\mu = \nu$.

Osserviamo che la relazione $\hat{\mu}(\theta) = \hat{\nu}(\theta)$ per ogni $\theta \in \mathbb{R}$ implica che si ha

$$\int f\,d\mu = \int f\,d\nu \tag{2.24}$$

per ogni funzione f della forma $f(x) = \mathrm{e}^{i\langle\theta,x\rangle}$. Per dimostrare il Teorema 2.20 invece basterebbe provare la (2.24) per ogni funzione f continua a supporto compatto (Lemma 1.19).

Nella dimostrazione del Teorema 2.20 useremo un risultato di analisi matematica che è una versione del Teorema di Stone-Weierstrass. Indichiamo con $\mathscr{C}_0 = \mathscr{C}_0(\mathbb{R}^n, \mathbb{C})$ lo spazio delle funzioni $\mathbb{R}^n \to \mathbb{C}$ che sono continue e nulle all'infinito, munito della norma

$$\|f\|_\infty = \sup_{x\in\mathbb{R}^n} |f(x)|\,.$$

Si ha allora

Teorema 2.21 *(Stone-Weierstrass) Sia $\mathscr{X} \subset \mathscr{C}_0$ una famiglia di funzioni tale che*

a) $\mathscr{X}$ è un'algebra, cioè combinazioni lineari e prodotti di funzioni di $\mathscr{X}$ appartengono ancora ad $\mathscr{X}$.

b) $\mathscr{X}$ è stabile per coniugazione: se $f \in \mathscr{X}$, allora $\bar{f} \in \mathscr{X}$.

c) $\mathscr{X}$ separa i punti, cioè dati $x, y \in \mathbb{R}^n$, esiste $f \in \mathscr{X}$ tale che $f(x) \neq f(y)$.

Allora $\mathscr{X}$ è densa in $\mathscr{C}_0$.

Dimostrazione del Teorema 2.20. Consideriamo la famiglia $\mathscr{X}$ formata dalle combinazioni lineari di funzioni della forma

$$f(x) = \mathrm{e}^{i\theta x}\mathrm{e}^{-\frac{1}{2}\sigma^2|x|^2}\,. \tag{2.25}$$

al variare di $\theta \in \mathbb{R}^n$ e $\sigma^2 > 0$. È chiaro che si tratta di funzioni nulle all'infinito ed inoltre che $\mathscr{X}$ è un'algebra di funzioni, stabile per coniugazione e che separa i punti.

Mostriamo che la (2.24) vale per le funzioni di $\mathscr{X}$. Supporremo per semplicità $m = 1$. Ricordando la funzione caratteristica di una legge normale di media θ e varianza σ^2, si ha

$$\mathrm{e}^{i\theta x}\mathrm{e}^{-\frac{1}{2}\sigma^2|x|^2} = \frac{1}{(2\pi)^{1/2}\sigma}\int \mathrm{e}^{iyx}\mathrm{e}^{-\frac{1}{2\sigma^2}|y-\theta|^2}\,dy$$

e dunque

$$\int \mathrm{e}^{i\theta x}\mathrm{e}^{-\frac{1}{2}\sigma^2|x|^2}\mu(dx) = \frac{1}{(2\pi)^{1/2}\sigma}\int \mu(dx)\int \mathrm{e}^{iyx}\mathrm{e}^{-\frac{1}{2\sigma^2}|y-\theta|^2}\,dy$$

e, per il teorema di Fubini,

$$\dots = \frac{1}{(2\pi)^{1/2}\sigma}\int e^{-\frac{1}{2\sigma^2}|y-\theta|^2}\,dy\int e^{iyx}\mu(dx) = \frac{1}{(2\pi)^{1/2}\sigma}\int \hat\mu(y)\,e^{-\frac{1}{2\sigma^2}|y-\theta|^2}\,dy$$

Dunque (2.24) è vera per ogni funzione f che sia combinazione lineare di funzioni della forma (2.25). Ora se $f \in \mathscr{C}_K$, per il Teorema 2.21 esiste una successione $(g_n)_n$ di combinazioni lineari di funzioni della forma (2.25) che converge a f uniformemente. Dunque, per n abbastanza grande si ha $|g_n(x)| \le |f(x)| + \varepsilon$ e si può dunque applicare il teorema di Lebesgue:

$$\int f\,d\mu = \lim_{n\to\infty}\int g_n\,d\mu = \lim_{n\to\infty}\int g_n\,d\nu = \int f\,d\nu\ .$$

e dunque la (2.24) è provata.

Esempio 2.22 Siano μ e ν leggi $N(a,\sigma^2)$ e $N(b,\tau^2)$ rispettivamente. Calcoliamo la legge di $\mu * \nu$.

Basta osservare che

$$\phi_{\mu*\nu}(\theta) = \hat\mu(\theta)\hat\nu(\theta) = e^{ia\theta}e^{-\frac12\sigma^2\theta^2}e^{ib\theta}e^{-\frac12\tau^2\theta^2} = e^{i(a+b)\theta}e^{-\frac12(\sigma^2+\tau^2)\theta^2}$$

e dunque $\mu * \nu$ è $N(a+b,\sigma^2+\tau^2)$. Lo stesso risultato si sarebbe potuto ottenere anche calcolando l'integrale di convoluzione della Proposizione 2.17, ma il calcolo dell'integrale relativo, peraltro elementare, non è né corto né divertente.

Il Teorema 2.20 è di grande importanza teorica, ma purtroppo non è costruttivo: esso cioè non dà indicazioni di come si possa, dalla funzione caratteristica $\hat\mu$, ricavare la funzione di ripartizione di μ oppure la sua densità, se esiste. Il teorema seguente, che non dimostriamo, dà una risposta in questo senso.

Teorema 2.23 *(D'inversione) Se $\hat\mu$ è una funzione integrabile allora μ è assolutamente continua rispetto alla misura di Lebesgue ed ha densità data da*

$$f(x) = \frac{1}{(2\pi)^d}\int_{-\infty}^{+\infty} e^{-i\langle\theta,x\rangle}\hat\mu(\theta)\,d\theta\ . \tag{2.26}$$

Siano $X_1,\dots,X_m$ v.a. a valori in $\mathbb{R}^{n_1},\dots,\mathbb{R}^{n_m}$ rispettivamente e consideriamo la v.a. a valori in $\mathbb{R}^k$, $k = n_1+\dots+n_m$, definita da $X = (X_1,\dots,X_m)$. Indichiamone con ϕ_X la funzione caratteristica. È allora facile calcolare la funzione caratteristica ϕ_{X_h} della h-esima marginale di X. In effetti, ricordando che ϕ_X è definita su $\mathbb{R}^m$ mentre ϕ_{X_h} è una funzione di variabile reale

$$\phi_{X_h}(\theta) = \mathrm{E}(e^{i\theta X_h}) = \mathrm{E}(e^{i\langle\tilde\theta,X\rangle}) = \phi_X(\tilde\theta)$$

dove $\tilde{\theta} = (0, \ldots, 0, \theta, 0, \ldots, 0)$ è il vettore di $\mathbb{R}^k$ le cui componenti sono tutte nulle tranne quelle corrispondenti alle coordinate dalla $n_1 + \ldots + n_{h-1} + 1$-esima alla $n_1 + \ldots + n_h$-esima.

Supponiamo che le v.a. $X_1, \ldots, X_m$ siano indipendenti; se $\theta_1 \in \mathbb{R}^{n_1}, \ldots, \theta_m \in \mathbb{R}^{n_m}$ e $\theta = (\theta_1, \ldots, \theta_m) \in \mathbb{R}^k$ allora

$$\phi_X(\theta) = \mathrm{E}(\mathrm{e}^{i\langle\theta, X\rangle}) = \mathrm{E}(\mathrm{e}^{i\langle\theta_1, X_1\rangle} \ldots \mathrm{e}^{i\langle\theta_m, X_m\rangle}) = \phi_{X_1}(\theta_1) \ldots \phi_{X_m}(\theta_m) \ . \tag{2.27}$$

La (2.27) si può anche esprimere in termini di leggi: se $\mu_1, \ldots, \mu_m$ sono leggi di probabilità su $\mathbb{R}^{n_1}, \ldots, \mathbb{R}^{n_m}$ rispettivamente e $\mu = \mu_1 \otimes \ldots \otimes \mu_m$ allora

$$\hat{\mu}(\theta) = \hat{\mu}_1(\theta_1) \ldots \hat{\mu}_m(\theta_m) \ . \tag{2.28}$$

Vale anzi il risultato seguente, più preciso

Proposizione 2.24 *Siano $X_1, \ldots, X_m$ v.a. a valori in $\mathbb{R}^{n_1}, \ldots, \mathbb{R}^{n_m}$ rispettivamente e poniamo $X = (X_1, \ldots, X_m)$. Allora le v.a. $X_1, \ldots, X_m$ sono indipendenti se e solo se per ogni $\theta_1 \in \mathbb{R}^{n_1}, \ldots, \theta_m \in \mathbb{R}^{n_m}$ e $\theta = (\theta_1, \ldots, \theta_m) \in \mathbb{R}^k$, posto $\theta = (\theta_1, \ldots, \theta_m)$ si ha*

$$\phi_X(\theta) = \phi_{X_1}(\theta_1) \ldots \phi_{X_m}(\theta_m) \ . \tag{2.29}$$

Dimostrazione. Se le X_i sono indipendenti abbiamo già visto che vale (2.29). Viceversa, se vale la (2.29), allora X ha la stessa funzione caratteristica che la legge prodotto delle leggi delle X_i. Quindi per il Teorema 2.20 la legge di X è la legge prodotto e le X_i sono indipendenti.

Sia $z \in C$. Se X è una v.a. m dimensionale chiamiamo *trasformata di Laplace complessa* (TLC) la funzione

$$H(z) = \mathrm{E}[\mathrm{e}^{\langle z, X\rangle}] = \int_{\mathbb{R}^m} \mathrm{e}^{\langle z, x\rangle}\, d\mu(x)$$

definita per quei valori $z \in \mathbb{C}$ per cui $\mathrm{e}^{\langle z, X\rangle}$ è integrabile. È chiaro che H è certamente definita sull'asse immaginario e anzi, se $\theta \in \mathbb{R}^m$,

$$H(i\theta) = \phi_X(\theta) \ .$$

Dunque la conoscenza della TLC H implica quella della funzione caratteristica ϕ_X. Si chiama *dominio* della TLC, l'insieme dei numeri $z \in \mathbb{C}^m$ tali che la v.a. $\mathrm{e}^{\langle z, X\rangle}$ sia integrabile, ovvero tali che

$$\int_{\mathbb{R}^m} |\mathrm{e}^{\langle z, x\rangle}|\, d\mu(x) \int_{\mathbb{R}^m} \mathrm{e}^{\mathrm{Re}\langle z, x\rangle}\, d\mu(x) < +\infty \ .$$

Indicheremo con $\mathcal{D}_\mu$ il dominio della TLC di μ.

Esempio 2.25 a) Supponiamo che X sia una v.a. di Cauchy, cioè di densità

$$f(x) = \frac{1}{\pi}\,\frac{1}{1+x^2}$$

allora, se $t \in \mathbb{R}$, si ha

$$H(t) = \frac{1}{\pi}\int_{-\infty}^{+\infty} \frac{e^{tx}}{1+x^2}\,dx$$

e dunque $H(t) = +\infty$ per ogni $t \neq 0$. In questo caso dunque il dominio è ridotto a $\mathrm{Re}\, z = 0$, cioè all'asse immaginario.

b) Supponiamo $X \sim N(0,1)$. Allora, sempre per $t \in \mathbb{R}$,

$$\begin{aligned} H(t) &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} e^{tx}e^{-x^2/2}\,dx = \frac{e^{t^2/2}}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} e^{-\frac{1}{2}(x-t)^2}\,dx = \\ &= \frac{e^{t^2/2}}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} e^{-y^2/2}\,dx = e^{t^2/2}\,. \end{aligned}$$

Dunque in questo caso $\mathcal{D}_\mu = \mathbb{R}$.

c) Se invece $X \sim \Gamma(1,\lambda)$ (esponenziale di parametro λ), allora, per $t \in \mathbb{R}$,

$$H(t) = \lambda\int_0^{+\infty} e^{tx}e^{-\lambda x}\,dx = \lambda\int_0^{+\infty} e^{-(\lambda-t)x}\,dx\,.$$

Questo integrale è convergente se e solo se $t < \lambda$ e dunque $\mathcal{D}_\mu = \{\mathrm{Re}\, z < \lambda\}$.

Per semplicità supporremo $m = 1$ d'ora in avanti. Vediamo come è fatto il dominio in generale. Si ha

$$\int_{-\infty}^{+\infty} e^{\mathrm{Re}\, z\, x}\,d\mu(x) = \int_0^{+\infty} e^{\mathrm{Re}\, z\, x}\,d\mu(x) + \int_{-\infty}^{0} e^{\mathrm{Re}\, z\, x}\,d\mu(x) = I_2 + I_1\,.$$

E' chiaro che se $\mathrm{Re}\, z \geq 0$, allora $I_1 < +\infty$, perché l'integrando è più piccolo di 1. Osserviamo inoltre che la funzione $t \to \int_0^{+\infty} e^{tx}\,d\mu(x)$ è crescente. Quindi se poniamo

$$x_2 = \sup\left\{t \geq 0, \int_0^{+\infty} e^{tx}\,d\mu(x) < +\infty\right\}$$

(eventualmente $x_2 = +\infty$), allora è chiaro che $x_2 \geq 0$ e $I_2 < +\infty$ se $\mathrm{Re}\, z < x_2$, mentre $I_2 = +\infty$ se $\mathrm{Re}\, z > x_2$. Dunque il dominio contiene la striscia $0 \leq \mathrm{Re}\, z < x_2$. Analogamente se $\mathrm{Re}\, z \leq 0$ si vede che esiste un numero $x_1 \leq 0$ tale che $\{0 \geq \mathrm{Re}\, z > x_1\} \subset \mathcal{D}_\mu$ e $z \notin \mathcal{D}_\mu$ se $\mathrm{Re}\, z < x_1$.

In conclusione $\mathcal{D}_\mu$ contiene la striscia $S = \{z; x_1 < \mathrm{Re}\, z < x_2\}$, mentre non contiene i numeri complessi z che si trovano al di fuori della chiusura di S, cioè tali che $\mathrm{Re}\, z > x_2$ oppure $\mathrm{Re}\, z < x_1$.

Si ha anzi il risultato seguente.

Teorema 2.26 *Esistono* $x_1, x_2 \in \overline{\mathbb{R}}$ *con* $x_1 \le 0 \le x_2$ *(eventualmente coincidenti) tali che* H *è definita nella striscia* $S = \{z; x_1 < \operatorname{Re} z < x_2\}$*, mentre non è definita per* $\operatorname{Re} z > x_2$ *oppure per* $\operatorname{Re} z < x_1$*. Inoltre* H *è olomorfa in* S*.* x_1 *e* x_2 *si chiamano le ascisse di convergenza.*

Dimostrazione. La prima affermazione è già stata dimostrata. Per mostrare che la TLC è olomorfa basta mostrare che valgono le equazioni di Cauchy-Riemann cioè, scrivendo $z = x + iy$ e $H = H_1 + iH_2$,

$$\begin{aligned}\frac{\partial H_1}{\partial x} &= \frac{\partial H_2}{\partial y}\\ \frac{\partial H_1}{\partial y} &= -\frac{\partial H_2}{\partial x}\ .\end{aligned}$$

Queste equazioni si ottengono applicando la Proposizione 1.16, di derivazione sotto il segno. Poiché supponiamo $x + iy \in S$, esiste $\varepsilon > 0$ tale che $x_1 + \varepsilon < x < x_2 - \varepsilon$. Si ha

$$H_1(x+iy) = \int \mathrm{e}^{xt} \cos(yt)\, d\mu(t)\ .$$

La derivata dell'integrando rispetto a x vale $t\mathrm{e}^{xt}\cos(yt)$. Ora

$$\begin{aligned}|t|\mathrm{e}^{xt} &\le c_1 \mathrm{e}^{(x_2-\varepsilon)t} \qquad \text{se } t \ge 0\\ |t|\mathrm{e}^{xt} &\le c_2 \mathrm{e}^{(x_1+\varepsilon)t} \qquad \text{se } t \le 0\ .\end{aligned}$$

Dunque la condizione del teorema di derivazione sotto il segno è soddisfatta con $g(x) = c_1\mathrm{e}^{(x_2-\varepsilon)t} + c_2\mathrm{e}^{(x_1+\varepsilon)t}$. Derivando si ottiene

$$\frac{\partial H_1}{\partial x}(x+iy) = \int t\mathrm{e}^{xt}\cos(yt)\, d\mu(t)\ .$$

Allo stesso modo si ragiona per H_2:

$$H_2(x+iy) = \int \mathrm{e}^{xt}\sin(yt)\, d\mu(t)\ ,$$

si verifica che si può derivare sotto il segno e quindi

$$\frac{\partial H_2}{\partial y}(x+iy) = \int t\mathrm{e}^{xt}\cos(yt)\, d\mu(t)\ .$$

Dunque è soddisfatta la prima delle equazioni di Cauchy-Riemann. Allo stesso modo si ragiona per la seconda.

Osserviamo che nel Teorema 2.26 non abbiamo mai utilizzato il fatto che μ sia una misura di probabilità. L'enunciato si applica quindi anche alla TLC di una misura finita qualunque.

Abbiamo visto che per alcune v.a., di Cauchy ad esempio, le ascisse di convergenza possono essere entrambe uguali a 0. Se invece esse non coincidono la proprietà di analiticità della TLC ha interessanti applicazioni.

Ricordiamo ad esempio che una funzione olomorfa è individuata non appena la si conosca su un insieme avente almeno un punto di accumulazione (unicità del prolungamento analitico). Tipicamente, quindi, la conoscenza della trasformata di Laplace sull'asse reale (o su un intervallo) ne determinano il valore su tutta la striscia di convergenza. Ciò fornisce un metodo di calcolo della funzione caratteristica.

Esempio 2.27 Sia X una v.a. $\Gamma(\alpha, \lambda)$ e calcoliamone la TLC. H è definita se

$$\int_{-\infty}^{+\infty} |e^{zx}|\, d\mu(x) = \int_{-\infty}^{+\infty} e^{\mathrm{Re}\, z\, x}\, d\mu(x) < +\infty$$

ovvero se

$$\frac{\lambda^\alpha}{\Gamma(\alpha)} \int_{-\infty}^{+\infty} e^{\mathrm{Re}\, z\, x} x^{\alpha-1} e^{-\lambda x}\, dx < +\infty\ .$$

Questo integrale è convergente se e solo se $\mathrm{Re}\, z < \lambda$. Dunque le ascisse di convergenza sono $x_1 = -\infty$, $x_2 = \lambda$. Calcoliamo la TLC per $t \in \mathbb{R}$, $z < \lambda$.

$$H(t) = \frac{\lambda^\alpha}{\Gamma(\alpha)} \int_0^{+\infty} x^{\alpha-1} e^{-(\lambda-t)x}\, dx = \frac{\lambda^\alpha}{\Gamma(\alpha)} \frac{\Gamma(\alpha)}{(\lambda-t)^\alpha} = \Big(\frac{\lambda}{\lambda - t}\Big)^\alpha\ .$$

Ora la funzione $H : \mathbb{C} \to \mathbb{C}$ definita da

$$H(z) = \Big(\frac{\lambda}{\lambda - z}\Big)^\alpha = e^{\alpha \log \frac{\lambda}{\lambda - z}} \tag{2.30}$$

è olomorfa sull'insieme $D \subset \mathbb{C}$ formato dai numeri complessi z tali che $\frac{\lambda}{\lambda - z}$ non sia un reale negativo. È facile però vedere che $\frac{\lambda}{\lambda - z}$ è reale negativo solo se z è reale $> \lambda$. Dunque H definita in (2.30) è olomorfa su $\{\mathrm{Re}\, z < \lambda\}$ e coincide con la trasformata di Laplace di X sul'asse reale. Per l'unicità del prolungamento analitico dunque H è la trasformata di Laplace di X *per ogni numero complesso* z tale che $\mathrm{Re}\, z < \lambda$.

In particolare la funzione caratteristica di una legge $\Gamma(\alpha, \lambda)$ è

$$\phi(t) = H(it) = \Big(\frac{\lambda}{\lambda - it}\Big)^\alpha\ .$$

Esempio 2.28 Un altro modo per calcolare la funzione caratteristica di una v.a. gaussiana. Se $X \sim N(0, 1)$, allora abbiamo visto nell'Esempio 2.25 b) che la sua trasformata di Laplace calcolata in $z \in \mathbb{R}$ vale

$$H(z) = e^{z^2/2}$$

Dunque, per l'unicità del prolungamento analitico, $H(z) = \mathrm{e}^{z^2/2}$ per ogni $z \in \mathbb{C}$. Quindi la funzione caratteristica è $\phi_X(\theta) = H(i\theta) = \mathrm{e}^{-\theta^2/2}$.

Se le ascisse di convergenza sono entrambe diverse da 0, allora la TLC è analitica in 0, grazie al Teorema 2.26. Ne segue che la funzione caratteristica $\phi_X(t) = H(it)$ è infinite volte derivabile e dunque la v.a. X ha finiti i momenti di tutti gli ordini, per il Teorema 2.17. Inoltre poiché

$$iH'(0) = \phi_X{}'(0) = i\mathrm{E}(X)$$

si ha che $H'(0) = \mathrm{E}(X)$. Anche gli altri momenti della v.a. X si possono ottenere derivando la TLC: si vede facilmente che

$$H^{(k)}(0) = \mathrm{E}(X^k) \ .$$

Osserviamo infine che se X è una v.a. a valori interi ≥ 0 e ψ_X è la relativa funzione generatrice, allora

$$\psi_X(t) = \mathrm{E}(t^X) = \mathrm{E}(\mathrm{e}^{X \log t}) = H(\log t) \ .$$

2.5 Leggi normali multivariate

Siano $X_1, \dots, X_m$ v.a. i.i.d di legge $N(0,1)$; allora il vettore $X = (X_1, \dots, X_m)$ ha densità

$$f(x) = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{1}{2}x_1^2} \dots \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{1}{2}x_m^2} = \frac{1}{(2\pi)^{m/2}} \mathrm{e}^{-\frac{1}{2}|x|^2} \ .$$

Inoltre per (2.29)

$$\phi_X(\theta) = \mathrm{e}^{-\frac{1}{2}\theta_1^2} \dots \mathrm{e}^{-\frac{1}{2}\theta_m^2} = \mathrm{e}^{-\frac{1}{2}|\theta|^2} \ .$$

Se ora A è una matrice $m \times m$ e $z \in \mathbb{R}^m$, e poniamo $Y = AX + z$, per (2.18)

$$\phi_Y(\theta) = \mathrm{e}^{i\langle\theta,z\rangle}\phi_X(A^*\theta) = \mathrm{e}^{i\langle\theta,z\rangle}\mathrm{e}^{-\frac{1}{2}|A^*\theta|^2} = \mathrm{e}^{i\langle\theta,z\rangle}\mathrm{e}^{-\frac{1}{2}\langle A^*\theta,A^*\theta\rangle} = \mathrm{e}^{i\langle\theta,z\rangle}\mathrm{e}^{-\frac{1}{2}\langle AA^*\theta,\theta\rangle} \ .$$

Osserviamo che la matrice AA^* è simmetrica e semi-definita positiva. Si tratta anzi della matrice di covarianza C_Y di Y: per l'Esercizio 3.17 in effetti $C_Y = AC_XA^*$, ed in questo caso C_X è la matrice identica I.

Definizione e Proposizione 2.29 *Dati un vettore $z \in \mathbb{R}$ ed una matrice C, $m \times m$ simmetrica e semi-definita positiva, esiste sempre una legge di probabilità μ su $\mathbb{R}^m$ tale che*

$$\hat{\mu}(\theta) = \mathrm{e}^{i\langle\theta,z\rangle}\mathrm{e}^{-\frac{1}{2}\langle C\theta,\theta\rangle} \ .$$

Si dirà che μ è $N(z,C)$ (normale di media z e di matrice di covarianza C).

Dimostrazione. Per quanto già osservato basta mostrare che esiste una matrice A tale che $AA^* = C$. È un classico risultato di algebra che una tale matrice esiste sempre (purché C sia simmetrica e semidefinita positiva) e che anzi essa può essere scelta simmetrica (e quindi tale che $A^2 = C$; in questo caso si dice che A è la radice quadrata di C). Infatti se C è diagonale

$$C = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_m \end{pmatrix}$$

poiché gli autovalori λ_i sono tutti ≥ 0 (C è semi-definita positiva) basta porre

$$A = \begin{pmatrix} \sqrt{\lambda_1} & & 0 \\ & \ddots & \\ 0 & & \sqrt{\lambda_m} \end{pmatrix} .$$

Altrimenti (cioè se C non è diagonale) esiste una matrice ortogonale O tale che OCO^{-1} sia diagonale. Si verifica subito che OCO^{-1} è ancora semi-definita positiva e quindi esiste una matrice B tale che $B^2 = OCO^{-1}$. Poniamo allora $A = O^{-1}BO$; A è simmetrica (perché $O^{-1} = O^*$) ed è la matrice cercata poiché

$$A^2 = O^{-1}BO \cdot O^{-1}BO = O^{-1}B^2O = C .$$

Mostriamo che se $Y \sim N(z, C)$, allora z è effettivamente la media di Y e C la matrice di covarianza. Questo fatto è ovvio se $z = 0$ e $C = I$, per come abbiamo definito le leggi $N(0, I)$. In generale invece possiamo scrivere $Y = z + AX$ dove A è la radice quadrata di C. Dunque $\mathrm{E}(Y) = \mathrm{E}(z + AX) = z + A\mathrm{E}(X) = z$. La matrice di covarianza di Y è invece uguale a $AIA^* = AA^* = C$ (vedi Esercizio 3.17). Media e matrice di covarianza si sarebbero potute calcolare anche derivando la funzione caratteristica.

Se C è invertibile allora la legge $N(z, C)$ ha densità; infatti in questo caso anche la radice quadrata A di C è invertibile e se Y è $N(z, C)$, allora Y è della forma $AX + z$, dove X è $N(0, I)$; quindi Y ha densità

$$g(y) = \frac{1}{|\det A|} f\big(A^{-1}(y - z)\big) = \frac{1}{(2\pi)^{m/2}(\det C)^{1/2}}\, \mathrm{e}^{-\frac{1}{2}\langle C^{-1}(y-z), y-z\rangle} .$$

Se $X \sim N(z, C)$ e A è una matrice $p \times m$, $b \in \mathbb{R}^p$, allora la v.a. p-dimensionale $Y = AX + b$ ha funzione caratteristica data da (vedi (2.18))

$$\begin{aligned} \phi_Y(\theta) &= \mathrm{e}^{i\langle\theta,b\rangle}\phi_X(A^*\theta) = \mathrm{e}^{i\langle\theta,b\rangle}\mathrm{e}^{i\langle A^*\theta,z\rangle}\mathrm{e}^{-\frac{1}{2}\langle CA^*\theta,A^*\theta\rangle} = \\ &= \mathrm{e}^{i\langle\theta,b+Az\rangle}\mathrm{e}^{-\frac{1}{2}\langle ACA^*\theta,\theta\rangle} \end{aligned} \tag{2.31}$$

e dunque $Y \sim N(b + Az, ACA^*)$. Le trasformazioni affini quindi trasformano leggi normali in leggi normali.

In particolare se $X = (X_1, \dots, X_m)$ è $N(z, C)$, allora per $1 \le i \le m$ si può scrivere $X_i = A_i X$ dove A è la matrice $1 \times m$ che ha tutte le componenti $= 0$ tranne la i-esima che è $= 1$ (e quindi A_i è la matrice corrispondente alla proiezione sulla i-esima coordinata). Dunque X_i, $i = 1, \dots, m$, ha legge normale. Ovvero *le marginali di una legge normale multivariata sono ancora normali*. Tenendo conto inoltre che X_i ha media z_i e covarianza c_{ii}, X_i è $N(z_i, c_{ii})$.

Se X è $N(0, I)$ e O è una matrice ortogonale allora $OIO^* = OO^* = I$ e quindi la v.a. OX è ancora $N(0, I)$. Ovvero le leggi $N(0, I)$ sono invarianti per trasformazioni ortogonali. In altre parole se $X = (X_1, \dots, X_m$ è $N(0, I)$ e $Y = (Y_1, \dots, Y_m)$ è la stessa v.a. in un'altra base ortonormale, allora Y è anch'essa $N(0, I)$.

Sia $X \sim N(z, C)$ e supponiamo che C sia diagonale. Allora, indicando con λ_h gli elementi sulla diagonale di C, si ha

$$\begin{aligned}\phi_X(\theta) &= \mathrm{e}^{i\langle\theta,z\rangle}\mathrm{e}^{-\frac{1}{2}\langle C\theta,\theta\rangle} = \mathrm{e}^{i\langle\theta,z\rangle}\exp\Big(-\frac{1}{2}\sum_{h=1}^m \lambda_h\theta_h^2\Big) = \\ &= \mathrm{e}^{i\theta_1 z_1}\mathrm{e}^{-\frac{1}{2}\lambda_1\theta_1^2}\dots\mathrm{e}^{i\theta_m z_m}\mathrm{e}^{-\frac{1}{2}\lambda_m\theta_m^2} = \phi_{X_1}(\theta_1)\dots\phi_{X_m}(\theta_m)\ .\end{aligned}$$

Per la Proposizione 2.24 quindi le v.a. $X_1, \dots, X_m$ sono indipendenti. Ricordando che C è la matrice di covarianza di X, abbiamo dunque provato che *variabili aleatorie non correlate sono anche indipendenti se la loro distribuzione congiunta è normale*.

Attenzione comunque perché le v.a. $X_1, \dots, X_m$ possono avere ciascuna distribuzione normale senza che la distribuzione congiunta lo sia (vedi Esercizio 2.36).

Il criterio d'indipendenza della Proposizione 2.24 applicato alle v.a. gaussiane multivariate dà anzi il seguente risultato più preciso. Supponiamo che le v.a. X, Y a valori in $\mathbb{R}^n$, $\mathbb{R}^m$ rispettivamente siano tali che la v.a. (X, Y) (a valori in $\mathbb{R}^k$, $k = n + m$) abbia distribuzione normale. Allora se per ogni $1 \le i \le n$ e $1 \le j \le m$ si ha

$$\mathrm{Cov}(X_i, Y_j) = 0 \tag{2.32}$$

le v.a. X e Y sono indipendenti.

Infatti la (2.32) è equivalente a supporre che la matrice di covarianza C di (X, Y) sia diagonale a blocchi

$$C = \begin{pmatrix} & & & 0 & \dots & 0 \\ & C_X & & \vdots & \ddots & \vdots \\ & & & 0 & \dots & 0 \\ 0 & \dots & 0 & & & \\ \vdots & \ddots & \vdots & & C_Y & \\ 0 & \dots & 0 & & & \end{pmatrix}$$

per cui se $\theta_1 \in \mathbb{R}^n, \theta_2 \in \mathbb{R}^m, (\theta_1, \theta_2) \in \mathbb{R}^k$ allora

$$e^{-\frac{1}{2}\langle C\theta,\theta\rangle} = e^{-\frac{1}{2}\langle C_X\theta_1,\theta_1\rangle}\, e^{-\frac{1}{2}\langle C_Y\theta_2,\theta_2\rangle}$$

che implica

$$\phi_{(X,Y)}(\theta) = \phi_X(\theta_1)\phi_Y(\theta_2)\ .$$

Dunque X e Y sono indipendenti per la proposizione 2.24. Richiamiamo ora alcune nozione sui proiettori ortogonali, di cui ci serviremo.

Due sottospazi vettoriali E e F di $\mathbb{R}^n$ si dicono *ortogonali* se ogni vettore di E è ortogonale ad ogni vettore di F. Se E è un sottospazio vettoriale di $\mathbb{R}^m$ si indica con $E^\perp$ il suo ortogonale, cioè l'insieme di tutti i vettori x di $\mathbb{R}^m$ tali che $\langle x, z\rangle = 0$ per ogni $z \in E$. $E^\perp$ è anch'esso un sottospazio vettoriale di $\mathbb{R}^m$ ed ha dimensione $m - k$, se k è la dimensione di E. Inoltre ogni $x \in \mathbb{R}^m$ si può scrivere in maniera unica nella forma $x = x_1 + x_2$, dove $x_1 \in E,\ x_2 \in E^\perp$.

Indichiamo con P_E il *proiettore ortogonale* su E, cioè l'applicazione $\mathrm{P}_E : x \to x_1$ che ad ogni $x \in \mathbb{R}^m$ associa la sua componente su E. È immediato verificare che P_E è un operatore lineare.

Esempio 2.30 Sia E il sottospazio di $\mathbb{R}^m$ dei vettori le cui ultime $m - k$ coordinate sono nulle, cioè dei vettori della forma $(x_1, \dots, x_k, 0, \dots, 0)$. Si tratta chiaramente di un sottospazio di dimensione k. Il suo ortogonale $E^\perp$ è costituito dai vettori della forma $(0, \dots, 0, x_{k+1}, \dots, x_m)$. L'ortogonalità dei due sottospazi è immediata perché facendo il prodotto scalare tutti i termini nella somma sono nulli. In questo esempio se $x = (x_1, \dots, x_m)$

$$\begin{aligned} P_E x &= (x_1, \dots, x_k, 0, \dots, 0) \\ P_{E^\perp} x &= (0, \dots, 0, x_{k+1}, \dots, x_m) \end{aligned}$$

Sono immediate le relazioni (I indica la matrice identità)

$$\begin{aligned} P_E P_E &= P_E \\ I - P_E &= P_{E^\perp} \end{aligned} \tag{2.33}$$

La prima delle ('2')08 segue dal fatto che $P_E x = x$ se $x \in E$. La seconda è particolarmente utile perché permette di calcolare immediatamente $P_{E^\perp}$ a partire da P_E.

Il lemma seguente dà un utile metodo di calcolo dei proiettori ortogonali.

Lemma 2.31 *$P_E x$ è il vettore di E che si trova a distanza minima da x*

Dimostrazione. Se y è un generico vettore in E e $x = x_1 + x_2$ con $x_1 \in E$, $x_2 \in E^\perp$ allora

$$\begin{aligned} |y - x|^2 &= |(y - x_1) - x_2|^2 = \langle (y - x_1) - x_2, (y - x_1) - x_2\rangle = \\ &= |y - x_1|^2 - 2\underbrace{\langle y - x_1, x_2\rangle}_{=0} + |x_2|^2 = |y - x_1|^2 + |x_2|^2 \end{aligned}$$

($\langle y - x_1, x_2 \rangle = 0$ perché $y - x_1 \in E$ mentre $x_2 \in E^{\perp}$). La quantità $|y - x|^2$ è dunque sempre $\geq |x_2|^2$ ed è esattamente uguale a $|x_2|^2$ se e solo se $y = x_1 = P_E x$.

Teorema 2.32 (Cochran) *Sia X una v.a. $N(0, I)$ a valori in $\mathbb{R}^m$ e siano $E_1, \ldots, E_k$ sottospazi vettoriali di $\mathbb{R}^m$ a due a due ortogonali. Per $i = 1, \ldots, k$ indichiamo con n_i la dimensione di E_i e con P_i il proiettore ortogonale su E_i. Allora le v.a. $P_i X$, $i = 1, \ldots k$ sono indipendenti e la v.a $|P_i X|^2$ ha distribuzione $\chi^2(n_i)$.*

Dimostrazione. Supponiamo per semplicità $k = 2$. A meno di una rotazione possiamo supporre che E_1 sia il sottospazio relativo alle prime n_1 coordinate e E_2 quello relativo alle successive n_2 (ricordiamo che le trasformazioni ortogonali lasciano invarianti le leggi $N(0, I)$). Dunque

$$
\begin{aligned}
&P_1 X = (X_1, \ldots, X_{n_1}, 0, \ldots, 0) \\
&P_2 X - (0, \ldots, 0, X_{n_1+1}, \ldots, X_{n_1+n_2}, 0, \ldots, 0)
\end{aligned}
$$

$P_1 X$ e $P_2 X$ sono congiuntamente normali (il vettore $(P_1 X, P_2 X)$ è una funzione lineare di X) ed è chiaro che (2.32) è verificata; dunque $P_1 X$ e $P_2 X$ sono indipendenti. Inoltre

$$
\begin{aligned}
&|P_1 X|^2 = (X_1^2 + \ldots + X_{n_1}^2) \sim \chi^2(n_1) \\
&|P_2 X|^2 = (X_{n_1+1}^2 + \ldots + X_{n_1+n_2}^2) \sim \chi^2(n_1) \ .
\end{aligned}
$$

Una prima applicazione importante del Teorema di Cochran è la seguente.

Indichiamo con V il sottospazio di $\mathbb{R}^m$ generato dal vettore $e = (1, 1, \ldots, 1)$ (cioè il sottospazio dei vettori aventi tutte le componenti uguali); il proiettore ortogonale $P_V : \mathbb{R}^m \to V$ è dato da $P_V x = (\bar{x}, \bar{x}, \ldots, \bar{x})$ dove

$$
\bar{x} = \frac{1}{m}(x_1 + \ldots + x_m) \ .
$$

Ricordiamo infatti che $P_V x$ è il vettore di V che ha distanza minima da x. Per determinare $P_V x$ occorre quindi calcolare $\lambda_0 \in \mathbb{R}$ tale che la funzione $\lambda \to |x - \lambda e|$ abbia minimo in $\lambda = \lambda_0$. Ovvero occorre calcolare il punto di minimo di

$$
\lambda \to \sum_{i=1}^{m} (x_i - \lambda)^2 \ .
$$

Derivando vediamo che deve essere $2 \sum_i (x_i - \lambda) = 0$ e cioè $\lambda = \bar{x}$.

Se $X \sim N(0, I)$ e $\bar{X} = \frac{1}{m}(X_1 + \ldots + X_m)$, allora $\bar{X} e$ è la proiezione ortogonale di X su V; quindi $X - \bar{X} e$ è la proiezione ortogonale di X sul sottospazio ortogonale a V. Per il Teorema 2.32 $\bar{X}e$ e $X - \bar{X}e$ sono indipendenti, che non è una cosa proprio evidente poiché entrambe queste v.a. dipendono da $\bar{X}$. Inoltre

$$\sum_{i=1}^{m}(X_i - \bar{X})^2 = |X - \bar{X}e|^2 \sim \chi^2(m-1)\ . \tag{2.34}$$

2.6 Statistica dei modelli gaussiani

In questo paragrafo vediamo dei problemi di stima per campioni gaussiani, che costituiscono una prima applicazione del teorema di Cochran.

Definiamo una nuova legge di probabilità. Si chiama *t di Student* con n gradi di libertà la legge della v.a.

$$Z = \frac{X}{\sqrt{Y}}\sqrt{n}$$

dove le v.a. X e Y sono indipendenti e di leggi $N(0,1)$ e $\chi^2(n)$ rispettivamente. Questa legge si indica con il simbolo $t(n)$.

Una proprietà importante delle leggi di Student è il fatto che sono simmetriche, cioè Z e $-Z$ hanno la stessa legge. Questo segue dal fatto che nella definizione le v.a. X, Y e $-X, Y$ hanno la stessa legge congiunta.

Non è difficile calcolare la densità di una legge $t(n)$, ma tralasceremo questo calcolo. Come vedremo tra poco la sola cosa realmente importante da conoscere delle leggi di Student è la funzione di ripartizione. Per questi ci sono delle tavole.

Abbiamo visto alla fine del paragrafo precedente che le v.a. $\bar{X}$ e $\sum_{i=1}^{m}(X_i - \bar{X})^2$ sono indipendenti e che $\sum_{i=1}^{m}(X_i - \bar{X})^2 \sim \chi^2(m-1)$. Poiché $\bar{X}$ ha legge $N(0, \frac{1}{m})$, $\sqrt{m}\bar{X}$ è $N(0,1)$ e dunque

$$T = \frac{\sqrt{m}\bar{X}}{\sqrt{\frac{1}{m-1}\sum_{i=1}^{m}(X_i - \bar{X})^2}} \sim t(m-1)\ . \tag{2.35}$$

Corollario 2.33 *Siano $Z_1, \dots, Z_m$ v.a. indipendenti e tutte di legge $N(z, \sigma^2)$. Poniamo*

$$\bar{Z} = \frac{1}{m}(Z_1 + \ldots + Z_m)$$

$$S^2 = \frac{1}{m-1}\sum_{i=1}^{m}(Z_i - \bar{Z})^2\ .$$

Allora le v.a. $\bar{Z}$ e S^2 sono indipendenti. Inoltre

$$\begin{aligned} &\frac{m-1}{\sigma^2}S^2 \sim \chi^2(m-1) \\ &\frac{\sqrt{m}(\bar{Z}-z)}{S} \sim t(m-1)\ . \end{aligned} \tag{2.36}$$

Dimostrazione. Si tratta semplicemente di ricondursi al caso di v.a. $N(0, I)$ che abbiamo già visto. Posto $X_i = \dfrac{Z_i - z}{\sigma}$, allora $X = (X_1, \ldots, X_m)$ è $N(0, I)$ e sappiamo già che $\bar{X}$ e $\sum_i (X_i - \bar{X})^2$ sono indipendenti. Tenendo conto che

$$\begin{gathered} \bar{Z} = \sigma \bar{X} + z \\ \sum_{i=1}^m (X_i - \bar{X})^2 = \frac{1}{\sigma^2} \sum_{i=1}^m (Z_i - \bar{Z})^2 = \frac{m-1}{\sigma^2} S^2 \; . \end{gathered} \tag{2.37}$$

e dato che $\bar{X}$ e $\sum_{i=1}^m (X_i - \bar{X})^2$ sono indipendenti, anche $\bar{Z}$ e S^2 sono indipendenti come funzioni di variabili indipendenti. Infine $\frac{m-1}{\sigma^2} S^2 \sim \chi^2(m-1)$ per (2.37) e (2.34), mentre poiché

$$\frac{\sqrt{m}(\bar{Z} - z)}{S} = \frac{\sqrt{m}\bar{X}}{\sqrt{\frac{1}{m-1} \sum_{i=1}^m (X_i - \bar{X})^2}}$$

(2.36) segue da (2.35).

Richiamo 2.34 Si chiama *quantile di ordine* α, $0 < \alpha < 1$, di una v.a. X l'estremo inferiore q_α dei numeri x tali che $F_X(x) = \mathrm{P}(X \le x) \ge \alpha$, ovvero

$$q_\alpha = \inf\{t, F_X(t) \ge \alpha\}$$

(in realtà è un minimo poiché F_X è continua a destra). Se X è una v.a. continua, allora F_X è continua e per il teorema dei valori intermedi l'equazione

$$F_X(x) = \alpha$$

ha sicuramente soluzione per ogni $0 < \alpha < 1$. Se per di più F_X è strettamente crescente (il che succede ad esempio se X ha densità strettamente positiva) allora la soluzione è unica. In questo caso q_α è dunque l'unico numero reale x tale che

$$\mathrm{P}(X \le x) = \alpha$$

Se per di più la v.a. X è simmetrica (cioè X e $-X$ hanno la stessa legge), come accade per le leggi $N(0, 1)$ e per le t di Student, allora si hanno le relazioni

$$1 - \alpha = \mathrm{P}(X \ge q_\alpha) = \mathrm{P}(-X \ge q_\alpha) = \mathrm{P}(X \le q_\alpha),$$

da cui si ricava che $q_{1-\alpha} = -q_\alpha$ e

$$\begin{gathered} \mathrm{P}(|X| \le q_{1-\alpha/2}) = \mathrm{P}(-q_{1-\alpha/2} \le X \le q_{1-\alpha/2}) = \\ = \mathrm{P}(X \le q_{1-\alpha/2}) - \mathrm{P}(X \le -q_{1-\alpha/2}) = 1 - \frac{\alpha}{2} - \frac{\alpha}{2} = 1 - \alpha \end{gathered} \tag{2.38}$$

Esempio 2.35 (Un po' di statistica ...) Siano $X_1, \dots, X_n$ v.a. indipendenti di legge $N(b, \sigma^2)$, dove però b e σ^2 sono sconosciute. È possibile, conoscendo i valori assunti da $X_1, \dots, X_n$ stimare i due parametri incogniti?

In effetti se poniamo

$$\bar{X} = \frac{1}{m}(X_1 + \ldots + X_m)$$
$$S^2 = \frac{1}{m-1}\sum_{i=1}^{m}(X_i - \bar{X})^2$$

allora sappiamo che

$$\frac{m-1}{\sigma^2}S^2 \sim \chi^2(m-1)$$
$$T = \frac{\sqrt{m}(\bar{X} - b)}{S} \sim t(m-1)\ .$$

Se indichiamo con $t_\alpha(n-1)$ il quantile di ordine α della legge $t(m-1)$, abbiamo

$$\mathrm{P}(|T| > t_{1-\alpha/2}(n-1)) = 1 - \alpha$$

(si usa il fatto che le leggi di Student sono simmetriche). D'altra parte $\{|T| > \eta\} = \{|\bar{X} - b| > t_{1-\alpha/2}(n-1)\frac{S}{\sqrt{m}}\}$. Quindi la probabilità che la media $\bar{X}$ dei valori osservati differisca dalla media b per una quantità superiore a $t_{1-\alpha/2}(n-1)\frac{S}{\sqrt{m}}$ è $\leq \alpha$. Ovvero, in altre parole la media b si trova nell'intervallo $I = [\bar{X} - t_{1-\alpha/2}(n-1)\frac{S}{\sqrt{m}}, \bar{X} + t_{1-\alpha/2}(n-1)\frac{S}{\sqrt{m}}]$ con probabilità $1-\alpha$. Si dice che I è *un intervallo di fiducia* per b *di livello* α.

La stessa idea permette di stimare la varianza σ^2, anche se con qualche cambiamento perché le leggi χ^2 non sono simmetriche come quelle di Student; se indichiamo con $\chi^2_\alpha(n-1)$ il quantile di ordine α di una legge $\chi^2(n-1)$,

$$\mathrm{P}(Z < \chi^2_{\alpha/2}(n-1)) = \frac{\alpha}{2}, \qquad\qquad \mathrm{P}(Z > \chi^2_{1-\alpha/2}(n-1)) = \frac{\alpha}{2}\ \cdot$$

Abbiamo dunque

$$\begin{aligned} 1 - \alpha &= \mathrm{P}\Big(\chi^2_{\alpha/2}(n-1) \leq \frac{m-1}{\sigma^2}S^2 \leq \chi^2_{1-\alpha/2}(n-1)\Big) = \\ &= \mathrm{P}\Big(\frac{m-1}{\chi^2_{1-\alpha/2}(n-1)}S^2 \leq \sigma^2 \leq \frac{m-1}{\chi^2_{\alpha/2}(n-1)}S^2\Big)\ . \end{aligned}$$

In altre parole $[\frac{m-1}{\chi^2_{1-\alpha/2}(n-1)}S^2, \frac{m-1}{\chi^2_{\alpha/2}(n-1)}S^2]$ è un intervallo di fiducia per σ^2 di livello α.

2.7 Leggi condizionali

Siano Y, X v.a. a valori negli spazi misurabili $(G, \mathscr{G})$, $(E, \mathscr{E})$ rispettivamente e indichiamo con ν_Y la legge di Y. Una famiglia di probabilità $(n(t, dx))_{t\in G}$ su $(E, \mathscr{E})$ si chiama una *legge condizionale di X dato Y* se,

i) Per ogni $A \in \mathscr{E}$, l'applicazione $t \to n(t, A)$ è $\mathscr{G}$-misurabile
ii) Per ogni $A \in \mathscr{E}$ e $B \in \mathscr{G}$,

$$\mathrm{P}(X \in A, Y \in B) = \int_B n(y, A)\, \nu_Y(dy).$$

Se l'insieme G è numerabile e $\mathrm{P}(Y = y) >$ per ogni $y \in G$, allora si si sceglie $B = y$, si trova

$$\mathrm{P}(X \in A, Y = y) = n(y, A)\mathrm{P}(Y = y)$$

ovvero

$$n(y, A) = \frac{\mathrm{P}(X \in A, Y = y)}{\mathrm{P}(Y = y)} = \mathrm{P}(X \in A \mid Y = y)$$

Intuitivamente, dunque, la legge condizionale è la legge che conviene attribuire alla v.a. X quando si disponga dell'informazione che $Y = y$.

La solita applicazione della Proposizione 1.2 implica che, se $g : G \to \mathbb{R}$ è misurabile limitata e $f : E \to \mathbb{R}$ è tale che la v.a. $f(X)$ sia integrabile (oppure se f e g sono positive),

$$\mathrm{E}[f(X)g(Y)] = \int_G \Big(\int_E f(x)\, n(y, dx)\Big) g(y)\, \nu_Y(dy)\ .$$

Poniamo

$$h(y) = \int_E f(x)\, n(y, dx) \tag{2.39}$$

allora la v.a. $h(Y)$ gode di alcune importanti proprietà. Intanto, per ogni funzione g misurabile limitata si ha

$$\mathrm{E}[f(X)g(Y)] = \int_G h(y)g(y)\, \nu_Y(dy) = \mathrm{E}[h(Y)g(Y)]\ . \tag{2.40}$$

Scegliendo $g = 1$, si trova che

$$\mathrm{E}[f(X)] = \mathrm{E}[h(Y)]$$

cioè le v.a. $f(X)$ e $g(Y)$ hanno la stessa speranza matematica.

Le applicazioni più interessanti si hanno quando la v.a. $f(X)$ è di quadrato integrabile allora si hanno delle proprietà interessanti. Intanto, in questo caso, anche la v.a. $h(Y)$ è in L^2. infatti, per la disuguaglianza di Jensen,

$$\begin{aligned}\mathrm{E}[h(Y)^2] &= \int_E \Big(\int_E f(x)\, n(y, dx)\Big)^2 \nu_Y(dy) \le\\ &\le \int_E \Big(\int_E f(x)^2\, n(y, dx)\Big) \nu_Y(dy) = \mathrm{E}[f(X)^2]\end{aligned}$$

Poi, usando il fatto che le v.a. limitate sono dense in L^2 (conseguenza della Proposizione 1.17), si vede facilmente che la relazione (2.40) è vera per ogni funzione g tale che $g(Y) \in L^2$. Si ha allora

$$\begin{aligned}
&\mathrm{E}[(f(X) - g(Y))^2] = \mathrm{E}[(\{f(X) - h(Y)\} + \{h(Y) - g(Y)\})^2] = \\
&= \mathrm{E}[(f(X) - h(Y))^2] + 2\underbrace{\mathrm{E}[(f(X) - h(Y))(h(Y) - g(Y))]}_{=0} + \mathrm{E}[(h(Y) - g(Y))^2] = \\
&= \mathrm{E}[(f(X) - h(Y))^2] + \mathrm{E}[(h(Y) - g(Y))^2] \geq \mathrm{E}[(f(X) - h(Y))^2]
\end{aligned}$$

Ovvero, in altre parole, h è la funzione di Y che meglio approssima $f(X)$ in L^2. La quantità indicata nella formula precedente si annulla perché

$$\mathrm{E}[(f(X)-h(Y))(h(Y)-g(Y))] = \mathrm{E}[f(X)(h(Y)-g(Y))]-\mathrm{E}[h(Y)(h(Y)-g(Y))] = 0$$

grazie alla (2.40), scritta con $h - g$ al posto di g.

Per vedere meglio il significato delle considerazioni precedenti, si può pensare che la v.a. $f(X)$ rappresenti un segnale che non può essere osservato direttamente. L'osservatore ha cioè solo accesso ad una osservazione Y. Il problema è di trovare la migliore stima di $f(X)$ che sia funzione dell'osservazione Y. Il calcolo di poco fa indica che, misurando la bontà della stima mediante la norma L^2 della differenza, la migliore approssimazione è data appunto da $h(Y)$, dove h è data dalla (2.39).

Osservazione 2.36 È utile confrontare il risultato ottenuto con l'Esempio 2.14 (la retta di regressione). In quell'esempio abbiamo calcolato la migliore approssimazione di una v.a. *X mediante una applicazione lineare-affine* di Y. Ora invece (scegliendo $f(x) = x$) abbiamo determinato la migliore approssimazione di X mediante una funzione (qualunque, purché boreliana) di Y.

Non è detto che una legge condizionale esista. I prossimi due esempi però mostrano delle situazioni in cui il calcolo della legge condizionale è facile (il che dimostra anche l'esistenza).

Esempio 2.37 Siano X, Y v.a. a valori in $\mathbb{R}^d$ e $\mathbb{R}^m$ rispettivamente, di densità congiunta $h(x, y)$ rispetto alla misura di Lebesgue di $\mathbb{R}^d \times \mathbb{R}^m$. Sia

$$h_Y(y) = \int_{\mathbb{R}^d} h(x, y)\, dx$$

la densità di Y e poniamo $Q = \{y; h_Y(y) = 0\}$. Evidentemente $\mathrm{P}(Y \in Q) = 0$. Se poniamo

$$\bar{h}(x; y) = \begin{cases} \dfrac{h(x, y)}{h_Y(y)} & \text{se } y \notin Q \\ \text{una densità arbitraria} & \text{se } y \in Q, \end{cases} \tag{2.41}$$

si vede subito che $n(y,dx) = \bar{h}(x;y)\,dx$ è una legge condizionale di X dato $Y = y$. Infatti, se f e g sono funzioni misurabili limitate su $\mathbb{R}^d$ e $\mathbb{R}^m$ rispettivamente,

$$
\begin{aligned}
\mathrm{E}[f(X)g(Y)] &= \int_{\mathbb{R}^m}\int_{\mathbb{R}^d} f(x)g(y)h(x,y)\,dy\,dx = \\
&= \int_{\mathbb{R}^m} g(y)h_Y(y)\,dy \int_{\mathbb{R}^d} f(x)\bar{h}(x;y)\,dx\,.
\end{aligned}
$$

Calcoliamo ora le leggi condizionali di una v.a. gaussiana multivariata. Ciò è naturalmente possibile usando l'Esempio 2.37, che però porta a dei calcoli non immediati, anche se elementari.

Useremo invece un argomento tipico delle leggi normali e che risulta utile anche in altre situazioni; esso si basa sul fatto che v.a. normali non correlate sono anche indipendenti. Cominciamo dal caso di due v.a. X e Y di legge congiunta normale e cerchiamo un numero a in modo che $X - aY$ e Y siano non correlate. deve cioè essere

$$
0 = \mathrm{Cov}(X - aY, Y) = \mathrm{Cov}(X,Y) - a\,\mathrm{Var}(Y)
$$

ovvero

$$
a = \frac{\mathrm{Cov}(Y,X)}{\mathrm{Var}(Y)}\,\cdot \tag{2.42}
$$

Dunque $Z = X - aY$ e Y sono indipendenti; vediamo ora che la legge condizionale di X dato $Y = y$ è appunto la legge della v.a. $Z + ay$, che indicheremo γ_y (il che è abbastanza intuitivo, dato il significato della legge condizionale. Intanto osserviamo che, se $g : \mathbb{R}^2 \to \mathbb{R}$ è misurabile limitata,

$$
\int g(z+ay,y)\,d\mu_Z(z) = \mathrm{E}[g(Z+ay,y)] = \int g(x,y)\,d\gamma_y(x)
$$

e dunque

$$
\begin{aligned}
\mathrm{E}[g(X,Y)] = \mathrm{E}[g(Z+aY,Y)] &= \int \mu_Y(dy)\int g(z+ay,y)\,d\mu_Z(z) = \\
&= \int \mu_Y(dy)\int g(x,y)\,d\gamma_y(z)
\end{aligned}
$$

che implica appunto che γ_y è la legge condizionale cercata. È importante osservare che questa legge condizionale è ancora gaussiana. $Z + ay = X - aY + ay$ è infatti normale (come funzione lineare-affine di X e Y) di varianza

$$
\mathrm{Var}(X - aY) = \mathrm{Var}(X) + a^2\,\mathrm{Var}(Y) - 2a\,\mathrm{Cov}(X,Y) = \mathrm{Var}(X) - \frac{\mathrm{Cov}(X,Y)^2}{\mathrm{Var}(Y)}
$$

e media

$$\mathrm{E}[Z + ay] = \mathrm{E}[X] + \frac{\mathrm{Cov}(X, Y)}{\mathrm{Var}(Y)}(y - \mathrm{E}[Y]) \cdot$$

Riprendendo l'Osservazione 2.36, vediamo quindi che, se le v.a. X e Y hanno legge congiunta normale, la migliore approssimazione di X mediante una funzione di Y coincide con la migliore approssimazione di X mediante una funzione lineare-affine di Y.

Queste argomentazioni si possono ripetere per delle v.a. X e Y di legge congiunta normale, ma rispettivamente a valori in $\mathbb{R}^{n_1}$ e $\mathbb{R}^{n_2}$. L'idea è di trovare una matrice A, $n_1 \times n_2$ e tale che ognuna delle componenti di Y sia non correlata rispetto a quelle di $X - AY$. Se $A = (a_{ij})_{ij}$ allora deve essere

$$0 = \mathrm{Cov}\Big(X_k - \sum_{i=1}^{n_1} a_{ki} Y_i, Y_h\Big) = \mathrm{Cov}(X_k, Y_h) - \sum_{i=1}^{n_1} a_{ki} \,\mathrm{Cov}(Y_i, Y_h) \ .$$

Se indichiamo con C_X la matrice di covarianza di X e $C_{X,Y}$ la matrice $n_1 \times n_2$ che ha come elementi $\mathrm{Cov}(X_k, Y_h)$, allora la relazione precedente diviene

$$C_{X,Y} - AC_Y = 0 \ .$$

Dunque, se supponiamo C_X invertibile,

$$A = C_{X,Y} C_Y^{-1} \ .$$

Ripetendo i ragionamenti sviluppati per il caso unidimensionale si ricava facilmente che la legge condizionale di X dato $Y = y$ è la legge di $X - AY + Ay$, cioè normale di media

$$\mathrm{E}(X) + C_{X,Y} C_Y^{-1}(y - \mathrm{E}(Y)) \ .$$

e matrice di covarianza

$$C_X - C_{X,Y} C_Y^{-1} C_{X,Y}^* \ .$$

Può essere utile segnalare che la matrice $C_{X,Y}$ non è altro che la matrice dei termini incrociati nella matrice di covarianza C di (X, Y), come chiarisce la formula seguente

$$C = \begin{pmatrix} \boxed{C_X} & \boxed{C_{X,Y}} \\ & \boxed{C_Y} \end{pmatrix} \ .$$

Esercizi

E2.1 Una v.a. reale X si dice *simmetrica* se X e $-X$ hanno la stessa legge. Dimostrare che una v.a. X di densità f è simmetrica se e solo se f è una funzione pari. (o, per essere precisi, se e solo se le funzioni $x \to f(x)$ e $x \to f(-x)$ sono equivalenti). Mostrare che se X è simmetrica per ogni $x \in \mathbb{R}$ $F(x) = 1 - F(-x)$

E2.2 a) Siano X una v.a. a valori positivi e $f : \mathbb{R}^+ \to \mathbb{R}$ una funzione derivabile con derivata continua e tale che $f(X)$ sia integrabile. Allora

$$\mathrm{E}[f(X)] = f(0) + \int_0^{+\infty} f'(t)\mathrm{P}(X \geq t)\,dt\,.$$

b) Sia X una v.a. a valori interi ≥ 0, allora

$$\mathrm{E}(X) = \sum_{k=1}^{\infty} \mathrm{P}(X \geq k)$$

[a) Se μ è la legge di X, allora $\int_0^{+\infty} f'(t)\,dt \int_t^{+\infty} \mu(dx) = \int_0^{+\infty} \mu(dx) \int_0^x f'(t)\,dt$ per il Teorema di Fubini.]

E2.3 Siano X e Y le coordinate di un punto scelto a caso con distribuzione uniforme sul quadrato di vertici $(1, 0)$, $(0, 1)$, $(-1, 0)$, $(0, -1)$.

a) Calcolare le leggi di X e Y. Ammettono una densità? Si tratta di v.a. indipendenti? Qual è la legge condizionale di Y dato X?

b) Calcolare

$$\mathrm{P}\Big(\frac{1}{\sqrt{3}} < \frac{X}{Y} < \sqrt{3}\Big)\,.$$

E2.4 Siano X e Y v.a. indipendenti uniformemente distribuite su $[0, 1]$ e poniamo $Z = X + Y$.

a) Qual è la legge di (X, Z)? È una legge uniforme?

b) Calcolare $\mathrm{P}(Z \geq 1 | X \leq \frac{1}{2})$. Qual è la legge condizionale di Z dato X?

c) Calcolare la legge di Z. È una legge uniforme?

E2.5 Una v.a. X ha densità data da

$$f(x) = \begin{cases} \frac{1}{x^2} & \text{se } x > 1 \\ 0 & \text{altrimenti}\,. \end{cases}$$

a) Qual è la densità di $W = \log X$? Quale la sua media?

b) Sia Y un'altra v.a. di densità f e indipendente da X. Calcolare la densità di $Z = \sqrt{X\,Y}$.

E2.6 Nell'intervallo $[0, R]$ vengono scelti, in maniera indipendente l'uno dall'altro, due punti X e Y con distribuzione uniforme. Indichiamo con U il punto più vicino a 0 e con V quello più vicino a R.

a) Calcolare le leggi di U, V e $V - U$.

b) Qual è la probabilità che con i tre segmenti OU, UV e VR si possa costruire un triangolo?

E2.7 a) Sia X una v.a. $N(0, 1)$.

a1) Quanto vale $\mathrm{E}(\mathrm{e}^{tX^2})$?

a2) Qual è la legge di X^2?

b) Sia W una v.a. $N(0, I)$ su $\mathbb{R}^m$ (normale multivariata di media 0 e di matrice di covarianza uguale alla matrice identità). Sia A una matrice $m \times m$ simmetrica e consideriamo la v.a. $Z = \frac{1}{2}\langle AW, W\rangle$ ($\langle\ ,\ \rangle$ è il prodotto scalare di $\mathbb{R}^m$). Indichiamo $\lambda_1, \dots, \lambda_m$ gli autovalori di A (eventualmente con ripetizione).

b1) Quali ipotesi occorre fare su $\lambda_1, \dots, \lambda_m$ perché la v.a. e^{tZ} sia integrabile? Quanto vale $\mathrm{E}(\mathrm{e}^{tZ})$?

b2) Come si deve modificare il risultato precedente se W (sempre centrata) avesse invece matrice di covarianza Γ, non necessariamente uguale alla matrice identità? E se la matrice A non fosse simmetrica?

E2.8 Siano $X_1, \dots, X_n$ della v.a. i.i.d. avente una legge μ che ammette densità rispetto alla misura di Lebesgue. Indichiamo con $Y_1, \dots, Y_n$ i rispettivi *ranghi*. Cioè $Y_i = 1$ se il valore di X_i è il più piccolo tra $X_1, \dots, X_n$, $Y_i = n$ se è il più grande, $Y_i = k$ se ci sono esattamente $k - 1$ indici j per i quali $X_j < X_i$ (e $n - k + 1$ per i quali $X_j > X_i$).

a) Mostrare che l'evento $A = \{X_i = X_j$ per qualche coppia di indici$i \neq j\}$ ha probabilità 0 e dunque i ranghi sono ben definiti.

b) Mostrare che il vettore $(Y_1, \dots, Y_n)$, a valori nel gruppo delle permutazioni su n elementi, ha legge uniforme.

E2.9 Date due misure μ e ν su $\mathbb{R}$ indichiamo con F_μ, F_ν le rispettive funzioni di ripartizione. Si dice che $\mu \preceq \nu$ (*μ è stocasticamente più piccola di ν*) se e solo se $F_\mu(x) \geq F_\nu(x)$ per ogni $x \in \mathbb{R}$.

a) Mostrare che $\mu \preceq \nu$ se e solo se, per ogni $a \in \mathbb{R}$,

$$\mu(]a, +\infty[) \leq \nu(]a, +\infty[) \ .$$

b) Siano μ, ν leggi di Poisson di parametri α e β rispettivamente con $\alpha \leq \beta$. Mostrare che $\mu \preceq \nu$.

c) Supponiamo che esista una probabilità γ su $\mathbb{R}^2$ tale che

i) γ ha μ e ν come prima e seconda marginale rispettivamente

ii) $\gamma(\Gamma) = 1$, dove $\Gamma = \{(x, y), x \leq y\}$.

Mostrare che allora $\mu \preceq \nu$.

d) Supponiamo $\mu \preceq \nu$ e supponiamo che F_μ e F_ν siano funzioni strettamente crescenti. Sia U una v.a. di legge uniforme su $[0, 1]$ e sia Z la v.a. a valori $\mathbb{R}^2$ definita

da
$$Z = (F_\mu^{-1}(U), F_\nu^{-1}(U)) \ .$$
Mostrare allora che la legge γ di Z soddisfa alle condizioni i) e ii) di c). Cosa si può dire se si toglie l'ipotesi che F_μ e F_ν siano strettamente crescenti?

e) Mostrare che $\mu \preceq \nu$ se e solo se esistono uno spazio di probabilità $(\Omega, \mathcal{F}, \mathrm{P})$ sul quale sono definite due v.a. reali X e Y aventi come legge μ e ν rispettivamente e tali che $\mathrm{P}(X \le Y) = 1$.

E2.11 Le v.a. $X_1, \dots, X_n$ si dicono *scambiabili* se e solo se la legge di $(X_1, \dots, X_n)$ è uguale alla legge di $(X_{\sigma_1}, \dots, X_{\sigma_n})$ dove $(\sigma_1, \dots, \sigma_n)$ è una qualunque permutazione di $(1, \dots, n)$.

a) Mostrare che se $X_1, \dots, X_n$ sono scambiabili allora esse hanno la stessa legge; anzi che la legge di (X_i, X_j) non dipende da $i, j,\ i \ne j$.

b) Mostrare che se $X_1, \dots, X_n$ sono indipendenti equidistribuite allora esse sono scambiabili.

Consideriamo ora un'urna contenente n palline di cui r rosse e b bianche e indichiamo con $X_1, \dots, X_n$ il risultato di n estrazioni senza rimpiazzo ($X_i = 1$ se la pallina i-esima è rossa, $X_i = 0$ se è bianca).

c) Mostrare che $X_1, \dots, X_n$ sono scambiabili.

d) Quanto vale $\mathrm{Cov}(X_1, X_2)$? Quanto vale $\mathrm{Cov}(X_i, X_j)$?

e) Sia X il numero di palline estratte in k estrazioni. Quanto vale $\mathrm{Var}(X)$?

f) Siano X, Y come nell'Esercizio 3.2 c). Mostrare che X e Y sono scambiabili.

E2.12 Siano X e Y v.a indipendenti di legge data rispettivamente dalle densità
$$f_X(x) = x\mathrm{e}^{-x^2/2} 1_{[0,+\infty[}(x) \qquad\qquad f_Y(y) = \frac{1}{\pi\sqrt{1-y^2}} 1_{]-1,1[}(y) \ .$$

a) Calcolare $\mathrm{E}(X)$, $\mathrm{E}(Y)$ e $\mathrm{Var}(X)$.

b) Posto $U = XY$, $V = X\sqrt{1-Y^2}$, calcolare le leggi di U e di V. Qual è la legge di (U, V)? Le v.a. U e V sono indipendenti?

E2.13 Siano μ una probabilità su $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, f una densità di probabilità su $\mathbb{R}$ e $\nu = f\,dx$. Mostrare che $\mu * \nu$ ha densità rispetto alla misura di Lebesgue e calcolarla.

E2.14 Consideriamo tre v.a. X, Y e Z dove: X ha legge uniforme su $[0, 1]$; Y ha densità condizionale se $X = x$ data da
$$g_x(y) = (y - x)\mathrm{e}^{-(y-x)} 1_{\{0 \le x \le 1, x \le y\}}$$
mentre Z ha legge condizionale per $X = x$, $Y = y$ di densità
$$h_{x,y}(z) = (y - x)\mathrm{e}^{-z(y-x)} 1_{\{0 \le x \le 1, x \le y, 0 < z\}} \ .$$

Qual è la legge di (X, Y, Z)? Determinare le leggi di Y e di Z. Se $U = Y - X$, $V = Z(Y - X)$ qual è la legge di (X, U, V)?

E2.15 Mostrare che se $\rho_{X,Y} = 1$ oppure $\rho_{X,Y} = -1$ allora esiste $a \in \mathbb{R}$ tale che $X = aY$ oppure $aX = Y$. Inoltre a è ≥ 0 oppure ≤ 0 a seconda che sia $\rho_{X,Y} = 1$ oppure $\rho_{X,Y} = -1$.

[Si usa il fatto che nella disuguaglianza di Schwartz si ha uguaglianza se e solo se i vettori sono collineari.]

E2.16 Siano X e Y v.a. indipendenti e supponiamo che $X+Y$ abbia speranza matematica finita. Allora lo stesso è vero anche per X e Y (ovvero, se μ e ν hanno media finita, lo stesso è vero per $\mu * \nu$).

E2.17 Siano A e B due matrici simmetriche $m \times m$ semidefinite positive. Consideriamo la matrice C i cui elementi si ottengono moltiplicando termine a termine quelli di A e di B; cioè $c_{ij} = a_{ij} b_{ij}$. Mostrare che C è anch'essa semidefinita positiva. (Chissà come c'entra la probabilità …)

[Siano X, Y v.a. n-dimensionali indipendenti di matrici di covarianza A e B rispettivamente (esistono certamente: basta prenderle gaussiane, ad esempio), e definiamo una v.a. Z n-dimensionale con $Z_i = X_i Y_i$, $i = 1, \dots, n$; allora C è la matrice di covarianza di Z ed è semi-definita positiva come tutte le matrici di covarianza.]

E2.18 Sia X una v.a. reale di densità

$$h_\theta(x) = \begin{cases} \frac{2x}{\theta}\, \mathrm{e}^{-x^2/\theta} & \text{se } x > 0 \\ 0 & \text{altrimenti} \end{cases}$$

dove θ è un parametro reale > 0 (legge di Rayleigh).

a) Calcolare media e varianza di X.

b) Se $Z = X^2$, mostrare che Z segue una legge Γ. Calcolarne i parametri.

c) Posto $W = \mathrm{e}^{-X^2/\theta}$, calcolare media e varianza di W. Qual è la densità di W?

c) Sia (U, V) una coppia di v.a. di densità $f(u, v) = \alpha h_\theta(u) v 1_{\{0<v<u\}}$ dove α è una opportuna costante. Determinare il valore di c al variare di θ. Le v.a. U e U/V sono indipendenti?

E2.19 a) Siano X e Y v.a. aleatorie indipendenti entrambe di legge esponenziale di parametro $\lambda > 0$. Calcolare le leggi delle v.a. $U = X - Y$, $V = \min(X, Y)$ e $W = \max(X, Y)$. Mostrare che U e V sono indipendenti.

b) Siano X e Y due v.a. positive indipendenti ed aventi entrambe una legge data da una densità f, dove f è una funzione strettamente positiva su $\mathbb{R}^+$. Calcolare la legge della coppia (U, V), dove $U = X - Y$ e $V = \min(X, Y)$. Mostrare che se U e V sono indipendenti allora X e Y sono esponenziali di parametro λ per qualche $\lambda > 0$.

[b): se h è una funzione boreliana positiva su $\mathbb{R}^2$, allora (integrazione rispetto a una legge immagine)

$$\mathrm{E}(h(U,V)) = \int\int h(x-y,x)f(x)f(y)1_{\{0<x<y\}}\,dx\,dy +$$
$$+\int\int h(x-y,x)f(x)f(y)1_{\{0<y<x\}}\,dx\,dy\ .$$

Nel primo integrale si fa il cambio di variabile $u = x - y$, $v = x$, nel secondo $u = x - y$, $v = y$. Mettendo insieme i pezzi si ottiene

$$\mathrm{E}(h(U,V)) = \int\int h(u,v)f(v)f(v+|u|)1_{\{v>0\}}\,du\,dv\ .$$

Dunque (U,V) ha densità $f(v)f(v+|u|)1_{\{v>0\}}$. Le marginali di U e di V sono date da

$$f_U(u) = \int_0^{+\infty} f(v)f(v+|u|)\,dv$$
$$f_V(v) = 1_{\{v>0\}}f(v)\int f(v+|u|)\,du = 2f(v)(1-F(v))$$

dove F è la funzione di ripartizione della legge di densità f. Se U e V sono indipendenti deve essere

$$f(v)f(v+|u|) = 2f(v)(1-F(v))f_U(u)$$

per ogni $u\in\mathbb{R}$, $v>0$.Ponendo $u=0$ si vede che F soddisfa all'equazione differenziale

$$\frac{F'(v)}{1-F(v)} = 2f_U(0)$$

ovvero $1-F(v) = C\,\mathrm{e}^{-2f_U(0)v}$ per $v>0$. Per $v\to 0$ si ha $C=1\ldots$]

E2.20 Siano X, Y, Z v.a. indipendenti tutte di legge $N(0,1)$.

a) Calcolare le leggi delle v.a.

$$\frac{X^2}{X^2+Y^2} \qquad \frac{|X|}{\sqrt{X^2+Y^2}}\ .$$

b) Calcolare la legge condizionale di X sapendo X^2+Y^2.

c) Calcolare le leggi delle v.a.

$$\frac{X^2}{Z^2+Y^2} \qquad \frac{|X|}{\sqrt{Z^2+Y^2}}\ .$$

c) Mostrare che

$$\frac{X}{\sqrt{X^2+Y^2}} \quad \text{e} \quad X^2+Y^2$$

sono indipendenti.

[È utile ricordare che il quadrato di una v.a. $N(0,1)$ segue una legge $\Gamma(\frac12,\frac12)$.]

E2.21 Siano $X_1, \dots, X_{12}$ v.a. indipendenti di legge $N(0, 1)$.

a) Qual è la legge della v.a. $W = \max(X_1, \dots, X_{12})$? Quanto vale $\mathrm{P}(W > 2)$?

b) Sia $Z = \min(X_1, \dots, X_{12})$. Quanto vale il quantile di ordine .05 di Z?

c) Un generatore aleatorio $N(0, 1)$ ha prodotto la seguente sequenza di valori:

0.78	−0.45	0.93	0.27	−0.57	0.45
0.19	2.03	−0.31	3.74	−4.23	−0.76

Cosa ne pensate?

E2.22 a) Sia ν la legge su $\mathbb{R}$ di densità $h(t) = \frac{1}{2}\,\mathrm{e}^{-|t|}$ rispetto alla misura di Lebesgue. Mostrare che

$$\hat{\nu}(\theta) = \frac{1}{1+\theta^2}\ .$$

b) Sia μ la probabilità di densità

$$f(t) = \frac{1}{\pi(1+t^2)}$$

(legge di Cauchy). Mostrare che $\hat{\mu}(\theta) = \mathrm{e}^{-|\theta|}$. Quanto vale la media di μ? $\hat{\mu}$ è derivabile?

c) Siano X, Y v.a. di Cauchy indipendenti. Mostrare che $2X$ e $X + Y$ hanno la stessa legge.

d) Mostrare che la legge di Cauchy è la legge di Student $t(1)$.

[b): usare il punto a) e il Teorema d'inversione 2.23.]

E2.24 Sia μ una misura di probabilità fissata su $\mathbb{R}^d$.

a) Mostrare che per ogni $\varepsilon > 0$ esiste $R = R_\varepsilon > 0$ tale che $\mu(B_R^C) \le \varepsilon$, dove B_R indica la palla di centro 0 e raggio R.

b) Mostrare che per ogni $\theta_1, \theta_2 \in \mathbb{R}^d$ si ha

$$|\mathrm{e}^{i\langle\theta_1,x\rangle} - \mathrm{e}^{i\langle\theta_2,x\rangle}| \le |x||\theta_1 - \theta_2|\ .$$

In particolare le funzioni $\theta \to \mathrm{e}^{i\langle\theta,x\rangle}$ sono uniformemente continue al variare di $x \in B_R$.

c) Mostrare che $\theta \to \hat{\mu}(\theta)$ è uniformemente continua.

E2.25 Siano X, Y delle v.a. congiuntamente gaussiane, centrate. Supponiamo che $\mathrm{E}(X^2) = 4$, $\mathrm{E}(Y^2) = 1$ e che le v.a. $2X + Y$ e $X - 3Y$ sono indipendenti.

a) Calcolare la matrice di covarianza di (X, Y).

b) Calcolare la legge del vettore $(X + Y, 2X - Y)$.

E2.26 Sia $X = (X_1, X_2, X_3)$ un vettore gaussiano centrato di matrice di covarianza

$$C = \begin{pmatrix} 3 & -1 & 0 \\ -1 & 3 & 0 \\ 0 & 0 & 2 \end{pmatrix}$$

1) X ha densità rispetto alla misura di Lebesgue di $\mathbb{R}^3$? Se si calcolarla.
2) X_3 è indipendente da $X_1 + X_2$?
3) Determinare la legge di $X_1 + 2X_2 - X_3$.
4) Determinare un operatore lineare $A : \mathbb{R}^3 \to \mathbb{R}^3$ tale che le componenti del vettore AX siano v.a. indipendenti.

E2.27 a) Sia $W = (W_1, \dots, W_m)$ una v.a. m-dimensionale di legge $N(b, C)$. Mostrare che la v.a. $\xi_1 W_1 + \ldots + \xi_m W_m$ è gaussiana di media $\langle \xi, b \rangle$ e varianza $\langle C\xi, \xi \rangle$.

b) Sia (X, Y, Z) un vettore gaussiano 3-dimensionale centrato di matrice di covarianza

$$C = \begin{pmatrix} 1 & 0 & \sqrt{\frac{2}{3}} \\ 0 & 1 & \sqrt{\frac{1}{3}} \\ \sqrt{\frac{2}{3}} & \sqrt{\frac{1}{3}} & 1 \end{pmatrix}$$

b1) Per ogni $(\alpha, \beta, \gamma) \in \mathbb{R}^3$, calcolare la legge della v.a. $\alpha X + \beta Y + \gamma Z$.

b2) Risolvere l'equazione $C\xi = 0$, $\xi \in \mathbb{R}^3$. La legge di (X, Y, Z) ha densità rispetto alla misura di Lebesgue di $\mathbb{R}^3$?

b3) Esiste un vettore $(\alpha, \beta, \gamma) \in \mathbb{R}^3 \setminus (0, 0, 0)$ tale che $\mathrm{Var}(\alpha X + \beta Y + \gamma Z) = 0$?

E2.28 Una v.a. $X = (X_1, X_2)$ segue una legge normale bivariata $N(0, C)$ dove

$$C = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} .$$

a) Qual è la legge di $X_1 + X_2$?
b) Qual è la legge condizionale di X_2 dato X_1?
c) Mostrare che la v.a. X ha densità e calcolarla.

E2.29 Siano X, Y v.a. indipendenti $N(0, 1)$.
a) Qual è la legge della v.a. $(X, X + Y)$?
b) Calcolare la legge condizionale di X dato $X + y = t$.

E2.30 Sia X un segnale di legge normale $N(0, 1)$. Un osservatore non ha accesso al valore di X, di cui conosce solo un'osservazione $Y = X + W$, dove W è un rumore, indipendente da X e di legge $N(0, \sigma^2)$.

a1) Qual è la vostra stima del valore X del segnale sapendo che $Y = y$?

a2) Supponiamo $\sigma^2 = 0.1$ e che il valore dell'osservazione sia $Y = 0.55$. Qual è la probabilità che il segnale X si trovi nell'intervallo $[\frac{1}{4}, \frac{3}{4}]$?

b) Lo stesso osservatore, per migliorare la stima, decide di effettuare due osservazioni $Y_1 = X + W_1$ e $Y_2 = X + W_2$, dove W_1 e W_2 sono v.a. $N(0, \sigma^2)$ e le tre v.a. X, W_1, W_2 sono indipendenti. Qual è ora la stima di X dato $Y_1 = y_1$ e $Y_2 = y_2$? Di quanto è diminuita la varianza della legge condizionale di X dato $(Y_1, Y_2) = (y_1, y_2)$?

E2.31 a) Sia X una v.a. gaussiana m-dimensionale $N(0, I)$. Qual è la legge della v.a. $|X|^2$?

b) Sia X una v.a. gaussiana m-dimensionale $N(0, C)$.

b1) Mostrare che la v.a. $|X|^2$ ha la stessa legge che una v.a. della forma

$$\sum_{k=1}^{m} \lambda_k Z_k$$

dove $Z_1, \dots, Z_m$ sono v.a. indipendenti di legge $\chi^2(1)$ e $\lambda_1, \dots, \lambda_m$ sono gli autovalori di C.

b2) Mostrare che $\mathrm{E}(|X|^2) = \mathrm{tr}\, A$.

E2.32 Sia $X = (X_1, \dots, X_n)$ un vettore aleatorio gaussiano di legge $N(0, I)$. Per ogni $k \in \{1, \dots, n\}$ poniamo $Y_k = X_1 + \dots + X_k - kX_{k+1}$ (con la convenzione $X_{n+1} = 0$). Le v.a. $Y_1, \dots, Y_n$ sono indipendenti?

E2.33 (Filtraggio di un segnale) Consideriamo delle v.a. X (il segnale) e W (il rumore) indipendenti e con W centrata. Supponiamo X e W entrambe di quadrato integrabile. Poniamo $Y = X + W$ (l'osservazione).

a) Qual è la migliore predizione lineare-affine $\phi(X)$ di Y come funzione di X nel senso della distanza in L^2?

b) Confrontare la distanza in L^2 da Y delle seguenti quantità

$\phi(X)$ ($\phi(X)$ è la migliore predizione lineare-affine di cui al punto a))

X

$\mathrm{E}(Y)$

E2.34 Un rivelatore viene usato per determinare l'istante di emissione di un fotone. Si sa, a priori, che il tempo T in cui il fotone viene emesso segue una legge esponenziale di parametro λ. Il rivelatore però ha un tempo di reazione che è a sua volta aleatorio, anch'esso esponenziale di media $\frac{1}{\mu}$ (che supporremo molto più piccola della media $\frac{1}{\lambda}$ di T). Più precisamente l'istante S in cui il rivelatore viene attivato è uguale a $T + W$ dove W è esponenziale di media $\frac{1}{\mu}$ e indipendente da T.

a) Qual è la legge di S? Qual è la legge congiunta di S e T?

b1) Quanto vale la speranza condizionale di T dato $S = s$? Supponiamo $\lambda = 1$, $\mu = 10$; se il rivelatore segnala l'emissione del fotone all'istante $s = 1.5$, qual è la vostra stima dell'istante T in cui il fotone è stato emesso? E se fosse $s = 0.1$?

b2) Qual è la migliore stima lineare-affine di T conoscendo S? Confrontarla con quella ottenuta in b1), per i valori numerici assegnati.

E2.35 Un generatore aleatorio ha prodotto 256 numeri a caso. La media $\bar{x}$ di questi numeri vale -0.25, mentre

$$s^2 = \frac{1}{255} \sum_{i=1}^{256} (x_i - \bar{x})^2 = 3.01$$

Calcolare un intervallo di fiducia di livello 0.05 per la media e per la varianza della legge di questi numeri, supponendo che si tratti di una legge normale. È ragionevole pensare che si tratti di una $N(0, 1)$?

E2.36 Siano X e Y due v.a. indipendenti, dove $X \sim N(0,1)$ mentre Y è tale che $\mathrm{P}(Y=1)=\mathrm{P}(Y=-1)=\frac{1}{2}$. Poniamo $Z=XY$.

a) Qual è la legge di Z? Z e X sono indipendenti?

b) Calcolare la funzione di ripartizione F di $X+Z$. Mostrare che X e Z non hanno legge congiunta normale.

[a): si calcola la funzione di ripartizione

$$\begin{aligned} F_Z(z) &= \mathrm{P}(Z \le z) = \mathrm{P}(Z \le z, Y=1) + \mathrm{P}(Z \le z, Y=-1) = \\ &= \frac{1}{2}\mathrm{P}(X \le z) + \frac{1}{2}\mathrm{P}(X \ge -z) = \mathrm{P}(X \le z)\ . \end{aligned}$$

La stessa idea si usa per b).]

3
Convergenza e approssimazione

3.1 Il Lemma di Borel-Cantelli

Se $(A_n)_n$ è una successione di eventi di $\mathscr{A}$, consideriamo l'evento

$$A = \overline{\lim_{n\to\infty}}\, A_n := \bigcap_{n=1}^{\infty} \bigcup_{k\ge n} A_k$$

che chiameremo il *limite superiore* degli eventi $(A_n)_n$. Da uno sguardo più attento a questa definizione si vede che $\omega \in A$ se e solo se per ogni n

$$\omega \in \bigcup_{k\ge n} A_k$$

ovvero se e solo se $\omega \in A_k$ per infiniti indici k; quindi si ha

$$\overline{\lim_{n\to\infty}}\, A_n = \{\omega;\, \omega \in A_k \text{ per infiniti indici } k\}$$

La terminologia $A = \overline{\lim}_{n\to\infty} A_n$ deriva dal fatto che

$$1_A = \overline{\lim_{n\to\infty}}\, 1_{A_n}$$

In maniera analoga si definisce

$$B = \underline{\lim_{n\to\infty}}\, A_n = \bigcup_{n=1}^{\infty} \bigcap_{k\ge n} A_k$$

Si vede che $\omega \in B$ se e solo se esiste n_0 tale che $\omega \in A_k$ per ogni $k \geq n_0$. È chiaro che

$$\overline{\lim_{n\to\infty}} A_n \supset \underline{\lim_{n\to\infty}} A_n$$

e, per la formula di De Morgan,

$$\Big(\overline{\lim_{n\to\infty}} A_n\Big)^C = \underline{\lim_{n\to\infty}} A_n^C$$

Evidentemente i limiti superiore ed inferiore sono eventi che appartengono alla σ–algebra terminale

$$\mathcal{B}^\infty = \bigcap_{i=1}^{\infty} \sigma(1_{A_i}, 1_{A_{i+1}}, \ldots)$$

Dunque per il Teorema di Kolmogorov 2.12, se gli eventi $A_1, A_2, \ldots$ sono indipendenti allora $\underline{\lim}_{n\to\infty} A_n$ e $\overline{\lim}_{n\to\infty} A_n$ possono avere solo probabilità 0 oppure 1. Il teorema seguente fornisce un modo pratico di stabilire quale di queste due eventualità sia vera.

Teorema 3.1 (Lemma di Borel-Cantelli)
a) Se $\sum_{n=1}^{\infty} \mathrm{P}(A_n) < +\infty$ *allora* $\mathrm{P}(\overline{\lim}_{n\to\infty} A_n) = 0$.
b) Se gli eventi $(A_n)_n$ *sono indipendenti e la serie* $\sum_{n=1}^{\infty} \mathrm{P}(A_n)$ *è divergente allora* $\mathrm{P}(\overline{\lim}_{n\to\infty} A_n) = 1$.

Dimostrazione. a) Per il teorema di Beppo Levi

$$\sum_{n=1}^{\infty} \mathrm{P}(A_n) = \mathrm{E}\Big[\sum_{n=1}^{\infty} 1_{A_n}\Big]$$

ma $\overline{\lim}_{n\to\infty} A_n$ è esattamente l'evento su cui $\sum_{n=1}^{\infty} 1_{A_n} = +\infty$ (se $\omega \in \overline{\lim}_{n\to\infty} A_n$ allora $\omega \in A_n$ per infiniti indici e dunque nella serie figurano infiniti termini uguali a 1). Quindi se $\sum_{n=1}^{\infty} \mathrm{P}(A_n) < +\infty$, la v.a. $\sum_{n=1}^{\infty} 1_{A_n}$ è integrabile e $\overline{\lim}_{n\to\infty} A_n$ è trascurabile (l'insieme degli ω sui quali una funzione integrabile prende il valore $+\infty$ è sempre trascurabile.

b) Per definizione la successione di eventi

$$\Big(\bigcup_{k\geq n} A_k\Big)_n$$

decresce a $\overline{\lim}_{n\to\infty} A_n$. Dunque

$$\mathrm{P}\Big(\overline{\lim_{n\to\infty}} A_n\Big) = \lim_{n\to\infty} \mathrm{P}\Big(\bigcup_{k\geq n} A_k\Big)$$

Basta ora dimostrare che, per ogni n, $\mathrm{P}\Big(\bigcup_{k\ge n} A_k\Big) = 1$ oppure, che è lo stesso, che

$$\mathrm{P}\Big(\bigcap_{k\ge n} A_k^C\Big) = 0$$

Ma, usando la disuguaglianza $\mathrm{e}^{-x} \ge 1 - x$,

$$\mathrm{P}\Big(\bigcap_{k\ge n} A_k^C\Big) = \lim_{N\to\infty} \mathrm{P}\Big(\bigcap_{k=n}^{N} A_k^C\Big) = \lim_{N\to\infty} \prod_{k=n}^{N} \mathrm{P}(A_k^C) =$$

$$= \lim_{N\to\infty} \prod_{k=n}^{N} \Big(1 - \mathrm{P}(A_k)\Big) \le \lim_{N\to\infty} \prod_{k=n}^{N} \mathrm{e}^{-\mathrm{P}(A_k)} = \lim_{N\to\infty} \exp\Big[-\sum_{k=1}^{N} \mathrm{P}(A_k)\Big] = 0$$

Esempio 3.2 Sia $(X_n)_n$ una successione di v.a. indipendenti, tutte di legge esponenziale di parametro λ. Sia c un numero positivo. Vogliamo calcolare quanto vale la probabilità dell'evento

$$\varlimsup_{n\to\infty} \{X_n \ge c \log n\} \tag{3.1}$$

Il lemma di Borel-Cantelli permette di rispondere immediatamente a questa domanda: basta determinare la natura della serie

$$\sum_{n=1}^{\infty} \mathrm{P}(X_n \ge c \log n)$$

Poiche' conosciamo la f.r. delle leggi esponenziali, sappiamo che

$$\mathrm{P}(X_n \ge c \log n) = \mathrm{e}^{-\lambda c \log n} = \frac{1}{n^{\lambda c}}$$

che è il termine generale di una serie convergente se e solo se $c > \frac{1}{\lambda}$. Dunque il limite superiore (3.1) ha probabilità 0 se $c > \frac{1}{\lambda}$ e probabilità 1 se $c \le \frac{1}{\lambda}$. Da notare il fatto apparentemente paradossale: in quest'ultimo caso gli eventi $\{X_n \ge c \log n\}$ hanno probabilità che tende a zero, ma ciononostante con probabilità 1 ogni $\omega \in \Omega$ appartiene a questi eventi per infiniti indici n

3.2 La convergenza quasi certa

Siano $X, X_1, \dots, X_n, \dots$ v.a. definite su uno stesso spazio di probabilità $(\Omega, \mathscr{A}, \mathrm{P})$.

Definizione 3.3 *Siano* $X, X_1, \ldots, X_n, \ldots$ *v.a. definite su uno stesso spazio di probabilità* $(\Omega, \mathscr{A}, \mathrm{P})$.

a) Se le v.a. $X, X_n, n \geq 1$ *sono a valori in uno spazio metrico* (E, d), *si dice che la successione* $(X_n)_n$ *converge a* X *in probabilità (e si scrive* $\lim_{n\to\infty} X_n \overset{\mathrm{P}}{=} X$*) se per ogni* $\delta > 0$

$$\lim_{n\to\infty} \mathrm{P}(d(X_n, X) > \delta) = 0$$

b) Se le v.a. $X, X_n, n \geq 1$ *sono a valori in uno spazio topologico* E *si dice che* $(X_n)_n$ *converge a* X *quasi certamente (q.c.) se esiste un evento* N *trascurabile (cioè tale che* $\mathrm{P}(N) = 0$*) tale che per ogni* $\omega \in N^C$

$$\lim_{n\to\infty} X_n(\omega) = X(\omega)$$

c) Se le v.a. $X, X_n, n \geq 1$ *sono a valori in* $\mathbb{R}^m$ *si dice che* $(X_n)_n$ *converge a* X *in* L^p *se* $X, X_n \in L^p$ *per ogni* n *e*

$$\lim_{n\to\infty} \mathrm{E}(|X_n - X|^p) = \lim_{n\to\infty} \|X_n - X\|_p = 0$$

Il resto di questo paragrafo è dedicato al confronto tra i diversi modi di convergenza introdotti nella definizione precedente. Per semplicità supporremo che tutte le v.a. siano a valori in $\mathbb{R}^m$, ma i risultati che seguono si possono immediatamente estendere al caso di uno spazio metrico (E, d), sostituendo d alla distanza euclidea nei ragionamenti che seguono.

Intanto è immediato che la convergenza in L^p, $p > 0$, implica quella in probabilità: per la disuguaglianza di Markov,

$$\mathrm{P}(|X_n - X| > \delta) \leq \frac{1}{\delta^p} \mathrm{E}(|X_n - X|^p)$$

Vedremo presto con degli esempi che le convergenze in L^p e q.c. non sono confrontabili (anche se i risultati di convergenze q.c. vengono in genere considerati più forti).

Vediamo invece ora di confrontare la convergenza q.c. e quella in probabilità. Per $\delta \geq 0$ poniamo $A_\delta = \overline{\lim}_{n\to\infty}\{|X_n - X| > \delta\}$. Se $X_n \overset{q.c.}{\to} X$, deve essere $\mathrm{P}(A_\delta) = 0$ per ogni $\delta > 0$. Infatti se $\omega \in A_\delta$ allora $|X_n(\omega) - X(\omega)| > \delta$ per infiniti indici n, e quindi non può essere $\lim_{n\to\infty} X_n(\omega) = X(\omega)$.

Viceversa l'insieme degli ω per cui $X_n(\omega)$ non converge a $X(\omega)$ è dato da

$$\big\{\omega; \overline{\lim_{n\to\infty}} |X_n(\omega) - X(\omega)| > 0\big\} = \bigcup_{k=1}^\infty \{\omega; \overline{\lim_{n\to\infty}} |X_n(\omega) - X(\omega)| > \tfrac{1}{k}\} =$$

$$= \bigcup_{k=1}^\infty \overline{\lim_{n\to\infty}} \{|X_n - X| > \tfrac{1}{k}\} = \bigcup_{k=1}^\infty A_{1/k}$$

Ne segue che se gli eventi $A_{1/k}$ sono trascurabili per ogni k, anche l'evento $\{\overline{\lim}_{n\to\infty} |X_n - X| > 0\}$ lo è e quindi $X_n \overset{q.c.}{\to} X$.

Abbiamo quindi dimostrato

Lemma 3.4 *$X_n \to X$ q.c. se e solo se*

$$\mathrm{P}\Big(\overline{\lim_{n\to\infty}}\{|X_n - X| > \delta\}\Big) = 0$$

per ogni $\delta > 0$.

Quindi il lemma di Borel-Cantelli fornisce un criterio di convergenza q.c: se per ogni $\delta > 0$ la serie di termine generale $\mathrm{P}(|X_n - X| > \delta)$ è sommabile, allora $X_n \overset{q.c.}{\to} X$. Questo criterio sarà molto utile nel seguito.

Inoltre, per il lemma di Fatou,

$$\begin{aligned}\mathrm{P}\Big(\overline{\lim_{n\to\infty}}\{|X_n - X| > \delta\}\Big) &= \mathrm{E}\Big[\overline{\lim_{n\to\infty}}\, 1_{\{|X_n - X|>\delta\}}\Big] \geq \\ &\geq \overline{\lim_{n\to\infty}}\, \mathrm{E}\big[1_{\{|X_n - X|>\delta\}}\big] = \overline{\lim_{n\to\infty}}\, \mathrm{P}(|X_n - X| > \delta)\end{aligned}$$

e dunque se $X_n \overset{q.c.}{\to} X$ allora $\overline{\lim}_{n\to\infty} \mathrm{P}(|X_n - X| > \delta) = 0$ per ogni δ. Ovvero

Proposizione 3.5 *La convergenza q.c. implica quella in probabilità.*

L'Esempio 3.2 mostra che il viceversa non è vero: le successione $(X_n/\log n)_n$ tende a zero in probabilità, e anzi in L^p per ogni $p > 0$,:

$$\mathrm{E}\Big[\Big(\tfrac{X_n}{\log n}\Big)^p\Big] \leq \frac{1}{(\log n)^p}\mathrm{E}(X_1^p) = \frac{1}{(\log n)^p} \underbrace{\int_0^{+\infty} x^p \mathrm{e}^{-\lambda x}\, dx}_{<+\infty}$$

La convergenza non ha però luogo q.c.: per l'Esempio 3.2 si ha con probabilità 1

$$\frac{X_n}{\log n} \geq \varepsilon$$

infinite volte non appena $\varepsilon \leq \frac{1}{\lambda}$.

Esempio 3.6 Consideriamo lo spazio di probabilità $([0, 1], \mathscr{B}[0, 1], dx)$ e su di esso la successione di v.a. $(X_n)_n$ definita da

$$X_{2^m+k} = 1_{[k/2^m,(k+1)/2^m]} \qquad\qquad \text{se } k = 0, \dots, 2^m - 1$$

È chiaro che $\mathrm{P}(|X_{2^m+k}| > 0) = 2^{-m}$, e quindi $X_n \overset{\mathrm{P}}{\to} 0$. D'altra parte, se $\omega \in [0, 1]$, è chiaro che per ogni $m \geq 0$ esiste k tale che $\omega \in [k/2^m, (k + 1)/2^m]$. Dunque $X_n(\omega) = 1$ per infiniti indici n e $\overline{\lim}_{n\to\infty} X_n(\omega) = 1$.

Proposizione 3.7 *Se $(X_n)_n$ converge a X in probabilità, allora esiste una sottosuccessione $(X_{n_k})_k$ tale che $X_{n_k} \stackrel{q.c.}{\to} X$.*

Dimostrazione. Per ogni k intero positivo si ha

$$\lim_{n\to\infty} \mathrm{P}(|X_n - X| > 2^{-k}) = 0$$

Esiste quindi una successione d'interi $(n_k)_k$, che possiamo supporre strettamente crescente, tale che

$$\mathrm{P}(|X_{n_k} - X| > 2^{-k}) \le 2^{-k}$$

Poiché per ogni $\delta > 0$ fissato esiste k_0 tale che per $k > k_0$ si abbia $2^{-k} \le \delta$, allora per $k > k_0$

$$\mathrm{P}(|X_{n_k} - X| > \delta) \le \mathrm{P}(|X_{n_k} - X| > 2^{-k}) \le 2^{-k}$$

Quindi la serie di termine generale $\mathrm{P}(|X_{n_k} - X| > \delta)$ è sommabile. Per il lemma di Borel-Cantelli $\mathrm{P}(\overline{\lim}_{n\to\infty}\{|X_{n_k} - X| > \delta\}) = 0$ che, per il criterio del Lemma 3.18, implica $X_{n_k} \stackrel{q.c}{\to} X$.

La Proposizione 3.19 implica, in particolare, che il limite in probabilità è unico, a meno di una P–equivalenza, cosa che non era ovvia dalla definizione (cioè se X e Y sono due limiti in probabilità di una stessa successione di v.a., allora $\mathrm{P}(X \ne Y) = 0$). Inoltre per la Proposizione 3.19 la convergenza in L^p implica la convergenza q.c. per una sottosuccessione. L'Esempio 3.6 mostra una successione di v.a. che converge in L^p e non q.c.

È chiaro infine che se $X_n \stackrel{q.c}{\to} X$ e le v.a. $(X_n)_n$ sono tutte maggiorate in modulo da una medesima v.a. $Y \in L^p$, allora $X_n \to X$ in L^p (in esercizio: si usa due volte il teorema di Lebesgue, prima per provare che $X \in L^p$ e poi la convergenza); è facile però costruire esempi di successioni che convergono q.c. ma non in L^p.

3.3 Le leggi forti dei grandi numeri

In questo paragrafo vedremo che, sotto ipotesi molto deboli, se $(X_n)_n$ è una successione di v.a. indipendenti (o almeno non correlate) e aventi speranza matematica m finita, allora la loro media empirica

$$\bar{X}_n = \frac{1}{n}(X_1 + \ldots + X_n)$$

converge quasi certamente a m. Questo tipo di risultati si chiama una *legge forte* dei grandi numeri, in contrapposizione alle leggi deboli nelle quali la tesi riguarda una convergenza in L^p o in probabilità.

Porremo nel seguito $S_n = X_1 + \ldots + X_n$. Inoltre osserviamo che si può supporre $m - 0$. Altrimenti si potrebbe porre $Y_n = X_n - m$; le v.a Y_n avrebbero media 0 e si avrebbe $\bar{Y}_n = \bar{X}_n - m$; infine dimostrare che $\bar{X}_n \stackrel{q.c}{\to} m$ è lo stesso che dimostrare che $\bar{Y}_n \stackrel{q.c.}{\to} 0$.

Teorema 3.8 (Legge forte di Rajchmann) *Se le* $(X_n)_n$ *è una successione di v.a. tutte di media* m, *aventi varianza finita e a due a due non correlate, allora se*

$$\sup_n \mathrm{Var}(X_n) = \alpha < +\infty \tag{3.2}$$

si ha $\bar{X}_n \overset{q.c.}{\to} m$.

Dimostrazione. Supponiamo,come abbiamo detto, $m = 0$. Per ogni $\varepsilon > 0$ si ha, per la disuguaglianza di Chebyshev

$$\mathrm{P}(|\bar{X}_{n^2}| > \delta) \le \frac{1}{\delta^2}\,\mathrm{Var}(\bar{X}_{n^2}) = \frac{1}{\delta^2 n^4}\sum_{k=1}^{n^2}\mathrm{Var}(X_k) \le \frac{\alpha}{\delta^2}\cdot\frac{1}{n^2}$$

Quindi la serie

$$\sum_{k=1}^{\infty}\mathrm{P}(|\bar{X}_{n^2}| > \delta)$$

è convergente ed il Lemma di Borel-Cantelli dà $\mathrm{P}(\overline{\lim}_{n\to\infty}\{|\bar{X}_{n^2}| > \delta\}) = 0$; ciò implica per il Lemma 3.18 che la sottosuccessione $(\bar{X}_{n^2})_n$ converge a 0 q.c. Resta ora da controllare il comportamento di $\bar{X}_n$ tra due istanti consecutivi della forma n^2. Per questo poniamo

$$D_n = \sup_{n^2 \le k < (n+1)^2} |S_k - S_{n^2}|$$

(ricordiamo che $S_k = X_1 + \ldots + X_k$) per cui se $n^2 \le k < (n+1)^2$

$$\frac{|S_k|}{k} \le \frac{|S_{n^2}| + D_n}{k} \le \frac{1}{n^2}(|S_{n^2}| + D_n) = \bar{X}_{n^2} + \frac{1}{n^2}D_n$$

e quindi basta dimostrare che $\dfrac{D_n}{n^2} \to 0$ q.c. Ma

$$D_n^2 = \sup_{n^2 \le k < (n+1)^2}(S_k - S_{n^2})^2 \le \sum_{n^2 \le k < (n+1)^2}(S_k - S_{n^2})^2$$

$$\mathrm{E}(D_n^2) \le \sum_{n^2 \le k < (n+1)^2}\mathrm{E}((S_k - S_{n^2})^2)$$

Poiché le X_n sono non correlate

$$\mathrm{E}((S_k - S_{n^2})^2) = \mathrm{E}((X_{n^2+1} + \ldots + X_k)^2) =$$

$$= \sum_{i=n^2+1}^{k}\mathrm{Var}(X_i) \le [(n+1)^2 - n^2 - 1]\cdot\alpha = 2n\alpha$$

Quindi

$$\mathrm{E}(D_n^2) \le [(n+1)^2 - n^2 - 1] \cdot 2n\alpha = 4n^2\alpha$$

e dunque, per ogni $\delta > 0$, per la disuguaglianza di Markov 2.13

$$\mathrm{P}\Big(\big|\frac{1}{n^2} D_n\big| > \delta\Big) \le \frac{1}{\delta^2 n^4}\mathrm{E}(D_n^2) \le \frac{4\alpha}{\delta^2}\,\frac{1}{n^2}$$

Come nella prima parte della dimostrazione, il Lemma di Borel-Cantelli ed il Lemma 3.18 permettono di concludere che $\frac{1}{n^2} D_n \to 0$ q.c.

Enunciamo infine, senza dimostrazione, la più celebre delle leggi dei grandi numeri. In esse l'ipotesi di esistenza di momenti sono più deboli, ma si suppone in compenso che le v.a. siano indipendenti ed equidistribuite.

Teorema 3.9 (Legge forte di Kolmogorov) *Sia $(X_n)_n$ una successione di v.a. indipendenti e tutte di legge μ. Allora*

a) Se $\int |x|\, d\mu < +\infty$ allora $\bar X_n \to m = \int x\, d\mu$ q.c.

b) Se $\int |x|\, d\mu = +\infty$, allora una almeno delle due v.a. terminali

$$\overline{\lim_{n\to\infty}}\ \bar X_n \qquad \text{e} \qquad \underline{\lim_{n\to\infty}}\ \bar X_n$$

sono q.c. infinite (cioè almeno una di esse prende il valore $+\infty$ q.c. oppure il valore $-\infty$ q.c.).

3.4 Convergenza in legge

Vedremo ora un altro tipo di convergenza di v.a. Siano $(E, \mathscr{E})$ uno spazio misurabile e $\mu, \mu_n, n \ge 1$ misure su $(E, \mathscr{E})$. Un modo tipico (non l'unico) di definire una nozione di convergenza $\mu_n \to \mu$ è il seguente: si fissa una classe $\mathscr{D}$ di funzioni misurabili $f : E \to \mathbb{R}$ e si definisce $\mu_n \to \mu$ se e solo se

$$\lim_{n\to\infty} \int f\, d\mu_n = \int f\, d\mu$$

per ogni $f \in \mathscr{D}$. Naturalmente a seconda della classe $\mathscr{D}$ prescelta si ottengono tipi di convergenza diversi (eventualmente non confrontabili tra loro).

Nel seguito, per semplificare le notazioni useremo talvolta la scrittura $\mu(f)$ al posto di $\int f\, d\mu$.

Definizione 3.10 *Siano E uno spazio topologico e μ, μ_n, $n \ge 1$ misure finite su $(E, \mathscr{B}(E))$. Diremo che $(\mu_n)_n$ converge a μ strettamente se e solo se per ogni funzione $f \in \mathscr{C}_b(E)$ (funzioni continue e limitate su E) si ha*

$$\lim_{n\to\infty} \int f\, d\mu_n = \int f\, d\mu \tag{3.3}$$

Supporremo sempre che lo spazio topologico E sia metrico e separabile.

Osserviamo intanto che il limite stretto è unico. Infatti se simultaneamente $\mu_n \to \mu$ e $\mu_n \to \nu$ allora necessariamente

$$\int f\,d\mu = \int f\,d\nu \tag{3.4}$$

per ogni funzione $f \in \mathscr{C}_b$ e dunque μ e ν coincidono.

Proposizione 3.11 *Siano $\mathscr{D}$ uno spazio vettoriale di funzioni misurabili limitate su $(E, \mathscr{E})$, μ, μ_n, $n \geq 1$ misure di probabilità su $(E, \mathscr{E})$. Allora perché la relazione*

$$\mu_n(f) \underset{n\to\infty}{\to} \mu(f) \tag{3.5}$$

sia vera per ogni $f \in \mathscr{D}$ basta che essa sia vera per ogni funzione f appartenente ad un sottoinsieme totale H di $\mathscr{D}$.

Dimostrazione. Per definizione H è totale in $\mathscr{D}$ se e solo se lo spazio vettoriale $\mathscr{H}$ delle combinazioni lineari di funzioni di H è denso in $\mathscr{D}$ nella norma uniforme.

Se (3.5) è vera per ogni $f \in H$, per linearità, essa lo è evidentemente per ogni $f \in \mathscr{H}$. Siano ora $f \in \mathscr{D}$ e $(g_k)_k$ una successione di funzioni di $\mathscr{H}$ convergente a f uniformemente. Per $\varepsilon > 0$, sia k abbastanza grande perché sia $\|f - g_k\|_\infty \leq \varepsilon$; dunque per ogni n

$$\int |f - g_k|\,d\mu_n \leq \varepsilon, \qquad \int |f - g_k|\,d\mu \leq \varepsilon\ .$$

Sia ora n_0 tale che $|\mu_n(g_k) - \mu(g_k)| \leq \varepsilon$ per $n \geq n_0$; allora per $n \geq n_0$

$$|\mu_n(f) - \mu(f)| \leq |\mu_n(f) - \mu_n(g_k)| + |\mu_n(g_k) - \mu(g_k)| + |\mu(g_k) - \mu(f)| \leq 3\varepsilon$$

che per l'arbitrarietà di ε implica la tesi.

Se per di più E è anche localmente compatto allora si può dimostrare che esiste una successione crescente $(h_p)_p$ di funzioni continue a supporto compatto tali che $\sup_p h_p = 1$. Se μ è una probabilità su $(E, \mathscr{B}(E))$, allora $\mu(h_p) \nearrow \mu(1) = \mu(E) = 1$ ed è chiaro che per ogni $\varepsilon > 0$ esiste p tale che $\mu(h_p) \geq 1 - \varepsilon$.

Questa proprietà permette di stabilire il criterio seguente, quando le misure che si considerano sono di probabilità.

Proposizione 3.12 *Date le misure di probabilità $\mu, \mu_n, n \geq 1$ sullo spazio metrico localmente compatto separabile E, allora $\mu_n \to \mu$ strettamente se e solo se $\mu_n(f) \to \mu(f)$ per ogni funzione continua a supporto compatto.*

Dimostrazione. Fissiamo $f \in \mathscr{C}_b$ e supponiamo $f \geq 0$. Allora $fh_p \nearrow f$ e le funzioni fh_p sono a supporto compatto. Ricordando che $\mu(1) = \mu_n(1) = 1$ si ha

$$\begin{aligned}|\mu_n(f) - \mu(f)| &\leq |\mu_n(fh_p) - \mu(fh_p)| + |\mu_n((1-h_p)f)| + |\mu((1-h_p)f)| \leq \\ &\leq |\mu_n(fh_p) - \mu(fh_p)| + |f|_\infty \mu_n(1-h_p) + |f|_\infty \mu(1-h_p) = \\ &= |\mu_n(fh_p) - \mu(fh_p)| + |f|_\infty(\mu(h_p) - \mu_n(h_p)) + 2|f|_\infty \mu(1-h_p)\end{aligned}$$

Basta ora scegliere prima p abbastanza grande perché $\mu(1-h_p)$ sia $\leq \varepsilon$ e poi n abbastanza grande perché gli altri due termini nell'ultima disuguaglianza siano anch'essi $\leq \varepsilon$ per ottenere

$$|\mu_n(f) - \mu(f)| \leq 2\varepsilon(1 + |f|_\infty)$$

e per l'arbitrarietà di ε si ha (3.3).

D'ora in avanti supporremo che E è uno spazio metrico localmente compatto separabile e μ_n, μ indicheranno delle probabilità su $(E, \mathscr{B}(E))$.

Osservazione 3.13 Combinando le Proposizioni 3.11 e 3.18 se E è uno spazio metrico localmente compatto e separabile per provare la convergenza stretta basta verificare (3.3) per ogni $f \in \mathscr{C}_K(E)$; ovvero per ogni $f \in \mathscr{C}_0(E)$ (funzioni nulle all'infinito) o comunque per ogni sottoinsieme totale in $\mathscr{C}_0(E)$.

Se $E = \mathbb{R}^d$, un sottoinsieme totale di cui ci serviremo è quello formato dalle funzioni f della forma $f(x) = \mathrm{e}^{-a|x|^2 + i\langle b,x\rangle}$ al variare di $a > 0$ e $b \in \mathbb{R}$. Che si tratti di un sottoinsieme totale in $\mathscr{C}_0$ è una conseguenza del teorema di Stone-Weierstrass, come abbiamo osservato nella dimostrazione del Teorema 3.21.

Osservazione 3.14 La Proposizione 3.18 ha un'importanza notevole perché gli spazi $\mathscr{C}_K$ e $\mathscr{C}_0$, per uno spazio metrico separabile localmente compatto E, sono separabili nella topologia uniforme (mentre invece $\mathscr{C}_b$ non lo è in generale). Questo implica che per verificare la convergenza stretta di misure basta provare (3.3) per f che varia in un sottoinsieme *numerabile* denso in $\mathscr{C}_0$. Questo fatto è cruciale e ne vedremo varie applicazioni, a cominciare dal prossimo risultato.

Osservazione 3.15 Se $\mu, \mu_n, n \geq 1$, sono leggi di probabilità sullo spazio topologico E, Ψ una applicazione continua da E allo spazio topologico F e se indichiamo con ν_n, ν rispettivamente le immagini di μ_n, μ tramite Ψ, allora $\nu_n \to \nu$.

Infatti se $f : F \to \mathbb{R}$ è continua limitata, allora $f \circ \Psi$ è continua e limitata da E in $\mathbb{R}$. Dunque

$$\nu_n(f) = \mu_n(f \circ \Psi) \to \mu(f \circ \Psi) = \nu(f)$$

Siano $\mu, \mu_n, n \geq 1$, probabilità su $(\mathbb{R}^d, \mathscr{B}(\mathbb{R}^d))$ e supponiamo che $\mu_n \to \mu$. Allora è chiaro che $\hat{\mu}_n(\theta) \to \hat{\mu}(\theta)$. Infatti per ogni $\theta \in \mathbb{R}^d$

$$\hat{\mu}(\theta) = \int \mathrm{e}^{i\langle x,\theta\rangle}\, d\mu(x)$$

cioè $\hat{\mu}(\theta)$ è l'integrale rispetto a μ della funzione $x \to e^{i\langle x,\theta\rangle}$ che è continua e limitata. Ci si può domandare viceversa se la convergenza delle funzioni caratteristiche implichi la convergenza stretta.

Proposizione 3.16 *Siano $\mu, \mu_n, n \ge 1$, leggi di probabilità su $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. Allora $(\mu_n)_n$ converge strettamente a μ se e solo se $\hat{\mu}_n(\theta) \to \hat{\mu}(\theta)$ per ogni $\theta \in \mathbb{R}^d$.*

Dimostrazione. Supponiamo che le funzioni caratteristiche di μ_n convergano alla funzione caratteristica di μ; per l'Osservazione 3.13, per provare che $\mu_n \to \mu$ basta verificare che $\mu_n(f) \to \mu(f)$ quando f è della forma $x \to e^{-a|x|^2+i\langle b,x\rangle}$. Abbiamo già visto, nella dimostrazione del Teorema 2.20, che per ogni probabilità μ su $\mathbb{R}$ si ha

$$\frac{1}{(2\pi)^{d/2}\sigma^d}\int \hat{\mu}(y)\exp\Big(-\frac{|y-\theta|^2}{2\sigma^2}\Big)\,dy = \int e^{i\langle\theta,x\rangle}e^{-\frac{1}{2}\sigma^2|x|^2}\,\mu(dx)$$

Dunque per il Teorema di Lebesgue, poichè $|\hat{\mu}_n(\theta)| \le 1$,

$$\int e^{i\langle\theta,x\rangle}e^{-\frac{1}{2}\sigma^2|x|^2}\,\mu_n(dx) = \frac{1}{(2\pi)^{d/2}\sigma^d}\int \hat{\mu}_n(y)\exp\Big(-\frac{|y-\theta|^2}{2\sigma^2}\Big)\,dy \to$$
$$\to \frac{1}{(2\pi)^{d/2}\sigma^d}\int \hat{\mu}(y)\exp\Big(-\frac{|y-\theta|^2}{2\sigma^2}\Big)\,dy = \int e^{i\langle\theta,x\rangle}e^{-\frac{1}{2}\sigma^2|x|^2}\,\mu(dx)$$

che conclude la dimostrazione.

Proposizione 3.17 *Siano $\mu, \mu_n, n \ge 1$, leggi di probabilità su $(E, \mathcal{B}(E))$. Allora $(\mu_n)_n$ converge strettamente a μ se e solo se è verificata una delle proprietà seguenti.*

a) Per ogni funzione $f : E \to \mathbb{R}$ semicontinua inferiormente e inferiormente limitata

$$\varliminf_{n\to\infty}\int f\,d\mu_n \ge \int f\,d\mu \tag{3.6}$$

b) Per ogni funzione $f : E \to \mathbb{R}$ semicontinua superiormente e superiormente limitata

$$\varlimsup_{n\to\infty}\int f\,d\mu_n \le \int f\,d\mu \tag{3.7}$$

c) Per ogni funzione f boreliana limitata e tale che l'insieme dei suoi punti di discontinuità sia μ–trascurabile

$$\lim_{n\to\infty}\int f\,d\mu_n = \int f\,d\mu \tag{3.8}$$

Dimostrazione. È chiaro che a) e b) sono equivalenti tra loro (basta considerare che se f è come in a), allora $-f$ è come in b)) e che insieme implicano la convergenza stretta, perché se $f \in \mathscr{C}_b$, allora a f si possono applicare simultaneamente (3.6) e (3.7), ottenendo (3.3) .

Viceversa, supponiamo che $\mu_n \to \mu$ strettamente e che f sia s.c.i. e inferiormente limitata. Allora (proprietà delle funzioni s.c.i.) esiste una successione crescente di funzioni continue limitate $(f_k)_k$ tale che $\sup_k f_k = f$. Poiché $f_k \le f$, per ogni k fissato abbiamo

$$\int f_k\,d\mu = \lim_{n\to\infty} \int f_k\,d\mu_n \le \underline{\lim}_{n\to\infty} \int f\,d\mu_n$$

e prendendo il sup in k di questa relazione, per il Teorema di B.Levi segue la (3.6).

Mostriamo ora che se $\mu_n \to \mu$ strettamente, allora vale c) (il viceversa è ovvio). Consideriamo le due funzioni f^* e f_* definite da

$$f_*(x) = \underline{\lim}_{y\to x} f(y) \qquad\qquad f^*(x) = \overline{\lim}_{y\to x} f(y)$$

Chiaramente $f_* \le f \le f^*$ e si può dimostrare che f_* è s.c.i. mentre f^* è s.c.s. Inoltre è chiaro che le tre funzioni sono limitate e coincidono nei punti di continuità di f; poiché supponiamo che questi ultimi costituiscono un insieme di misura 0 per μ

$$\int f_*\,d\mu = \int f\,d\mu = \int f^*\,d\mu$$

(3.6) e (3.7) danno quindi

$$\int f\,d\mu = \int f_*\,d\mu \le \underline{\lim}_{n\to\infty} \int f_*\,d\mu_n \le \underline{\lim}_{n\to\infty} \int f\,d\mu_n$$
$$\int f\,d\mu = \int f^*\,d\mu \ge \overline{\lim}_{n\to\infty} \int f^*\,d\mu_n \ge \overline{\lim}_{n\to\infty} \int f\,d\mu_n$$

che insieme permettono di concludere.

Se $\mu_n \to \mu$ e $A \in \mathscr{B}(E)$, si può dire che $\mu_n(A) \to \mu(A)$? La proposizione precedente permette di rispondere a questa questione. Se $G \subset E$ è un aperto, allora la sua funzione indicatrice 1_G è s.c.i. e (3.6) implica che se $\mu_n \to \mu$ strettamente allora

$$\underline{\lim}_{n\to\infty} \mu_n(G) = \underline{\lim}_{n\to\infty} \int 1_G\,d\mu_n \ge \mu(G) \tag{3.9}$$

e analogamente se F è chiuso

$$\overline{\lim}_{n\to\infty} \mu_n(F) = \overline{\lim}_{n\to\infty} \int 1_F\,d\mu_n \le \mu(F) \tag{3.10}$$

Naturalmente in (3.9) vale il segno di uguaglianza e si ha un vero limite, che G sia aperto o no, se l'insieme ∂G è μ-trascurabile. Infatti ∂G è l'insieme dei punti di discontinuità di 1_G.

Se $E = \mathbb{R}$ vale anche il seguente criterio.

Proposizione 3.18 *Siano μ, μ_1, $n \geq 1$, misure di probabilità su $\mathbb{R}$ e indichiamo con F, $F_1, F_2, \dots$ le rispettive funzioni di ripartizione. Allora $\mu_n \to \mu$ strettamente se e solo se per ogni punto $x \in \mathbb{R}$ di continuità per F si ha*

$$\lim_{n\to\infty} F_n(x) = F(x) \tag{3.11}$$

Dimostrazione. Supponiamo μ_n converga a μ strettamente. Se x è un punto di continuità per F allora sappiamo che $\mu(\{x\}) = 0$. Poiché $\{x\}$ è la frontiera di $]-\infty, x]$

$$F_n(x) = \mu_n(]-\infty, x]) \to \mu(]-\infty, x]) = F(x)$$

Viceversa supponiamo che valga la (3.11) per ogni punto di continuità x di F. Se a e b sono punti di continuità per F allora

$$\mu_n(]a, b]) = F_n(b) - F_n(a) \to F(b) - F(a) = \mu(]a, b]) \tag{3.12}$$

Poiché i punti di discontinuità della funzione crescente F sono al più una infinità numerabile, ne segue che (3.12) è vera per almeno un insieme di a, b in un insieme S denso in $\mathbb{R}$. Per linearità dunque $\mu_n(g) \to \mu(g)$ per ogni funzione g che sia combinazione lineare di funzioni indicatrici di intervalli $]a, b]$ con $a, b \in S$. È facile ora vedere che ogni funzione $f \in \mathscr{C}_K(\mathbb{R})$ si può approssimare uniformemente con funzioni di questo tipo.

Esempi 3.19

a) $\mu_n = \delta_{1/n}$ (massa di Dirac nel punto $\frac{1}{n}$). Allora $\mu_n \to \delta_0$ strettamente. Infatti se $f \in \mathscr{C}_b$

$$\int f\, d\mu_n = f(\tfrac{1}{n}) \to f(0) = \int f\, d\delta_0$$

Da notare che se $G =]0, 1[$, allora $\mu_n(G) = 1$ per ogni n e quindi

$$\lim_{n\to\infty} \mu_n(G) = 1$$

mentre $\delta_0(G) = 0$. Nella (3.9) vale dunque, in questo caso, una disuguaglianza stretta, cosa possibile perché $\partial G = \{0, 1\}$ e $\delta_0(\partial G) > 0$.

b) $\mu_n = \frac{1}{n}\sum_{k=0}^{n-1}\delta_{k/n}$. Cioè μ_n è una somma di masse di Dirac, ciascuna di peso $\frac{1}{n}$ poste nei punti $0, \frac{1}{n}, \dots, \frac{n-1}{n}$. Se $f \in \mathscr{C}_b$ allora

$$\int f\,d\mu_n = \sum_{k=0}^{n-1}\frac{1}{n}f(\tfrac{k}{n})$$

Nel termine a destra riconosciamo la somma di Riemann di f sull'intervallo $[0, 1]$ rispetto alla partizione $0, \frac{1}{n}, \dots, \frac{n-1}{n}$. Poiché f è continua le somme di Riemann convergono all'integrale e quindi

$$\lim_{n\to\infty}\int f\,d\mu_n = \int_0^1 f(x)\,dx$$

che prova che $(\mu_n)_n$ converge strettamente verso la distribuzione uniforme su $[0, 1]$. Si può giungere allo stesso risultato anche calcolando il limite delle funzioni caratteristiche o delle funzioni di ripartizione.

c) $\mu_n \sim B(n, \frac{\lambda}{n})$. Mostriamo che $(\mu_n)_n$ converge ad una legge di Poisson di parametro λ; cioè l'approssimazione di una legge binomiale con parametro n grande che abbiamo visto nell'Esempio 2.24 era in realtà una convergenza stretta.

Ciò si può vedere in modi diversi. Conosciamo infatti ormai tre metodi per verificare la convergenza stretta: la definizione, la convergenza delle funzioni di ripartizione (Proposizione 3.18) e la convergenza delle funzioni caratteristiche.

Ad esempio in questo caso la funzione di ripartizione F del limite è continua ovunque tranne che per gli x interi positivi. Dunque se $x \notin \mathbb{N}$ e $x > 0$

$$F_n(x) = \sum_{k=0}^{\lfloor x\rfloor}\binom{n}{k}\Big(\frac{\lambda}{n}\Big)^k\Big(1-\frac{\lambda}{n}\Big)^{n-k} \to \sum_{k=0}^{\lfloor x\rfloor}\mathrm{e}^{-\lambda}\frac{\lambda^k}{k!} = F(x)$$

poiché nella somma compaiono solo un numero finito di termini ($\lfloor\ \rfloor$ indica al solito la funzione parte intera). Se invece $x < 0$ non c'è niente da dimostrare poiché $F_n(x) = 0 = F(x)$. Da notare che in questo caso $F_n(x) \to F(x)$ per ogni x, e non solo per gli x che sono punti di continuità.

Avremmo anche potuto calcolare le funzioni caratteristiche ed il loro limite:

$$\hat{\mu}_n(\theta) = \Big(1 - \frac{\lambda}{n} + \frac{\lambda}{n}\mathrm{e}^{i\theta}\Big)^n = \Big(1 + \frac{\lambda}{n}(\mathrm{e}^{i\theta} - 1)\Big)^n \to \mathrm{e}^{\lambda(\mathrm{e}^{i\theta}-1)}$$

che è la funzione caratteristica di una legge di Poisson di parametro λ. Quindi per il Teorema di P.Lévy $\mu_n \to \mathrm{Poiss}(\lambda)$.

d) $\mu_n \sim N(b, \frac{1}{n})$. Sappiamo che le leggi μ_n hanno densità date da curve a campana centrate tutte nel punto b e che tendono ad essere sempre più alte e più strette al crescere di n. Ciò suggerisce che le μ_n tendono a concentrarsi sempre più vicino a b.

Anche in questo caso per studiare la convergenza si può sia calcolare il limite delle funzioni di ripartizione, sia usare le funzioni caratteristiche. Quest'ultimo metodo è in questo caso più semplice:

$$\hat{\mu}_n(\theta) = \mathrm{e}^{ib\theta}\mathrm{e}^{-\theta^2/2n} \to \mathrm{e}^{ib\theta}$$

che è la funzione caratteristica di una legge δ_b.

e) $\mu_n \sim N(0, n)$. La densità delle μ_n è data

$$g_n(x) = \frac{1}{\sqrt{2\pi n}}\,\mathrm{e}^{-x^2/2n}$$

Poiché g_n è maggiorata, per ogni x, da $\frac{1}{\sqrt{2\pi n}}$, se fosse $\mu_n \to \mu$ si avrebbe per ogni intervallo $]a, b[$

$$\mu(]a, b[) \le \lim_{n\to\infty} \mu_n(]a, b[) = \lim_{n\to\infty} \int_a^b g_n\,dx \le \lim_{n\to\infty} \frac{b-a}{\sqrt{2\pi n}} = 0$$

Dunque μ darebbe probabilità 0 ad ogni intervallo aperto limitato. Poiché $\mathbb{R}$ si può ottenere come riunione numerabile di tali intervalli si avrebbe $\mu(\mathbb{R}) = 0$ e dunque μ non potrebbe essere una misura di probabilità. La stessa cosa si sarebbe potuta vedere con le funzioni caratteristiche: infatti

$$\hat{\mu}_n(\theta) = \mathrm{e}^{-\frac{1}{2}n\theta^2} \to \psi(\theta) = \begin{cases} 1 & \text{se } \theta = 0 \\ 0 & \text{se } \theta \ne 0 \end{cases}$$

La funzione ψ non può essere una funzione caratteristica (non è continua in 0). Consideriamo delle v.a. $X, X_1, X_2, \dots$ e indichiamo con $\mu_X, \mu_{X_1}, \mu_{X_2}, \dots$ le leggi rispettive. La convergenza di leggi di probabilità permette di definire una forma di convergenza di v.a.

Definizione 3.20 *Si dice che la successione* $(X_n)_n$ *converge a* X *in legge* ($X_n \stackrel{\mathcal{L}}{\to} X$) *se e solo se* $\mu_{X_n} \to \mu_X$ *strettamente.*

Osservazione 3.21 Per provare che $X_n \stackrel{\mathcal{L}}{\to} X$ basta verificare che

$$\lim_{n\to\infty} \mathrm{E}[f(X_n)] = \mathrm{E}[f(X)]$$

per ogni funzione $f \in \mathscr{C}_K$. Basta infatti osservare che

$$\mathrm{E}[f(X_n)] = \int f(x)\,d\mu_{X_n}(x) \qquad \mathrm{E}[f(X)] = \int f(x)\,d\mu_X(x)$$

Proposizione 3.22 *Se* $X_n \stackrel{\mathrm{P}}{\to} X$ *allora* $X_n \stackrel{\mathcal{L}}{\to} X$.

Dimostrazione. Continuiamo a supporre $E = \mathbb{R}^m$; il caso di uno spazio metrico generale si tratta in maniera assolutamente simile. Per l'Osservazione 3.13 basta dimostrare che

$$\mu_{X_n}(f) = \mathrm{E}(f(X_n)) \to \mathrm{E}(f(X)) = \mu_X(f)$$

per ogni funzione $f \in \mathscr{C}_K$. Poiché si tratta di funzioni uniformemente continue e limitate, per $\varepsilon > 0$ fissato esistono $\delta > 0$ tale che $|f(x) - f(y)| \le \varepsilon$ se $|x - y| \le \delta$; inoltre esiste n_0 tale che $\mathrm{P}(|X_n - X| > \delta) \le \varepsilon$ per $n > n_0$. Dunque

$$\begin{aligned}|\mu_{X_n}(f) - \mu_X(f)| &= |\mathrm{E}(f(X_n)) - \mathrm{E}(f(X))| \le \mathrm{E}\big(|f(X_n) - f(X)|\big) = \\ &= \mathrm{E}\big(|f(X_n) - f(X)|\, 1_{\{|X_n - X| \le \delta\}}\big) + \mathrm{E}\big(|f(X_n) - f(X)| 1_{\{|X_n - X| > \delta\}}\big) \le \\ &\le \varepsilon \mathrm{P}(|X_n - X| \le \delta) + 2\|f\|_\infty \mathrm{P}(|X_n - X| > \delta) \le \varepsilon(1 + 2\|f\|_\infty)\end{aligned}$$

da cui per l'arbitrarietà di ε si ha la tesi.

La convergenza in legge è dunque più debole di tutte quelle già viste: q.c, in probabilità e in L^p. Anzi, perché essa abbia luogo non è nemmeno necessario che le variabili aleatorie siano definite sullo stesso spazio di probabilità.

Esempio 3.23 Sia $(X_n)_n$ una successione di v.a. tale che $X_n \sim t(n)$. Allora $X_n \xrightarrow{\mathscr{L}} X$ dove $X \sim N(0, 1)$.

Siano $Z, Y_n, n = 1, 2, \dots$ delle v.a. indipendenti con $Z \sim N(0, 1)$ e $Y_n \sim \chi^2(1)$ per ogni n. Allora $S_n = Y_1 + \dots + Y_n \sim \chi^2(n)$ e S_n è indipendente da Z. Dunque la v.a. T_n definita da

$$T_n = \frac{Z}{\sqrt{S_n}}\sqrt{n} = \frac{Z}{\sqrt{\frac{S_n}{n}}}$$

segue una legge $t(n)$. D'altra parte per la legge dei grandi numeri $S_n/n \to \mathrm{E}[Y_1] = 1$ e dunque $T_n \xrightarrow{q.c.} Z$. Poiché la convergenza q.c. implica quella in legge ciò conclude la dimostrazione.

C'è però un caso in cui la convergenza in legge implica in probabilità.

Proposizione 3.24 *Se $(U_n)_n$ è una successione di v.a. definite su uno stesso spazio di probabilità e $U_n \xrightarrow{\mathscr{L}} U$ dove U è una v.a. costante, allora $U_n \xrightarrow{\mathrm{P}} U$.*

Dimostrazione. Sia $u_0 \in E$ tale che $U = u_0$ q.c. Indichiamo con B_η la palla aperta di centro u_0 e raggio η; allora si può scrivere

$$\mathrm{P}(d(U_n, u_0) \ge \eta) = \mathrm{P}(X_n \in B_\eta^c)$$

Ma B_η^c è un chiuso che ha probabilità 0 per la legge di U che è la massa di Dirac δ_{u_0}. Dunque $\overline{\lim}_{n\to\infty} \mathrm{P}(d(U_n, u_0) \ge \eta) = 0$ per la Proposizione 3.17.

3.5 Il teorema limite centrale, il test del χ^2

In questo capitolo vediamo il risultato di convergenza in legge più classico e importante.

Teorema 3.25 (Teorema limite centrale) *Sia $(X_n)_n$ una successione di v.a. k-dimensionali i.i.d., di media $m \in \mathbb{R}^k$ e di matrice di covarianza C. Allora posto*

$$S_n^* = \frac{X_1 + \ldots + X_n - nm}{\sqrt{n}},$$

S_n^* converge in legge a una v.a. normale multivariata $N(0, C)$.

Dimostrazione. Se $Y_i = X_i - m$, allora le Y_i sono centrate, hanno la stessa matrice di covarianza C e $S_n^* = \frac{1}{\sqrt{n}}(Y_1 + \ldots + Y_n)$. Se indichiamo con ϕ la funzione caratteristica delle Y_i, allora

$$\phi_{S_n^*}(\theta) = \phi\big(\tfrac{\theta}{\sqrt{n}}\big)^n = \Big(1 + \Big(\phi\big(\tfrac{\theta}{\sqrt{n}}\big) - 1\Big)\Big)^n$$

Calcolando lo sviluppo di Taylor intorno a $\theta = 0$ e, ricordando che

$$\phi'(0) = i\mathrm{E}(Y_1) = 0, \qquad \phi''(0) = -C_Y = -C,$$

si ha

$$\phi(\theta) = 1 - \frac{1}{2}\langle C\theta, \theta\rangle + o(|\theta|^2)$$

Quindi per $n \to +\infty$

$$\phi\big(\tfrac{\theta}{\sqrt{n}}\big) - 1 = -\frac{1}{2n}\langle C\theta, \theta\rangle + o(\tfrac{1}{n})$$

e, poiché $\log(1 + z) \sim z$ per $z \to 0$

$$\begin{aligned}
\lim_{n\to\infty} \phi_{S_n^*}(\theta) &= \lim_{n\to\infty}\Big(1 - \frac{1}{2n}\langle C\theta, \theta\rangle + o(\tfrac{1}{n})\Big)^n = \\
&= \lim_{n\to\infty} \exp\Big[n \log\Big(1 + \Big(\phi\big(\tfrac{\theta}{\sqrt{n}}\big) - 1\Big)\Big)\Big] = \lim_{n\to\infty} \exp\Big[n\Big(\phi\big(\tfrac{\theta}{\sqrt{n}}\big) - 1\Big)\Big] = \\
&= \lim_{n\to\infty} \exp\Big[n\Big(-\frac{1}{2n}\langle C\theta, \theta\rangle + o(\tfrac{1}{n})\Big)\Big] = \mathrm{e}^{-\frac{1}{2}\langle C\theta,\theta\rangle}
\end{aligned}$$

che è la funzione caratteristica di una v.a. $N(0, C)$. Basta ora applicare La Proposizione 3.16.

Corollario 3.26 *Sia $(X_n)_n$ una successione di v.a. reali i.i.d., di media m e varianza σ^2. Allora posto*

$$S_n^* = \frac{X_1 + \ldots + X_n - nm}{\sigma\sqrt{n}}$$

$S_n^* \xrightarrow{\mathcal{L}} N(0, 1)$.

Dimostrazione. Basta osservare che $S_n^* = \frac{1}{\sqrt{n}}(Y_1 + \ldots + Y_n)$, dove $Y_i = \frac{1}{\sigma}(X_i - m)$, e applicare il Teorema 3.25.

Vediamo ora una classica applicazione del teorema limite centrale alla statistica.

Sia $(X_n)_n$ una successione di v.a. indipendenti equidistribuite a valori in un insieme finito composto da m elementi, che supporremo essere $\{1, \ldots, m\}$ e poniamo $p_i = \mathrm{P}(X_1 = i)$, $i = 1, \ldots, n$. Supponiamo che i numeri p_i siano tutti > 0 e poniamo, per ogni $n > 0$, $i = 1, \ldots, m$,

$$N_i^{(n)} = \#\{k; k \le n,\ X_k = i\}, \qquad \bar{p}_i^{(n)} = \frac{N_i^{(n)}}{n}\cdot$$

Naturalmente $\sum_{i=1}^m N_i^{(n)} = n$, $\sum_{i=1}^m \bar{p}_i^{(n)} = 1$. Nel seguito ometteremo il sopraindice $^{(n)}$ e scriveremo N_i, $\bar{p}_i$ per semplicità. Consideriamo, per ogni n, la v.a.

$$T_n = \sum_{i=1}^m \frac{1}{np_i}(N_i - np_i)^2 = n \sum_{i=1}^m \frac{(\bar{p}_i - p_i)^2}{p_i}\cdot$$

Allora

Teorema 3.27 (Pearson)

$$T_n \xrightarrow{\mathcal{L}} \chi^2(m-1)$$

Dimostrazione. Consideriamo i vettori aleatori m-dimensionali Y_n definiti da

$$Y_n(\omega) = e_i \qquad \text{se } X_n(\omega) = i$$

dove $e_i = (0, \ldots, 0, 1, 0, \ldots, 0)^*$ è il vettore colonna di $\mathbb{R}^m$ avente tutte le coordinate nulle meno la i-esima, che è uguale a 1. Indichiamo con N, p e $\sqrt{p}$ i vettori di $\mathbb{R}^m$ di componenti N_i, p_i e $\sqrt{p_i}$ $i = 1, \ldots, m$ rispettivamente; quindi il vettore $\sqrt{p}$ ha modulo $= 1$. È chiaro che i vettori aleatori Y_n sono indipendenti e che $\mathrm{E}(Y_n) = p$; inoltre, facendo direttamente il calcolo, si vede che la matrice di covarianza di Y_n è $C = (c_{ij})_{ij}$ con $c_{ij} = p_i\delta_{ij} - p_i p_j$. Infine $\sum_{k=1}^n Y_k = N$. Quindi per il Teorema limite centrale le v.a.

$$Z_n = \frac{1}{\sqrt{n}}(N - np) = \frac{1}{\sqrt{n}} \sum_{k=1}^n (Y_k - \mathrm{E}(Y_k))$$

convergono in legge, per $n \to \infty$, verso una v.a. $N(0, C)$. Ora si può scrivere $T_n = f(Z_n)$ dove $f : \mathbb{R}^m \to \mathbb{R}^+$ è la funzione

$$f(z) = \sum_{i=1}^m \frac{z_i^2}{p_i}\cdot$$

Dunque $(T_n)_n$ converge in legge alla v.a. $f(Z)$, dove $Z \sim N(0, C)$. Ora $f(Z) = |W|^2$, dove W è il vettore gaussiano di componenti $W_i = Z_i/\sqrt{p_i}$. Indichiamo con K la matrice di covarianza di W. K ha per elementi i numeri $k_{ij} = c_{ij}/\sqrt{p_i p_j}$. Dunque $k_{ij} = \delta_{ij} - \sqrt{p_i}\sqrt{p_j}$. Si verifica subito che, per ogni $x \in \mathbb{R}^m$,

$$Kx = x - \langle\sqrt{p}, x\rangle\sqrt{p}$$

Dunque $K\sqrt{p} = 0$, mentre $Kx = x$ per ogni x ortogonale a $\sqrt{p}$. Dunque K non è altro che il proiettore sul sottospazio ortogonale a $\sqrt{p}$, che ha dimensione $m-1$. K ha quindi come autovalori 1 con molteplicità $m-1$ e 0 con molteplicità 1. Sia O una matrice ortogonale formata da autovettori di K. Se $\tilde{W} = OW$, allora la matrice di covarianza di $\tilde{W}$ è OKO^*, che è una matrice diagonale che ha sulla diagonale $m-1$ volte 1 ed una volta 0. Quindi le v.a. $\tilde{W}_i$ sono indipendenti e di esse $m-1$ sono $N(0, 1$, mentre una è uguale a 0 q.c. Quindi $|\tilde{W}|^2 = \tilde{W}_1^2 + \ldots + \tilde{W}_m^2 \chi^2(m-1)$. Basta ora osservare che $|W|^2 = |\tilde{W}|^2$.

Supponiamo di essere in presenza di v.a. $X_1, X_2, \ldots$ i.i.d. a valori in $\{1, \ldots, m\}$ che si suppone seguano una legge data da $\mathrm{P}(X_n = i) = p_i$ dove $p = (p_1, \ldots, p_m)$ è assegnato. Il Teorema di Pearson fornisce un modo per verificare questa ipotesi.

In effetti se l'ipotesi è vera $T_n \sim \chi^2(n-1)$, mentre se la loro legge fosse individuata da un altro vettore $q = (q_1, \ldots, q_m)$ diverso da p, avremmo per $n \to \infty$ $\bar{p}_i \to q_i$ per la legge dei grandi numeri, e dunque

$$T_n \simeq n \sum_{i=1}^{m} \frac{(q_i - p_i)^2}{p_i}$$

per cui T_n tenderebbe ad assumere valori grandi.

Esempio 3.28 Un dado viene lanciato 2000 volte ottenendo i seguenti risultati

1	2	3	4	5	6
388	322	314	316	344	316

Che ne pensate ?

Effettivamente il risultato 1 è apparso un numero di volte superiore agli altri: le frequenze sono

$\bar{p}_1$	$\bar{p}_2$	$\bar{p}_3$	$\bar{p}_4$	$\bar{p}_5$	$\bar{p}_6$
0.196	0.161	0.157	0.158	0.172	0.158

Prima di concludere che il dado non è equilibrato bisogna però stabilire se i risultati osservati si possono attribuire a normali fluttuazioni oppure sono significativamente lontani da quelli teorici. Sappiamo però che, sotto l'ipotesi che il dado sia equilibrato, la quantità

$$T_n = 2000 \times \sum_{i=1}^{6} (\bar{p}_i - \tfrac{1}{6})^2 \times 6 = 12.6$$

segue una legge che è approssimativamente $\chi^2(5)$, altrimenti tenderebbe ad assumere valori grandi. La questione è dunque: il valore osservato di T_n può essere considerato un valore tipico per una v.a. $\chi^2(5)$? Oppure è troppo grande? Si può affrontare la questione nel modo seguente: si fissa una soglia α (ad esempio $\alpha = 0.05$). Se $\chi^2_\alpha(5)$ indica il quantile di ordine α della legge $\chi^2(5)$, allora $\mathrm{P}(X > \chi^2_{1-\alpha}(5)) = \alpha$. Si decide quindi di respingere l'ipotesi se il valore di T_n osservato supera $\chi^2_{1-\alpha}(5)$. Uno sguardo alle tavole del χ^2 con 5 gradi di libertà mostra che $\chi^2_{0.95}(5) = 11.07$. Se ne conclude che il dado è molto probabilmente truccato. Nel linguaggio della statistica matematica il Teorema di Pearson ha permesso di respingere l'ipotesi che il dado fosse equilibrato al livello 5%. Il valore 12.6 corrisponde al quantile di ordine 97.26% di una legge $\chi^2(5)$. Dunque se il dado fosse equilibrato, un valore di T_n superiore a 12.6 si sarebbe potuto verificare con la probabilità del 2.7%

(I dati di questo esempio sono stati simulati con delle distribuzioni teoriche $q_1 = 0.2$, $q_2 = \ldots = q_6 = 0.16$).

Il Teorema di Pearson ha dunque un'importanza applicativa notevole in problemi di Statistica quando si voglia verificare se i dati seguono effettivamente una data distribuzione teorica. Per questo occorre sapere quanto debba essere grande n perché si possa supporre che T_n segua una legge vicina ad una $\chi^2(k-1)$. Una regoletta pratica, della cui validità teorica non discuteremo, richiede che debba essere $np_i \geq 5$ per ogni $i = 1, \ldots, k$.

Esempio 3.29 I dati del Riquadro 3.1 riguardano 6115 famiglie di 12 figli. Per ognuna di esse è stato riportato il numero N_k dei figli maschi. Un'ipotesi abbastanza naturale consiste nel supporre che ogni nascita dia luogo ad un maschio oppure ad una femmina con probabilità $\frac{1}{2}$, ed inoltre che gli esiti di parti diversi siano tra di loro indipendenti. Si può dire che questa ipotesi sia confermata dalle osservazioni?

Sotto l'ipotesi, la v.a. X =''numero di figli maschi'' segue una legge binomiale $B(12, \frac{1}{2})$, ovvero la probabilità di osservare una famiglia con k figli maschi dovrebbe essere pari a

$$p_k = \binom{12}{k}\Big(\frac{1}{2}\Big)^k\Big(1-\frac{1}{2}\Big)^{12-k} = \binom{12}{k}\Big(\frac{1}{2}\Big)^{12}$$

Siamo in una situazione classica di applicazione del teorema di Pearson, cioè di confronto tra una distribuzione empirica (i $\bar{p}_k$) e una teorica (la binomiale $B(12, \frac{1}{2})$). La condizione di applicabilità del Teorema di Pearson non è però soddisfatta poiché per $i = 1$ oppure $i = 12$ abbiamo $p_i = 2^{-12}$ e dunque

$$np_i = 6115 \cdot 2^{-12} = 1.49$$

k	N_k	p_k	$\bar{p}_k$	$p_k/\bar{p}_k$
0	3	0.000244	0.000491	0.49764
1	24	0.002930	0.003925	0.74646
2	104	0.016113	0.017007	0.94743
3	286	0.053711	0.046770	1.14840
4	670	0.120850	0.109567	1.10298
5	1033	0.193359	0.168929	1.14462
6	1343	0.225586	0.219624	1.02715
7	1112	0.193359	0.181848	1.06330
8	829	0.120850	0.135568	0.89143
9	478	0.053711	0.078168	0.68712
10	181	0.016113	0.029599	0.54438
11	45	0.002930	0.007359	0.39811
12	7	0.000244	0.001145	0.21327

Riquadro 3.1 Valori numerici delle osservazioni (N_k=numero di famiglie con k figli maschi) delle probabilità teoriche (p_k), di quelle empiriche ($\bar{p}_k = N_k/6115$) e del loro rapporto $p_k/\bar{p}_k$.

che è una quantità più piccola di 5 e dunque insufficiente all'applicazione del teorema di Pearson. Questa difficoltà si supera nel modo seguente. Consideriamo una nuova v.a. Y definita da

$$Y = \begin{cases} 1 & \text{se } X = 0 \\ i & \text{se } X = i \text{ per } i = 1, \dots, 11 \\ 11 & \text{se } X = 12 \end{cases}$$

In altre parole Y coincide con X se X prende i valori $1, \dots, 11$ mentre vale 1 anche su $\{X = 0\}$ e 11 su $\{X = 12\}$. Chiaramente la legge di Y è data da

$$\mathrm{P}(Y = i) = q_i = \begin{cases} p_0 + p_1 & \text{se } i = 1 \\ p_i & \text{se } i = 2, \dots, 10 \\ p_{11} + p_{12} & \text{se } i = 11 \end{cases}$$

È chiaro ora che se nelle osservazioni *raggruppiamo* le osservazioni delle classi 0 e 1 e delle classi 11 e 12, nelle ipotesi fatte le nuove distribuzioni empiriche così ottenute dovranno seguire la distribuzione di Y. Ovvero dovremo confrontare usando il teorema di Pearson le distribuzioni

k	q_k	$\bar{q}_k$
1	0.003174	0.004415
2	0.016113	0.017007
3	0.053711	0.046770
4	0.120850	0.109567
5	0.193359	0.168929
6	0.225586	0.219624
7	0.193359	0.181848
8	0.120850	0.135568
9	0.053711	0.078168
10	0.016113	0.029599
11	0.003174	0.008504

Ora il prodotto nq_1 vale $6115 \cdot .003174 = 19.41$, e l'approssimazione di Pearson è applicabile. Il calcolo numerico dà

$$T = 6115 \cdot \sum_{i=1}^{11} \frac{(\bar{q}_i - q_i)^2}{q_i} = 242.05$$

che è molto più grande dei quantili usuali della distribuzione $\chi^2(10)$. L'ipotesi che i dati seguissero una distribuzione $B(12, \frac{1}{2})$ è dunque respinta. Del resto qualche sospetto in questo senso sarebbe stato suscitato anche da un istogramma per confrontare valori teorici ed empirici, come nella Figura 3.2.

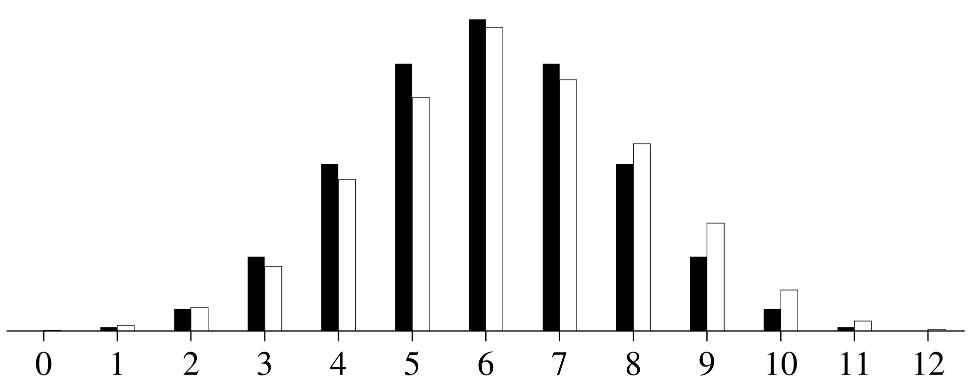

Figura 3.2 Le sbarre scure indicano i valori teorici p_k, quelle chiare i valori empirici $\bar{p}_k$.

In effetti, più che grosse discrepanze tra i valori teorici e quelli empirici, ciò che insospettisce è il fatto che i valori teorici superano quelli empirici per valori estremi e viceversa ne sono più piccoli per valori centrali. Ciò è messo ancor più in evidenza da un istogramma del quoziente $p_k/\bar{p}_k$, come nella Figura 3.3. In effetti se la differenza

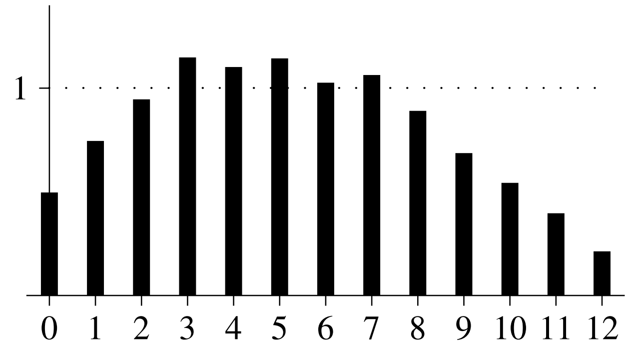

Figura 3.3 Istogramma dei valori del rapporto $p_k/\bar{p}_k$.

fosse attribuibile a fluttuazioni aleatorie piuttosto che ad una inadeguatezza del modello ci dovremmo aspettare una maggiore irregolarità nelle differenze tra i due tipi di valori.

Il modello proposto all'inizio per spiegare i dati, che prevedeva indipendenza tra gli esiti di parti diversi ed uguale probabilità di ottenere un maschio o una femmina deve quindi essere respinto.

3.6 Il lemma di Slutski

Vediamo in questo paragrafo delle trasformazioni che preservano la convergenza in legge. Un primo risultato di questo tipo è stato già visto nell'Osservazione 3.21.

Lemma 3.30 (Slutsky) *Siano* Z_n, U_n, $n \ge 1$ *v.a. definite su* $(\Omega, \mathscr{A}, \mathrm{P})$ *ed a valori in* $\mathbb{R}^p$ *e* $\mathbb{R}^q$ *rispettivamente e supponiamo che* $Z_n \xrightarrow{\mathscr{L}} Z$, $U_n \xrightarrow{\mathscr{L}} U$ *dove* U *è una v.a. costante che prende il solo valore* $u_0 \in \mathbb{R}^q$. *Allora*

i) $(Z_n, U_n) \xrightarrow{\mathscr{L}} (Z, u_0)$.
ii) Se $p = q$, *allora* $(Z_n + U_n)_n$ *converge a* $Z + u_0$.
iii) Se $q = 1$ *(cioè la successione* $(U_n)_n$ *è a valori reali) allora* $Z_n U_n \to u_0 Z$.

Dimostrazione. i) Se $\tau \in \mathbb{R}^p$, $\theta \in \mathbb{R}^q$, la funzione caratteristica di (Z_n, U_n) calcolata in $(\tau, \theta) \in \mathbb{R}^{p+q}$ è

$$\mathrm{E}[\mathrm{e}^{i\langle \tau, Z_n\rangle}\mathrm{e}^{i\langle \theta, U_n\rangle}] = \mathrm{E}[\mathrm{e}^{i\langle \tau, Z_n\rangle}\mathrm{e}^{\langle \theta, u_0\rangle}] + \mathrm{E}[\mathrm{e}^{i\langle \tau, Z_n\rangle}(\mathrm{e}^{i\langle \theta, U_n\rangle} - \mathrm{e}^{i\langle \theta, u_0\rangle})]$$

Il primo termine a secondo membro converge a $\mathrm{E}[\mathrm{e}^{i\langle \tau, Z\rangle}\mathrm{e}^{i\langle \theta, u_0\rangle}]$, basterà dunque provare che l'altro termine tende a 0. In effetti

$$\begin{aligned}|\mathrm{E}[\mathrm{e}^{i\langle \tau, Z_n\rangle}(\mathrm{e}^{i\langle \theta, U_n\rangle} - \mathrm{e}^{i\langle \theta, u_0\rangle})]| &\le \mathrm{E}[|\mathrm{e}^{i\langle \tau, Z_n\rangle}(\mathrm{e}^{i\langle \theta, U_n\rangle} - \mathrm{e}^{i\langle \theta, u_0\rangle})|] = \\ &= \mathrm{E}[|\mathrm{e}^{i\langle \theta, U_n\rangle} - \mathrm{e}^{i\langle \theta, u_0\rangle}|] = \mathrm{E}[f(U_n)]\end{aligned}$$

dove $f(x) = |\mathrm{e}^{i\langle \theta, x\rangle} - \mathrm{e}^{i\langle \theta, u_0\rangle}|$; f è una funzione continua limitata e quindi $\mathrm{E}[f(U_n)] \to \mathrm{E}[f(U)] = f(u_0) = 0$.

I punti ii) e iii) sono conseguenza di i) e dell'Osservazione 3.15: le applicazioni $(z, u) \to z + u$ e $(z, u) \to zu$ sono continue e $(Z_n, U_n) \xrightarrow{\mathscr{L}} (Z, u_0)$.

Teorema 3.31 *(Il metodo delta) Sia $(Z_n)_n$ una successione di v.a. a valori in $\mathbb{R}^p$, tale che*

$$\sqrt{n}(Z_n - z) \underset{n\to\infty}{\overset{\mathscr{L}}{\to}} Z \sim N(0, C)$$

Sia $\Phi : \mathbb{R}^p \to \mathbb{R}^q$ una funzione derivabile con derivata continua in z. Allora

$$\sqrt{n}(\Phi(Z_n) - \Phi(z)) \underset{n\to\infty}{\overset{\mathscr{L}}{\to}} N(0, \Phi'(z)C\Phi'(z)^*)$$

Dimostrazione. Grazie al Lemma di Slutski 3.30 ii), si ha

$$Z_n - z = \frac{1}{\sqrt{n}}\sqrt{n}(Z_n - z) \underset{n\to\infty}{\overset{\mathscr{L}}{\to}} 0 \cdot Z = 0$$

Dunque, per la Proposizione 3.24, $Z_n \overset{\mathrm{P}}{\to} z$. Per il teorema della media, si può scrivere

$$\sqrt{n}(\Phi(Z_n) - \Phi(z)) = \sqrt{n}\Phi'(Z_n^*)(Z_n - z) \tag{3.13}$$

dove Z_n^* è un punto che si trova nel segmento che congiunge z a Z_n e, dunque, tale che $|Z_n^* - z| \le |Z_n - z|$. Ne segue che $|Z_n^* - z| \to 0$ in probabilità e in legge. Per l'Osservazione 3.15, $\Phi(Z_n^*) \overset{\mathscr{L}}{\to} \Phi(z)$. Dalla (3.13), applicando ancora il Lemma di Slutski e ricordando come si trasformano le leggi gaussiane rispetto ad una trasformazione lineare, si ha la tesi.

Esempio 3.32 Sia $(X_n)_n$ una successione di v.a. reali i.i.d. di media x e varianza σ^2 e indichiamo con $\bar{X}_n$ le medie empiriche. La successione delle v.a.

$$\sqrt{n}(\mathrm{e}^{\bar{X}_n} - \mathrm{e}^x)$$

converge in legge?

Osserviamo che, per il Teorema Limite Centrale,

$$\sqrt{n}(\bar{X}_n - x) = \frac{X_1 + \ldots + X_N - nx}{\sqrt{n}} \underset{n\to\infty}{\overset{\mathscr{L}}{\to}} N(0, \sigma^2)$$

Si può dunque applicare il metodo delta, da cui si ricava

$$\sqrt{n}(\mathrm{e}^{\bar{X}_n} - \mathrm{e}^x) \underset{n\to\infty}{\overset{\mathscr{L}}{\to}} N(0, \sigma^2\mathrm{e}^{2x})$$

Esercizi

E3.1 Sia $(X_n)_n$ una successione di v.a. i. i.d. su uno stesso spazio di probabilità $(\Omega, \mathcal{F}, \mathrm{P})$, tali che $0 < \mathrm{E}(X_1) < +\infty$. Per ogni $\omega \in \Omega$ consideriamo la serie di potenze

$$\sum_{n=1}^{\infty} X_n(\omega)x^n$$

e indichiamo con $R(\omega)$ il suo raggio di convergenza.

Ricordiamo che $R(\omega) = \big(\overline{\lim}_{n\to\infty} |X_n(\omega)|^{1/n}\big)^{-1}$.

a) Mostrare che R è una v.a. costante q.c.

b) Mostrare che esiste un numero $a > 0$ tale che $\mathrm{P}(|X_n| \geq a$ per infiniti indici $n) = 1$ e dedurre che $R \leq 1$ q.c.

c) Sia $b > 1$. Mostrare che $\sum_{n=1}^{\infty} \mathrm{P}(|X_n| \geq b^n) < +\infty$ e dedurre il valore di R q.c.

E3.2 Sia $(X_n)_n$ una successione di v.a. positive i.i.d. Poniamo

$$\bar\theta = \sup\{\theta \geq 0; \mathrm{E}(\mathrm{e}^{\theta X_1}) < +\infty\} \in \mathbb{R}^+ \cup \{+\infty\}.$$

a) Mostrare che $\mathrm{E}(\mathrm{e}^{\theta X_1}) < +\infty$ se $\theta < \bar\theta$ e $\mathrm{E}(\mathrm{e}^{\theta X_1}) = +\infty$ se $\theta > \bar\theta$. Mostrare la formula

$$\mathrm{E}(\mathrm{e}^{\theta X_1}) = \int_0^{+\infty} \mathrm{P}(X \geq \tfrac{1}{\theta}\log t)\, dt.$$

b) Quanto vale

$$\overline{\lim_{n\to\infty}} \frac{X_n}{\log n}\ ?$$

c) Sia $(X_n)_n$ una successione di v.a. i.i.d. di legge $N(0, 1)$. Quanto vale

$$\overline{\lim_{n\to\infty}} \frac{|X_n|}{\sqrt{\log n}}\ ?$$

E3.3 a) La successione di v.a. $(X_n)_n$ converge in probabilità alla v.a. X se e solo se da ogni sottosuccessione $(X_{n_k})_k$ si può estrarre una ulteriore sottosuccessione $(X_{n_{k_h}})_h$ tale che $X_{n_{k_h}} \xrightarrow{\mathrm{P}} X$ per $h \to \infty$.

b) Perché l'argomento precedente non si applica alla convergenza q.c.?

[a): si ricorda il seguente fatto elementare ma utilissimo: una successione reale $(a_n)_n$ converge verso un limite ℓ se e solo se da ogni sua sottosuccessione si può estrarre una ulteriore sottosuccessione convergente a ℓ.

b): attenzione agli eventi trascurabili ! Non sempre li si può trascurare.]

E3.4 (Teorema di Lebesgue per la convergenza in probabilità) Sia $(X_n)_n$ una successione di v.a. tutte maggiorate in modulo da una medesima v.a. integrabile Z e tale che $X_n \xrightarrow{P} X$. Allora

$$\lim_{n\to\infty} \mathrm{E}(X_n) = \mathrm{E}(X)\ .$$

[Vedi il suggerimento al punto a) dell'Esercizio 3.3.]

E3.5 Sia $(X_n)_n$ una successione di v.a. i.i.d. e aventi varianza finita σ^2. Poniamo, per ogni n, $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ e

$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X}_n)^2$$

Mostrare che la successione $(S_n^2)_n$ converge q.c. e determinarne il limite.

E3.6 Sia $(X_n)_n$ una successione di v.a. i.i.d. di varianza finita.

a) Consideriamo la v.a. X_1X_2. Quanto vale la sua media? Ha varianza finita?

b1) Poniamo

$$V_n = \frac{1}{n}(X_1X_2 + X_3X_4 + \ldots + X_{2n-1}X_{2n})$$

La succesione $(V_n)_n$ converge in probabilità? q.c.? Quanto vale il limite?

b2) E se fosse

$$V_n = \frac{1}{n}(X_1X_2 + X_2X_3 + \ldots + X_{2n-1}X_{2n})\,?$$

c) Supponiamo per di più che sia $\mathrm{E}(X_i^4) < +\infty$. Le successioni

$$W_n = \frac{1}{n}(X_1^4 + \ldots + X_n^4)$$
$$U_n = \frac{X_1^2 + \ldots + X_n^2}{X_1^4 + \ldots + X_n^4}$$

sono convergenti in probabilità? q.c.? A che limite?

E3.7 Sia $(X_n)_n$ una successione di v.a. indipendenti di Poisson di parametro λ e poniamo $\bar{X}_n = \frac{1}{n}(X_1 + \ldots + X_n)$.

a) Stimare con la disuguaglianza di Chebyshev la probabilità

$$\mathrm{P}(|\bar{X}_n - \lambda| \ge \eta) \tag{3.14}$$

b) Stimare la stessa quantità usando l'approssimazione normale.

c) Confrontare le due stime per $\lambda = 1$, $\eta = 10^{-2}$, $n = 10000$.

E3.8 Sia $(X_n)_n$ una successione di v.a. e supponiamo $X_n \sim \Gamma(n, \lambda)$.

a) Quanto vale $\mathrm{P}(X_1 > \frac{1}{\lambda})$? E $\mathrm{P}(X_3 > \frac{3}{\lambda})$?

b) Calcolare quanto vale approssimativamente

$$\mathrm{P}(\tfrac{1}{n}X_n > \tfrac{1}{\lambda})$$

per n grande.

E3.9 Sia $(X_n)_n$ una successione di v.a. indipendenti, tutte di legge uniforme sull'intervallo $[0, 2a]$.

a) Calcolare media e varianza delle X_i.

b) Calcolare, per $n \to \infty$ e per $x \in \mathbb{R}$ fissato, il limite della probabilità

$$\mathrm{P}(X_1 + \ldots + X_n > na + x\sqrt{n})$$

Quanto vale questo limite per $a = 2$, $x = 2$?

E3.10 Un calcolatore addiziona un milione di numeri e in ognuna di queste operazioni viene effettuato un errore di arrotondamento; supponiamo che i singoli errori siano tra loro indipendenti e abbiano distribuzione uniforme su $[-0.5 \cdot 10^{-10}, 0.5 \cdot 10^{-10}]$ (cioè supponiamo che la decima cifra decimale sia significativa). Qual è la probabilità che l'errore finale sia più piccolo in valore assoluto di $0.5 \cdot 10^{-7}$? (cioè qual è la probabilità che la settima cifra decimale sia significativa?) Qual è la probabilità che l'errore sia più piccolo in valore assoluto di $0.5 \cdot 10^{-8}$?

E3.11 Un dado equilibrato viene lanciato 900 volte e indichiamo con X il numero di volte in cui compare il 6.

a) Quanto vale $\mathrm{E}(X)$? Quanto vale $\mathrm{P}(X \geq 180)$?

b) Supponiamo di sapere dell'esistenza di una partita di dadi truccati che producono il 6 con probabilità $\frac{2}{9}$. Per decidere se un dado è di questi ultimi usiamo la procedura seguente: esso viene lanciato 900 volte e decidiamo che esso è truccato se si ottiene il 6 più ($\geq$) di 180 volte. Qual è la probabilità che un dado truccato venga effettivamente individuato?

E3.12 n carte numerate da 1 a n vengono girate successivamente. Diciamo che al tempo i si ha una *coincidenza* (*matching*, in inglese) se la i-esima carta girata è proprio la numero i. Indichiamo con X il numero totale di matching. Indichiamo con X_i la v.a. indicatrice dell'evento $A_i = \{\text{si ha un matching al tempo } i\}$.

a1) Qual è la legge della v.a. X_i?

a2) Quanto vale $\mathrm{E}(X)$?

a3) Quanto vale $\mathrm{P}(A_1 \cap A_2 \cap \ldots \cap A_k)$, $k \leq n$?

a4) Quanto vale $\mathrm{E}(X_i X_j)$? Quanto vale $\mathrm{Var}(X)$?

b) Indichiamo con G la funzione generatrice delle probabilità di X (che naturalmente dipende da n).

b1) Quanto vale $G'(1)$? E $G''(1)$?

b2) Mostrare che, per ogni $k \leq n$, si ha

$$X(X-1)\ldots(X-k+1) = \sum 1_{A_{i_1} \cap \ldots \cap A_{i_k}}$$

dove la somma viene fatta su tutte le k-uple di indici distinti $(i_1, \ldots, i_k) \in \{1, \ldots, n\}^k$.

b3) Mostrare che $G^{(k)} = 1$ per ogni $k \leq n$.

b3) Mostrare che, per $n \to \infty$, la legge di X converge a una legge notevole e determinarla.

E3.13 a) Consideriamo una v.a. reale Z di densità

$$f(t) = \frac{1}{2t}\, 1_{[\mathrm{e}^{-1},\mathrm{e}]}(t). \tag{3.15}$$

Calcolare la legge di $X = \log Z$.

b) Sia X una v.a. $N(0,1)$; mostrare che $\mathrm{E}(\mathrm{e}^{\theta X}) = \mathrm{e}^{-\theta^2/2}$. Sia Y una v.a. reale di legge $N(\mu,\sigma^2)$; calcolare la media e la varianza di e^Y. Calcolare la legge di e^Y (*legge lognormale di parametri μ e σ^2*).

c) Sia $(Z_n)_n$ una successione di v.a. indipendenti, tutte di legge data dalla densità (3.15). Mostrare che $(Z_1 \dots Z_n)^{1/\sqrt{n}}$ converge in legge ha una legge lognormale di parametri μ e σ^2 e calcolarli. Quanto vale il limite $\lim_{n\to\infty} \mathrm{E}[(Z_1 \dots Z_n)^{1/\sqrt{n}}]$?

E3.14 Un segnale consiste in una parola di n bit, ciascuno dei quali può assumere i valori 0 oppure 1. Nel corso della trasmissione ogni bit con probabilità $p = 0.01$ può essere distorto (cioè può essere mutato da 0 a 1 oppure da 1 a 0).

a) Qual è il numero medio di bit distorti? Qual è la probabilità che un segnale di 1000 bit contenga bit distorti? Qual è la probabilità che contenga almeno 10 bit distorti?

b) Per ridurre la distorsione si usa il seguente protocollo: ogni bit viene trasmesso tre volte ed il vero valore viene deciso a maggioranza: il bit viene posto uguale ad A ($A = 0$ oppure 1) se vi sono almeno due valori A tra quelli ricevuti. Qual è ora la probabilità che un singolo bit sia distorto? Qual è la probabilità che un segnale di 1000 bit contenga bit distorti?

E3.15 Nella trasmissione di un'immagine il colore di ogni pixel è descritto da 8 bit, cioè da un vettore $(a_1,\dots,a_8)$ dove $a_1,\dots,a_8$ possono essere 0 oppure 1. Durante la trasmissione di ogni singolo bit si può avere una distorsione con probabilità $p = 0.0002 = 2\cdot 10^{-4}$; cioè ogni bit trasmesso può venire alterato (da 0 a 1 o da 1 a 0) con probabilità $p = 2\cdot 10^{-4}$ e per di più indipendentemente da un bit all'altro.

a) Qual è la probabilità che un singolo pixel venga trasmesso correttamente?

b) Un'immagine è composta da $512 \times 256 = 131\,072$ pixel. Qual è il numero medio di pixel distorti in un'immagine? Qual è la probabilità che vi siano più ($\geq$) di 200 pixel distorti?

E3.16 Sia $(X_n)_n$ una successione di v.a., dove per ogni n $X_n \sim \chi^2(n)$. Qual è il comportamento della successione $(\frac{1}{n}X_n)_n$? Si può dire che converge in legge? In probabilità?

E3.17 Sia $(X_n)_n$ una successione di v.a. rispettivamente di legge geometrica di parametro $p_n = \frac{\lambda}{n}$. La successione $(\frac{1}{n}X_n)_n$ converge in legge? In caso affermativo, qual è la legge limite?

E3.18 Sia $(X_n)_n$ una successione di v.a. indipendenti tutte di legge di Poisson di parametro λ. Quanto vale il limite

$$\lim_{n\to\infty} \mathrm{P}(X_1 + \ldots + X_n \le n)\,?$$

al variare di $\lambda > 0$?

E3.19 Sia $(X_n)_n$ una successione di v.a. indipendenti tali che

$$\mathrm{P}(X_i > x) = \begin{cases} x^{-\lambda} & \text{se } x > 1 \\ 1 & \text{se } x \le 1 \end{cases}$$

dove λ è un numero > 1.

a) Calcolare media e varianza delle v.a. X_i.

b) Poniamo $Y_i = \log X_i$. Qual è la legge di Y_i?

c) Mostrare che la successione di v.a. $((X_1 X_2 \ldots X_n)^{1/n})_n$ converge q.c. e determinarne il limite.

E3.20 a) Sia $(X_n)_n$ una successione di v.a. tutte di legge normale e supponiamo che

$$\mathrm{E}(X_n) = b_n \overset{n\to\infty}{\to} b \qquad\qquad \mathrm{Var}(X_n) = \sigma_n^2 \overset{n\to\infty}{\to} \sigma^2$$

Mostrare che $X_n \overset{\mathscr{L}}{\to} N(b, \sigma^2)$ per $n \to \infty$.

b) Sia $(Z_n)_n$ una successione di v.a. indipendenti e $N(0, \sigma^2)$. Consideriamo la successione definita per ricorrenza da

$$X_0 = x \in \mathbb{R} \qquad X_{n+1} = \alpha X_n + Z_n$$

dove $\alpha < 1$ (cioè $X_1, \ldots, X_n, \ldots$ sono le posizioni successive di un mobile che ad ogni istante si sposta dalla posizione attuale X_n in αX_n ma subisce anche una perturbazione Z_n). Qual è la legge di X_1? E quella di X_2? Mostrare che, per $n \to \infty$, X_n converge in legge ad una v.a. di cui si preciserà la distribuzione. Se $\sigma^2 = 1, \alpha = 0.5$, quanto vale la probabilità che X_n disti dall'origine meno di 1 per n grande?

E3.21 Sia $X_1, X_2, \ldots$ una successione di v.a. indipendenti, tutte di legge uniforme su $[0, 1]$ e poniamo

$$Z_n = \min(X_1, \ldots, X_n)$$

a) La successione $(Z_n)_n$ converge in legge per $n \to \infty$? Converge in probabilità?

b) Mostrare che la successione $(n\, Z_n)_n$ converge in legge per $n \to \infty$ e determinare la legge limite. Dare un'approssimazione della probabilità

$$\mathrm{P}(\min(X_1, \ldots, X_n) \le \tfrac{2}{n})$$

per n grande.

E3.22 Sia $(X_n)_n$ una successione di v.a. indipendenti aventi la stessa legge, tutte di media 0 e varianza σ^2. Mostrare che la successione di v.a.

$$Z_n = \frac{(X_1 + \ldots + X_n)^2}{n}$$

converge in legge e determinarne il limite.

E3.23 Descrivere una procedura per simulare
a) Le leggi $\chi^2(n)$, $\Gamma(n, \lambda)$, $\Gamma(\frac{n}{2}, \lambda)$, $t(n)$, $F(n_1, n_2)$.
b) Una legge di Poisson di parametro λ.

E3.24 (Leggi Beta) Sappiamo che per ogni $\alpha, \beta > 0$ la funzione definita da

$$f(t) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\, t^{\alpha-1}(1-t)^{\beta-1} \qquad 0 \le t \le 1$$

e da $f(t) = 0$ se $t \notin [0, 1]$ è una densità di probabilità; essa si chiama legge Beta di parametri α e β e si indica con $\mathrm{Beta}(\alpha, \beta)$.

a) Mostrare che se $X \sim \mathrm{Beta}(\alpha, \beta)$ allora

$$\begin{aligned}
\mathrm{E}[X] &= \frac{\alpha}{\alpha+\beta}\\
\mathrm{E}[X^2] &= \frac{\alpha(\alpha+1)}{(\alpha+\beta)(\alpha+\beta+1)}\\
\mathrm{Var}(X) &= \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}
\end{aligned}$$

b) Mostrare che se $n \to \infty$ e $X_n \sim \mathrm{Beta}(n\alpha, n\beta)$ allora X_n converge in probabilità e determinare il limite.

c) Si sa a priori che una moneta dà testa con probabilità p ignota. Si saperò che p segue una legge $\beta(\alpha, \lambda)$. La moneta viene lanciata n volte. Qual è la probabilità di ottenere testa k volte? Qual è la legge condizionale di p sapendo che è stato ottenuto testa k volte? Calcolare media e varianza di questa legge condizionale e confrontarle con quelle di una $\beta(\alpha, \lambda)$. Qual è secondo voi una buona stima di p (sempre sapendo che in n lanci è stato ottenuto testa k volte)?

d) Come pensate che si possa simulare una legge $\mathrm{Beta}(\alpha, \beta)$?

E3.25 a) Sia $(\mu_n)_n$ la successione di probabilità su $\mathbb{R}$ data da

$$\mu_n = (1-\alpha_n)\delta_0 + \alpha_n\delta_n$$

dove $(\alpha_n)_n$ è una successione di numeri reali compresi tra 0 e 1. Mostrare che $(\mu_n)_n$ converge strettamente se e solo se $\lim_{n\to\infty} \alpha_n = 0$ e, in questo caso, calcolarne il limite.

b) Costruire un esempio di successione $(\mu_n)_n$ convergente strettamente ma tale che le medie e le varianza di μ_n non convergano alla media e alla varianza del limite.

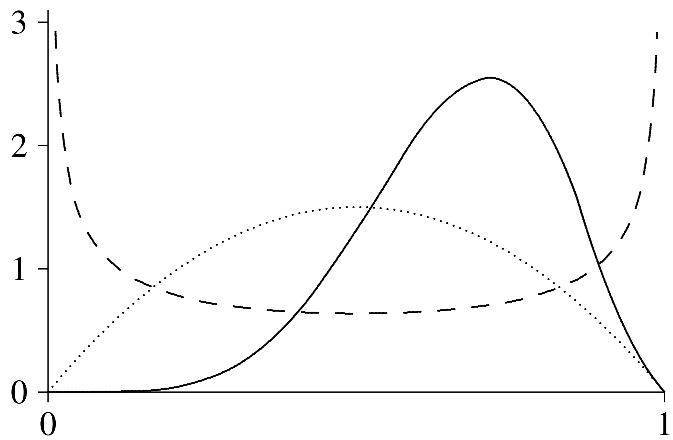

Figura 3.4 Grafico di densità beta per tre valori del parametro: $\beta(2,2)$ (puntini), $\beta(6,3)$ (tratto continuo) e $\beta(\frac{1}{2},\frac{1}{2})$ (trattini).

c) Sia $(\mu_n)_n$ una successione di probabilità su $\mathbb{R}$. Mostrare che se $\mu \to \mu$ strettamente allora

$$\underline{\lim}_{n\to\infty} \int x^2\, d\mu_n \ge \int x^2\, d\mu$$

E3.26 In questo esercizio vediamo che la convergenza in legge, con una ipotesi addizionale, implica la convergenza delle medie.

Siano X, X_n, $n \ge 1$ v.a. a valori $\mathbb{R}^m$.

a) Mostrare che, per ogni $p > 0$, $\|X\|_p \le \underline{\lim}_{n\to\infty} \|X_n\|_p$.

b) Supponiamo che esista una costante $M \in \mathbb{R}$ tale che $\|X_n\|_p \le M$.

b1) Mostrare che, per ogni $R > 0$, $\mathrm{P}(|X_n| > R) \le MR^{-p}$.

b2) Sia $\phi : \mathbb{R}^m \to \mathbb{R}$ una funzione tale che

- $0 \le \phi(x) \le 1$ per ogni $x \in \mathbb{R}^m$.
- $\phi(x) = 1$ per $|x| \le R$.
- $\phi(x) = 0$ per $|x| \ge R+1$.

Mostrare che

$$|\mathrm{E}(X_n) - \mathrm{E}(\phi(X_n))| \le \varepsilon$$
$$|\mathrm{E}(X) - \mathrm{E}(\phi(X))| \le \varepsilon$$

b3) Mostrare che $\lim_{n\to\infty} \mathrm{E}(X_n) = \mathrm{E}(X)$.

E3.27 a) Mostrare che se $X_n \sim \chi^2(n)$ allora

$$\frac{X_n - n}{\sqrt{2n}} \xrightarrow{\mathcal{L}} N(0,1)$$

b) (Approssimazione di Fisher) Mostrare che

$$\sqrt{2X_n} - \sqrt{2n-1} \xrightarrow{\mathcal{L}} N(0,1)$$

c) Derivare a partire sia da a) che da b) delle approssimazioni della funzione di ripartizione delle leggi $\chi^2(n)$. Utilizzarle per ricavare valori approssimati del quantile di ordine 0.95 di una v.a. $\chi^2(100)$ e confrontarli con il valore esatto 124.34. Quale delle due approssimazioni è migliore?

[a) Basta scrivere X_n come somma di n v.a. $\chi^2(1)$. b) Si usa il Lemma di Slutsky (Proposizione 3.30). c) a) dà

$$F_n(x) \sim \Phi\Big(\frac{x-n}{\sqrt{n}}\Big)$$

mentre b)

$$F_n(x) \sim \Phi(\sqrt{2x} - \sqrt{2n-1})$$

Tenendo conto che il quantile di ordine 0.95 di una $N(0,1)$ è 1.65, la prima approssimazione dà

$$x = 1.65 \cdot \sqrt{200} + 100 = 123.334$$

mentre la seconda

$$x = \tfrac{1}{2}\,(1.65 + \sqrt{199})^2 = 124.137$$

L'approssimazione di Fisher, trovata in b), resta migliore anche per valori di n più grandi. Ecco i valori dei quantili per alcuni valori di n e le loro approssimazioni.

	200	300	400	500
$\chi^2(n)$	233.99	341.40	447.63	553.13
$\frac{1}{2}\,(\phi_\alpha + \sqrt{2n-1})^2$	233.71	341.11	447.35	552.84
$\sqrt{2n}\,\phi_\alpha + n$	232.90	340.29	446.52	552.01

E3.28 (Teorema di Scheffè)Siano μ, μ_n, $n \ge 1$ misure di probabilità su uno spazio misurabile $(E, \mathscr{E})$ e supponiamo che esista una misura γ su $(E, \mathscr{E})$ tale che μ e le μ_n abbiano densità f e f_n rispettivamente rispetto a γ. Supponiamo che sia

$$\lim_{n\to\infty} f_n(x) = f(x)$$

tranne al più per un insieme di valori x di γ-misura nulla.

a) Mostrare che

$$\sup_{A\in\mathscr{E}} |\mu_n(A) - \mu(A)| = \frac{1}{2}\int_E |f_n - f|\,d\gamma \to 0$$

b) Mostrare che, se per di più E è uno spazio topologico, allora $\mu_n \to \mu$ strettamente.

c) Mostrare, con un esempio, che si può avere $\mu_n \to \mu$ strettamente senza che si abbia convergenza delle densità.

[a): se $\alpha_n = f - f_n$, allora $\int \alpha_n\,d\gamma = 0$ e se $A \in \mathscr{E}$

$$\int_A \alpha_n\,d\gamma = -\int_{A^C} \alpha_n\,d\gamma \qquad]$$

E3.29 a) Siano $(\mu_n)_n$ e $(\nu_n)_n$ successioni di probabilità su $\mathbb{R}^d$, convergenti strettamente alle probabilità μ e ν rispettivamente. Allora

$$\mu_n \otimes \nu_n \to \mu \otimes \nu$$
$$\mu_n * \nu_n \to \mu * \nu$$

b) Se ν_σ indica una probabilità $N(0, \sigma^2)$, mostrare che $\mu * \nu_\sigma \overset{\sigma\to 0}{\to} \mu$

E3.30 a) Sia $(X_n)_n$ una successione di v.a. reali tutte di legge normale e supponiamo che $X_n \overset{L^2}{\to} X$. Mostrare che anche X ha legge normale.

b) Sia $(\Omega, \mathscr{A}, \mathrm{P})$ uno spazio di probabilità. Mostrare che l'insieme delle v.a. reali di legge normale definite su $(\Omega, \mathscr{A}, \mathrm{P})$ costituisce un chiuso di $L^2(\Omega, \mathscr{A}, \mathrm{P})$. Costituiscono anche uno spazio vettoriale?

c) (Più difficile) Mostrare che l'affermazione del punto a) resta vera anche se la convergenza ha luogo solo in legge.

[a): se $X_n \sim N(m_n, \sigma_n^2)$ allora

$$\phi_{X_n}(\theta) = \mathrm{e}^{i\langle\theta, m_n\rangle}\mathrm{e}^{-\langle\sigma_n^2\theta,\theta\rangle}$$

e la convergenza in L^2 implica la convergenza della media e della varianza.

b) implica a) ma è più difficile. Si suppone prima che le X_n siano centrate e si mostra che le varianze devono restare limitate. Se $(Z_n)_n$ sono v.a. indipendenti dalle $(X_n)_n$ ma tali che $Z_n \sim X_n$ allora $(X_n - Z_n)_n$ converge in legge, è centrata e dunque le varianze di $X_n - Z_n$ devono restare limitate. Poiché la varianza di $X_n - Z_n$ è due volte la varianza di X_n, si ha che se $\mu_{X_n} \to \mu$ necessariamente le varianze σ_n^2 devono restare limitate. Si può dimostrare ora che anche le medie b_n sono limitate. Riprendendo l'argomento di a) si vede che per ogni valore di aderenza σ^2 e b di $(\sigma_n^2)_n$ e $(b_n)_n$ rispettivamente vi è una sottosuccessione di $(\mu_{X_n})_n$ che converge a una legge $N(b, \sigma^2)$. Poiché $(\mu_{X_n})_n$ converge è chiaro che $b_n \to b$, $\sigma_n^2 \to \sigma^2$ e $\mu_{X_n} \to N(b, \sigma^2)$.]

E3.31 (Dimostrazione del Teorema 2.23, d'inversione) Indichiamo con ν_σ una legge $N(0, \sigma^2)$ e con γ_σ la sua densità. Allora

a) Mostrare che

$$\gamma_\sigma(x) = \frac{1}{2\pi}\int \hat{\nu}_\sigma(\theta)\mathrm{e}^{-i\theta x}\, d\theta$$

b) Sia μ una probabilità su $\mathbb{R}$. Mostrare che $\mu * \nu_\sigma$ ha densità rispetto alla misura di Lebesgue data da

$$f_\sigma(x) = \frac{1}{2\pi}\int \mathrm{e}^{-\frac{1}{2}\sigma^2\theta^2}\hat{\mu}(\theta)\mathrm{e}^{-i\theta x}\, d\theta$$

c) Supponiamo $\hat{\mu} \in L^1$. Mostrare che f_σ converge puntualmente verso

$$f(x) = \frac{1}{2\pi}\int \hat{\mu}(\theta)\mathrm{e}^{-i\theta x}\, d\theta \tag{3.16}$$

d) Mostrare che, se $\hat{\mu}$ è integrabile, allora μ ha densità rispetto alla misura di Lebesgue data dalla funzione f in (3.16).

E3.32 (Convergenza delle leggi empiriche) Sia $(X_n)_n$ una successione di v.a. indipendenti, tutte di legge μ a valori nello spazio metrico E, localmente compatto e separabile. Per ogni $\omega \in \Omega$ e per ogni n consideriamo la misura su E $Z_n = \frac{1}{n}\sum_{i=1}^n \delta_{X_i}$.

a) Mostrare che per ogni n Z_n è una v.a. a valori in $\mathcal{M}_1(E)$, munito della σ-algebra di Borel della topologia della convergenza stretta (vedi l'esercizio precedente).

b) Mostrare che $Z_n \xrightarrow{\mathcal{L}} \delta_\mu$.

[Si tratta di provare che $\nu_n(f) \to \mu(f)$ per ogni ω tranne al più per un insieme di ω trascurabile (le probabilità ν_n dipendono da ω). Ma per ogni $f \in \mathcal{C}_b$ si ha

$$\nu_n(f) = \frac{1}{n}\sum_{i=1}^n f(X_i) \to \mu(f) \tag{3.17}$$

q.c. per la legge forte di Rajchman. Attenzione però, nella $(*)$ l'insieme di misura nulla dipende dalla funzione f: per terminare occorre mostrare l'esistenza di un insieme di misura nulla tale che (3.17) valga qualunque sia f; usare l'Osservazione 3.14.]

4
Problemi

In questo capitolo sviluppiamo alcuni esercizi che sono più articolati di quelli proposti nei capitoli precedenti, anche se comunque non fanno che applicare la teoria sviluppata finora. Si consiglia al lettore di cimentarsi comunque, magari in gruppo, prima di ricorrere alla soluzione, che viene fornita nell'ultimo paragrafo.

4.1 Problemi al capitolo 1

Problema 4.1 Sia μ una misura *finita* sullo spazio misurabile $(E, \mathscr{E})$.

a1) Mostrare che, se $0 \le p \le q$, allora $|x|^p \le 1 + |x|^q$ per ogni $x \in \mathbb{R}$.

a2) Mostrare che, se $f \in L^q$,

$$\lim_{p \to q^-} \|f\|_p = \|f\|_q \tag{4.1}$$

a3) Mostrare che $\lim_{p \to q^+} \|f\|_p \ge \|f\|_q$.

a4) Mostrare che si ha sempre (cioè anche se $f \notin L^q$) $\lim_{p \to q^-} \int |f|^p 1_{\{|f| \ge 1\}} \, d\mu = \int |f|^q 1_{\{|f| \ge 1\}} \, d\mu$. Mostrare che la (4.1) vale anche senza l'ipotesi $f \in L^q$.

a5) Costruire un esempio di funzione che appartiene a L^q per un dato valore di q, ma che non appartiene a L^p per ogni $p > q$. Mostrare che in generale non si ha $\lim_{p \to q^+} \|f\|_p \ge \|f\|_q$.

b1) Sia $f : E \to \mathbb{R}$ una funzione misurabile. Mostrare che

$$\overline{\lim_{p \to +\infty}} \|f\|_p \le \|f\|_\infty$$

b2) Sia $M \ge 0$. Mostrare che, per ogni $p \ge 0$,

$$\int |f|^p \, d\mu \ge M^p \mu(|f| \ge M)$$

b3) Quanto vale $\lim_{p\to+\infty} \|f\|_p$?

4.2 Problemi al capitolo 2

Problema 4.1 Siano $X_1, \dots, X_n$ v.a. indipendenti di legge esponenziale di parametro λ e poniamo

$$Z_n = \max(X_1, \dots, X_n) \ .$$

a) Mostrare che la v.a. Z_n ha una legge data da una densità rispetto alla misura di Lebesgue e calcolarla. Quanto vale la media di Z_2 ? E di Z_3 ?

b) Mostrare che la trasformata di Laplace di Z_n vale

$$n\Gamma(n)\frac{\Gamma(1-\frac{\theta}{\lambda})}{\Gamma(n+1-\frac{\theta}{\lambda})}$$

c) Dimostrare che per la funzione

$$\alpha \to \frac{\Gamma'(\alpha)}{\Gamma(\alpha)}$$

(è la derivata del logaritmo della funzione Γ) vale la relazione

$$\frac{\Gamma'(\alpha+1)}{\Gamma(\alpha+1)} = \frac{1}{\alpha} + \frac{\Gamma'(\alpha)}{\Gamma(\alpha)} \ . \tag{4.2}$$

Quanto vale $\mathrm{E}(Z_n)$?

Problema 4.2 a1) Mostrare che

$$\int_{-\infty+i\pi}^{+\infty+i\pi} \frac{\mathrm{e}^{i\theta z}}{\cosh z}\,dz = -\mathrm{e}^{-\theta\pi}\int_{-\infty}^{+\infty} \frac{\mathrm{e}^{i\theta x}}{\cosh x}\,dx \ .$$

a2) Usando il metodo dei residui e il contorno della Figura 4.1 calcolare l'integrale

$$\int_{-\infty}^{+\infty} \frac{\mathrm{e}^{i\theta x}}{\cosh x}\,dx \ .$$

Determinare la costante c tale che la funzione

$$x \to \frac{c}{\cosh x} \tag{4.3}$$

sia una densità di probabilità.

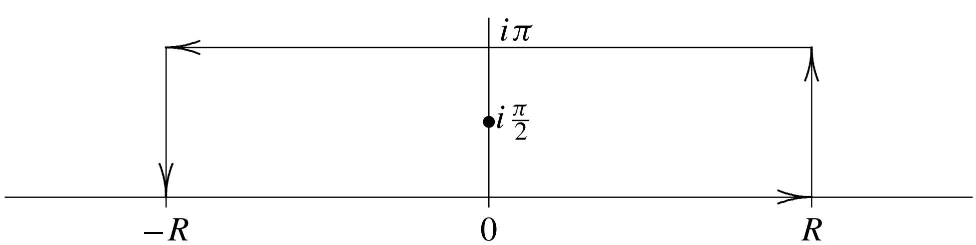

Figura 4.1

b1) Mostrare che

$$\frac{1}{\cosh^2 z} = -\frac{1}{(z-i\frac{\pi}{2})^2}\Big(1-\tfrac{1}{3}(z-i\tfrac{\pi}{2})^2+o((z-i\tfrac{\pi}{2})^2)\Big).$$

b2) Calcolare

$$\int_{-\infty}^{+\infty}\frac{e^{i\theta x}}{\cosh^2 x}\,dx\ .$$

Determinare una costante c_2 in modo che la funzione

$$x \to \frac{c_2}{\cosh^2 x} \tag{4.4}$$

sia una densità di probabilità.

c) Siano X, Y v.a. indipendenti, entrambe di legge data da (4.3). Qual è la funzione caratteristica di $X+Y$? E la densità?

d) Mostrare che

$$\frac{2}{\pi^2}\,\frac{x}{\sinh x}$$

è una densità di probabilità.

4.3 Problemi al capitolo 3

Problema 4.1 Sia $(X_n)_n$ una successione di v.a. i. i.d. su uno stesso spazio di probabilità $(\Omega, \mathscr{F}, \mathrm{P})$, tali che $0 < \mathrm{E}(X_1) < +\infty$. Per ogni $\omega \in \Omega$ consideriamo la serie di potenze

$$\sum_{n=1}^{\infty} X_n(\omega)x^n$$

e indichiamo con $R(\omega)$ il suo raggio di convergenza.

Ricordiamo che $R(\omega) = \big(\overline{\lim}_{n\to\infty}|X_n(\omega)|^{1/n}\big)^{-1}$.

a) Mostrare che R è una v.a. costante q.c.

b) Mostrare che esiste un numero $a > 0$ tale che $\mathrm{P}(|X_n| \ge a$ per infiniti indici $n) = 1$ e dedurre che $R \le 1$ q.c.

c) Sia $b > 1$. Mostrare che $\sum_{n=1}^{\infty}\mathrm{P}(|X_n| \ge b^n) < +\infty$ e dedurre il valore di R q.c.

Problema 4.2 Sia $(X_n)_n$ una successione di v.a. positive i.i.d. Poniamo

$$\bar\theta = \sup\{\theta \ge 0; \mathrm{E}(e^{\theta X_1}) < +\infty\} \in \mathbb{R}^+ \cup \{+\infty\}.$$

a) Mostrare che $\mathrm{E}(e^{\theta X_1}) < +\infty$ se $\theta < \bar\theta$ e $\mathrm{E}(e^{\theta X_1}) = +\infty$ se $\theta > \bar\theta$. Mostrare la formula

$$\mathrm{E}(e^{\theta X_1}) = \int_0^{+\infty}\mathrm{P}(X \ge \tfrac{1}{\theta}\log t)\,dt.$$

b) Quanto vale

$$\varlimsup_{n\to\infty} \frac{X_n}{\log n}\,?$$

c) Sia $(X_n)_n$ una successione di v.a. i.i.d. di legge $N(0,1)$. Quanto vale

$$\varlimsup_{n\to\infty} \frac{|X_n|}{\sqrt{\log n}}\,?$$

Problema 4.3 n carte numerate da 1 a n vengono girate successivamente. Diciamo che al tempo i si ha una *coincidenza* (*matching*, in inglese) se la i-esima carta girata è proprio la numero i. Indichiamo con X il numero totale di matching. Indichiamo con X_i la v.a. indicatrice dell'evento $A_i = \{\text{si ha un matching al tempo } i\}$.

a1) Qual è la legge della v.a. X_i ?

a2) Quanto vale $\mathrm{E}(X)$?

a3) Quanto vale $\mathrm{P}(A_1 \cap A_2 \cap \ldots \cap A_k)$, $k \le n$?

a4) Quanto vale $\mathrm{E}(X_i X_j)$? Quanto vale $\mathrm{Var}(X)$?

b) Indichiamo con G la funzione generatrice delle probabilità di X (che naturalmente dipende da n).

b1) Quanto vale $G'(1)$? E $G''(1)$?

b2) Mostrare che, per ogni $k \le n$, si ha

$$X(X-1)\ldots(X-k+1) = \sum 1_{A_{i_1}\cap\ldots\cap A_{i_k}} \tag{4.5}$$

dove la somma viene fatta su tutte le k-uple *ordinate* di indici distinti $(i_1,\ldots,i_k) \in \{1,\ldots,n\}^k$.

b3) Mostrare che $G^{(k)} = 1$ per ogni $k \le n$. Quanto vale $G^{(k)}(1)$ per $k > n$?

b3) Mostrare che, per $n \to \infty$, la legge di X converge a una legge nota e determinarla.

4.4 Soluzioni

S4.1 a1) Se $|x| \le 1$, allora $|x|^p \le 1$, se invece $|x| \ge 1$, allora $|x|^p \le |x|^q$. Dunque, in ogni caso, $|x|^p \le 1 + |x|^q$.

a2) Se $p \uparrow q$, allora $|f|^p \to |f|^q$. Inoltre, per a1), $|f|^p \le 1 + |f|^q$. Poiché $|f|^q$ è integrabile, così come pure la funzione 1, poiché la misura è supposta finita, si può applicare il teorema di Lebsgue, per cui

$$\lim_{p\to q^-} \int |f|^p \, d\mu = \int |f|^q \, d\mu$$

a3) Se $p \downarrow q$, allora $|f|^p \to |f|^q$ e basta applicare il lemma di Fatou.

a4) Si ha $f^p 1_{\{|f|\ge 1\}} \uparrow f^q 1_{\{|f|\ge 1\}}$ se $p \uparrow q$. Basta quindi applicare il teorema di Beppo Levi. Si ha quindi

$$\lim_{p\to q^-} \int |f|^p \, d\mu = \lim_{p\to q^-} \int |f|^p 1_{\{|f|\ge 1\}} \, d\mu + \lim_{p\to q^-} \int |f|^p 1_{\{|f|<1\}} \, d\mu =$$
$$= \int |f|^q 1_{\{|f|\ge 1\}} \, d\mu + \int |f|^q 1_{\{|f|<1\}} \, d\mu = \int |f|^q \, d\mu,$$

dove, per passare al limite per entrambi gli integrali si è usato il teorema di Beppo Levi (per il second0 si può anche usare il teorema di Lebesgue).

a5) La funzione

$$f(x) = \frac{1}{x \log^2 x} \, 1_{[0,\frac{1}{2}]}(x)$$

è integrabile (la primitiva di $(x \log^2 x)^{-1}$ è $(-\log x)^{-1}$). Ma $|f|^p$ non è integrabile, per ogni $p > 1$. Per questa funzione, dunque, $\|f\|_1 < +\infty$, mentre $\lim_{p\to 1+} \|f\|_p = +\infty$.

b1) Si ha, q.o., $|f|^p \le \|f\|_\infty^p$. Dunque

$$\int |f|^p \, d\mu \le \|f\|_\infty^p \mu(E).$$

Dunque

$$\overline{\lim_{p\to+\infty}} \|f\|_p \le \|f\|_\infty \lim_{p\to+\infty} \mu(E)^{1/p} = \|f\|_\infty$$

b2) Si ha $|f|^p \ge |f|^p 1_{\{|f|\ge M\}} \ge M^p 1_{\{|f|\ge M\}}$. Dunque

$$\int |f|^p \, d\mu \ge \int M^p 1_{\{|f|\ge M\}} \, d\mu = M^p \mu(|f| \ge M).$$

b3) Per definizione, se $M < \|f\|_\infty$, $\mu(|f| \ge M) > 0$. Dunque, per ogni $M < \|f\|_\infty$ e grazie a c2),

$$\|f\|_p \ge M \mu(|f| \ge M)^{1/p}$$

da cui $\underline{\lim}_{p\to+\infty} \|f\|_p \ge M$ e, per l'arbitrarietà di M,

$$\underline{\lim_{p\to+\infty}} \|f\|_p \ge \|f\|_\infty \, .$$

Si conclude combinando con b1).

S4.1 a) Il metodo della funzione di ripartizione dà immediatamente, per $x \ge 0$,

$$F_n(x) = \mathrm{P}(Z_n \le x) = \mathrm{P}(X_1 \le x, \dots, X_n \le x) = \mathrm{P}(X_1 \le x)^n = (1 - \mathrm{e}^{-\lambda x})^n$$

Derivando si trova la densità

$$f_n(x) = n\lambda \mathrm{e}^{-\lambda x}(1 - \mathrm{e}^{-\lambda x})^{n-1}$$

per $x \geq 0$ e $f_n(x) = 0$ per $x < 0$. Ricordando, dall'espressione della media delle leggi esponenziali, che $\int_{-\infty}^{+\infty} x\mathrm{e}^{-\lambda x}\,dx = \lambda^{-2}$,

$$\begin{aligned}\mathrm{E}(Z_2) = 2\lambda \int_{-\infty}^{+\infty} x\mathrm{e}^{-\lambda x}(1-\mathrm{e}^{-\lambda x})\,dx = 2\lambda \int_{-\infty}^{+\infty} x\mathrm{e}^{-\lambda x} 1 - x\mathrm{e}^{-2\lambda x})\,dx = \\ = 2\lambda\Big(\frac{1}{\lambda^2} - \frac{1}{4\lambda^2}\Big) = \frac{1}{\lambda}\frac{3}{2}\cdot\end{aligned}$$

Inoltre

$$\begin{aligned}\mathrm{E}(Z_3) = 3\lambda \int_{-\infty}^{+\infty} x\mathrm{e}^{-\lambda x}(1-\mathrm{e}^{-\lambda x})^2\,dx = 3\lambda \int_{-\infty}^{+\infty} x\mathrm{e}^{-\lambda x} - 2x\mathrm{e}^{-2\lambda x} + x\mathrm{e}^{-3\lambda x}\,dx = \\ = 3\lambda\Big(\frac{1}{\lambda^2} - \frac{2}{4\lambda^2} + \frac{1}{9\lambda^2}\Big) = \frac{1}{\lambda}\frac{11}{6}\cdot\end{aligned}$$

b) Si ha, per $z \in \mathbb{R}$,

$$\mathrm{E}(\mathrm{e}^{zZ_n}) = n\lambda \int_{-\infty}^{+\infty} \mathrm{e}^{zx}\mathrm{e}^{-\lambda x}(1-\mathrm{e}^{-\lambda x})^{n-1}\,dx$$

Questo integrale chiaramente diverge se $z \geq \lambda$. Altrimenti poniamo $\mathrm{e}^{-\lambda x} = t$, quindi $-\lambda\mathrm{e}^{-\lambda x} = dt$, $\mathrm{e}^{zx} = t^{-z/\lambda}$ e, ricordando la relazione

$$\int_0^1 t^{\alpha-1}(1-t)^{\beta-1}\,dt = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

(vedi l'espressione delle leggi Beta),

$$\psi_n(z) = \mathrm{E}(\mathrm{e}^{zZ_n}) = n\int_0^1 t^{-z/\lambda}(1-t)^{n-1}\,dt = n\Gamma(n)\frac{\Gamma(1-\frac{z}{\lambda})}{\Gamma(n+1-\frac{z}{\lambda})}\cdot$$

c) Dalla relazione fondamentale della funzione Γ,

$$\Gamma(\alpha+1) = \alpha\Gamma(\alpha), \tag{4.6}$$

derivando si trova $\Gamma'(\alpha+1) = \Gamma(\alpha) + \alpha\Gamma'(\alpha)$ e dividendo per $\Gamma(\alpha+1)$ ambo i membri e usando la (4.6), si trova la (4.2).

Per calcolare la media di Z_n calcoliamo la derivata della trasformata di Laplace all'origine. Si ha

$$\begin{aligned}\psi'_n(z) = n\Gamma(n)\frac{-\frac{1}{\lambda}\Gamma(n+1-\frac{z}{\lambda})\Gamma'(1-\frac{z}{\lambda}) + \frac{1}{\lambda}\Gamma'(n+1-\frac{z}{\lambda})\Gamma(1-\frac{z}{\lambda})}{\Gamma(n+1-\frac{z}{\lambda})^2} = \\ = -\frac{n\Gamma(n)}{\lambda\Gamma(n+1-\frac{z}{\lambda})}\Big(\frac{\Gamma'(n+1-\frac{z}{\lambda})\Gamma(1-\frac{z}{\lambda})}{\Gamma(n+1-\frac{z}{\lambda})} - \Gamma'(1-\tfrac{z}{\lambda})\Big).\end{aligned}$$

Per $z = 0$,

$$\psi'_n(0) = \frac{n\Gamma(n)}{\lambda\Gamma(n+1)}\Big(\frac{\Gamma'(n+1)}{\Gamma(n+1)} - \Gamma'(1)\Big). \tag{4.7}$$

Ricordiamo che $n\Gamma(n) = \Gamma(n+1)$, mentre, per (4.2),

$$\frac{\Gamma'(n+1)}{\Gamma(n+1)} = \frac{1}{n} + \frac{\Gamma'(n)}{\Gamma(n)} = \ldots = \frac{1}{n} + \frac{1}{n-1} + \ldots + 1 + \Gamma'(1)$$

e quindi sosituendo nella (4.7),

$$\mathrm{E}(Z_n) = \frac{1}{\lambda}\Big(1 + \frac{1}{2} + \ldots + \frac{1}{n}\Big).$$

S4.2 a1) Si ha $\mathrm{e}^{i\pi} = \mathrm{e}^{-i\pi} = -1$ e quindi

$$\cosh(x + i\pi) = \frac{\mathrm{e}^{x+i\pi} + \mathrm{e}^{-x-i\pi}}{2} = -\frac{\mathrm{e}^{x} + \mathrm{e}^{-x}}{2} = -\cosh x\ .$$

Dunque

$$\frac{\mathrm{e}^{i\theta(x+i\pi)}}{\cosh(x+i\pi)} = -\mathrm{e}^{-\theta\pi}\ \frac{\mathrm{e}^{i\theta x}}{\cosh x}\cdot$$

a2) Nel semipiano superiore la funzione cosh si annulla per $z = i\frac{\pi}{2}$. Calcoliamo il residuo, per $z = i\frac{\pi}{2}$ della funzione $z \to \mathrm{e}^{i\theta z}(\cosh z)^{-1}$. Poiché $\cosh' = \sinh$ e $\sinh i\frac{\pi}{2} = i$, $\cosh z = i(z - i\frac{\pi}{2}) + o(z - i\frac{\pi}{2})$. Dunque

$$\lim_{z\to\frac{\pi}{2}} \frac{\mathrm{e}^{i\theta z}}{\cosh z}\,(z - i\tfrac{\pi}{2}) = \mathrm{e}^{-\theta\frac{\pi}{2}}\ \frac{1}{i}\cdot$$

Se integriamo sul contorno della Figura 4.1, i contributi dell'integrale sui lati corti tendono a 0 per $R \to +\infty$ perché

$$|\cosh z| = \frac{1}{2}\,(|\mathrm{e}^{z} + \mathrm{e}^{-z}|) \geq \frac{1}{2}\,(|\mathrm{e}^{z}| - |\mathrm{e}^{-z}|) \geq \frac{1}{2}\,(\mathrm{e}^{R} - 1)$$

e dunque l'integrale sul lato corto è $\leq \frac{R}{2}(\mathrm{e}^{R} - 1)$. Dunque, al limite per $R \to +\infty$, usando a1) e la formula dei residui,

$$2\pi i \cdot \frac{1}{i} \cdot \mathrm{e}^{-\frac{1}{2}\theta\pi} = \int_{-\infty}^{+\infty} \frac{\mathrm{e}^{i\theta x}}{\cosh x}\,dx - \int_{i\pi-\infty}^{i\pi+\infty} \frac{\mathrm{e}^{i\theta z}}{\cosh z}\,dz = (1 + \mathrm{e}^{-\theta\pi})\int_{-\infty}^{+\infty} \frac{\mathrm{e}^{i\theta x}}{\cosh x}\,dx$$

Dunque

$$\int_{-\infty}^{+\infty} \frac{\mathrm{e}^{i\theta x}}{\cosh x}\,dx = \frac{2\pi}{\mathrm{e}^{\frac{1}{2}\theta\pi}(1 + \mathrm{e}^{-\theta\pi})} = \frac{\pi}{\cosh(\theta\frac{\pi}{2})}\cdot$$

Poiché questa funzione vale π per $\theta = 0$, ne segue che

$$\frac{1}{\pi \cosh x}$$

è una densità di probabilità e che la sua funzione caratteristica è

$$\phi_1(\theta) = \frac{1}{\cosh(\theta \frac{\pi}{2})}\cdot$$

b) Con un po' di pazienza si trova successivamente

$$\begin{aligned}
\cosh z &= i(z - i\tfrac{\pi}{2}) + \tfrac{i}{6}(z - i\tfrac{\pi}{2})^3 + o((z - i\tfrac{\pi}{2})^3)\\
\cosh^2 z &= -(z - i\tfrac{\pi}{2})^2 - \tfrac{1}{3}(z - i\tfrac{\pi}{2})^4 + o((z - i\tfrac{\pi}{2})^4)\\
\frac{1}{\cosh^2 z} &= -\frac{1}{(z - i\frac{\pi}{2})^2}\, \frac{1}{1 + \frac{1}{3}(z - i\frac{\pi}{2})^2 + o((z - i\frac{\pi}{2})^2)} =\\
&= -\frac{1}{(z - i\frac{\pi}{2})^2}\Big(1 - \tfrac{1}{3}(z - i\tfrac{\pi}{2})^2 + o((z - i\tfrac{\pi}{2})^2)\Big).
\end{aligned}$$

Poiché, sviluppando in $z = i\frac{\pi}{2}$,

$$e^{i\theta z} = e^{-\theta\frac{\pi}{2}} + i\theta e^{-\theta\frac{\pi}{2}}(z - i\tfrac{\pi}{2}) + o(z - i\tfrac{\pi}{2})$$

moltiplicando i due sviluppi si trova che il residuo in $z = i\frac{\pi}{2}$ vale

$$-i\theta e^{-\theta\frac{\pi}{2}}.$$

Dunque

$$\begin{aligned}
2\pi i(-i\theta e^{-\theta\frac{\pi}{2}}) &= \int_{-\infty}^{+\infty} \frac{e^{i\theta x}}{\cosh^2 x}\, dx + \int_{-\infty+i\pi}^{+\infty+i\pi} \frac{e^{i\theta z}}{\cosh^2 z}\, dz =\\
&= (1 - e^{-\theta\pi})\int_{-\infty}^{+\infty} \frac{e^{i\theta x}}{\cosh^2 x}\, dx
\end{aligned}$$

ovvero, se $\theta \neq 0$,

$$\int_{-\infty}^{+\infty} \frac{e^{i\theta x}}{\cosh^2 x}\, dx = \frac{2\pi\theta e^{-\theta\frac{\pi}{2}}}{1 - e^{-\theta\pi}} = \frac{\pi\theta}{\sinh(\theta\frac{\pi}{2})}\cdot \tag{4.8}$$

Poiché $\sinh z \sim z$ per $z \to 0$, si ha, mandando $\theta \to 0$ nella relazione precedente,

$$\int_{-\infty}^{+\infty} \frac{1}{\cosh^2 x}\, dx = 2$$

e dunque

$$f(x) = \frac{1}{2\cosh^2 x}$$

è una densità di probabilità.

c) La funzione caratteristica di $X + Y$ è $(\cosh(\theta\frac{\pi}{2}))^{-2}$. Poiché si tratta di una funzione integrabile, per il Teorema 2.23 d'inversione, essa ha densità data da

$$g(x) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \frac{e^{-\theta x}}{\cosh^2(\theta\frac{\pi}{2})}\, d\theta = \frac{1}{\pi^2}\int_{-\infty}^{+\infty} \frac{e^{-\frac{2x}{\pi}y}}{\cosh^2 y}\, dy\ .$$

Grazie a (4.8),

$$g(x) = \frac{2}{\pi^2}\,\frac{x}{\sinh x}\cdot \tag{4.9}$$

d) Conseguenza di (4.9).

S4.1 a) R è $\mathcal{B}^\infty$-misurabile poiché

$$\{R \ge b\} = \bigcap_{q\in\mathbb{Q},|q|<b} \Big\{\sum\nolimits_n |X_n||q|^n < +\infty\Big\}$$

e quindi per il Lemma 0-1 è q.c. costante.

b) Sia $a > 0$ tale che $P(|X_n| \ge a) > 0$, allora $\sum_n P(|X_n| \ge a) = \sum_n P(|X_1| \ge a) = +\infty$. Quindi per il lemma di Borel-Cantelli

$$\mathrm{P}(\overline{\lim_{n\to\infty}}\{|X|_n \ge a\}) = 1,$$

e quindi

$$\mathrm{P}(\overline{\lim_{n\to\infty}}\, |X_n|^{1/n} \ge 1) = 1,$$

cioè $\mathrm{P}(R \le 1) = 1$.

c) per ogni $b > 1$, per la disuguaglianza di Markov $\sum_{n=1}^\infty \mathrm{P}(|X_n| \ge b^n)$ è maggiorata da una serie geometrica convergente, quindi per il lemma di Borel-Cantelli

$$\mathrm{P}\Big(\overline{\lim_{n\to\infty}}\{|X_n| \ge b^n\}\Big) = 0$$

cioè

$$\mathrm{P}(|X_n|^{1/n} < b \text{ definitivamente}) = 1$$

e quindi

$$\mathrm{P}\Big(\overline{\lim_{n\to\infty}}\, |X_n|^{1/n} \le 1\Big) = 1$$

cioè $\mathrm{P}(R \geq 1) = 1$.

S4.2 Per definizione di estremo superiore, $\theta < \bar{\theta}$ implica che esiste δ tale che $\mathrm{E}(\mathrm{e}^{\theta X_1}) \leq \mathrm{E}(\mathrm{e}^{\delta X_1}) < +\infty$, mentre $\theta > \bar{\theta}$ implica $\mathrm{E}(\mathrm{e}^{\delta X_1}) = +\infty$. Inoltre

$$\mathrm{E}(\mathrm{e}^{\theta X_1}) = \int_0^{+\infty} \mathrm{P}(\mathrm{e}^{\theta X_1} > t)\, dt = \int_0^{+\infty} \mathrm{P}(X_1 \geq \tfrac{1}{\theta} \log t)\, dt$$

b) Sia $\delta > \frac{1}{\bar{\theta}}$. Studiamo il comportamento della serie

$$\sum_{n=1}^{\infty} \mathrm{P}(X_n > \delta \log n) \tag{4.10}$$

Se $\frac{1}{\delta} < \theta < \bar{\theta}$, per la disuguaglianza di Markov,

$$\mathrm{P}(X_n > \delta \log n) = \mathrm{P}(\theta X_n > \theta\delta \log n) = \mathrm{P}(\mathrm{e}^{\theta X_n} > n^{\theta\delta}) \leq \frac{\mathrm{E}(\mathrm{e}^{\theta X_n})}{n^{\delta\theta}}\cdot$$

che, poiché $\theta\delta > 1$, è il termine generale di una serie convergente.

Se invece $\delta < \frac{1}{\bar{\theta}}$ la serie diverge. Infatti

$$\sum_{n=1}^{\infty} \mathrm{P}(X_n > \delta \log n) \geq \sum_{n=1}^{\infty} \int_n^{n+1} \mathrm{P}(X_1 > \delta \log n)\, dt =$$
$$\geq \sum_{n=1}^{\infty} \int_n^{n+1} \mathrm{P}(X_1 > \delta \log t)\, dt = \int_1^{+\infty} \mathrm{P}(X_1 > \delta \log t)\, dt.$$

Poiché

$$\mathrm{P}(X_1 > \delta \log t) = \mathrm{E}(\mathrm{e}^{\frac{1}{\delta} X_1}) - 1.$$

La serie (4.10) diverge se $\delta < \frac{1}{\bar{\theta}}$. Dunque se $\delta < \frac{1}{\bar{\theta}}$, per il lemma di Borel-Cantelli,

$$\mathrm{P}\Big(\overline{\lim_{n\to\infty}} \{\frac{X_n}{\log n} \geq \delta\}\Big) = 1$$

e quindi $\mathrm{P}(\overline{\lim}_{n\to\infty} \frac{X_n}{\log n} \geq \frac{1}{\bar{\theta}}) = 1$. Se invece $\delta > \frac{1}{\bar{\theta}}$, il lemma di Borel-Cantelli dà $\mathrm{P}(\overline{\lim}_{n\to\infty}\{\frac{X_n}{\log n} \geq \delta\}) = 0$, cioè

$$\mathrm{P}\Big(\underline{\lim_{n\to\infty}} \Big\{\frac{X_n}{\log n} \leq \delta\Big\}\Big) = 1,$$

e quindi $\mathrm{P}(\overline{\lim}_{n\to\infty} \frac{X_n}{\log n} \le \frac{1}{\bar\theta}) = 1$. Poiché q.c. valgono sia $\overline{\lim}_{n\to\infty} \frac{X_n}{\log n} \ge \frac{1}{\bar\theta}$ che $\overline{\lim}_{n\to\infty} \frac{X_n}{\log n} \le \frac{1}{\bar\theta}$, si conclude che

$$\mathrm{P}\Big(\overline{\lim_{n\to\infty}} \frac{X_n}{\log n} = \frac{1}{\bar\theta}\Big) = 1.$$

b) Se $\bar\theta = +\infty$, per il lemma di Borel-Cantelli $\mathrm{P}(\overline{\lim}_{n\to\infty}\{\frac{X_n}{\log n} \ge \delta\}) = 0$ per ogni $\delta > 0$ e quindi $\lim_{n\to\infty} \frac{X_n}{\log n} = 0$ q.c.

c) Si ha

$$\overline{\lim_{n\to\infty}} \frac{|X_n|}{\sqrt{\log n}} = \overline{\lim_{n\to\infty}} \sqrt{\frac{|X_n|^2}{\log n}} \tag{4.11}$$

e poiché $|X_n|^2$ ha legge $\chi^2(1)$, basta calcolare il valore di $\frac{1}{\bar\theta}$ per le leggi $\chi^2(1)$. Queste hanno densità $f(x) = \frac{1}{2\sqrt{\pi}} \mathrm{e}^{-\frac{1}{2}x}$, per $x > 0$ e $f(x) = 0$ per $x < 0$. Dunque

$$\mathrm{E}(\mathrm{e}^{\theta X_1}) = \frac{1}{2\sqrt{\pi}} \int_0^{+\infty} \mathrm{e}^{(\theta-\frac{1}{2})x}\, dx.$$

Questo integrale converge per $\theta < \frac{1}{2}$ e diverge per $\theta \ge \frac{1}{2}$. Dunque $\bar\theta = \frac{1}{2}$ ed il $\overline{\lim}$ nella (4.11) vale $\sqrt{2}$.

S4.3 1. a1) X_i è di Bernoulli $B(1, \frac{1}{n})$.

a2) Poichè $X = X_1 + \ldots + X_n$, dal punto precedente segue che $\mathrm{E}(X) = 1$.

a3)

$$\mathrm{P}(A_1 \cap \ldots \cap A_n) = \frac{(n-k)!}{n!} = \frac{1}{n(n-1)\ldots(n-k+1)}$$

a4) Si ha

$$\mathrm{Var}(X) = \sum_{i=1}^{n} \mathrm{Var}(X_i) + \sum_{i\ne j} \mathrm{Cov}(X_i, X_j)$$

Si ha chiaramente $\mathrm{Var}(X_i) = \frac{1}{n}(1 - \frac{1}{n})$. Inoltre la v.a. $X_i X_j$ vale 1 se si hanno matching ai posti i e j e 0 se no. Si tratta dunque ancora di una v.a. di Bernoulli. Con una semplie applicazione della legge ipergeometrica, si vede che la probabilità di avere matching ai posti i e j vale $p = \frac{1}{n(n-1)}$. Dunque $\mathrm{E}(X_i X_j) = \frac{1}{n(n-1)}$ e

$$\mathrm{Cov}(X_i, X_j) = \mathrm{E}(X_i X_j) - \mathrm{E}(X_i)\mathrm{E}(X_j) = \frac{1}{n(n-1)} - \frac{1}{n^2} = \frac{1}{n^2(n-1)}$$

Riprendendo il calcolo si trova

$$\mathrm{Var}(X) = n \cdot \frac{1}{n}\Big(1 - \frac{1}{n}\Big) - n(n-1) \cdot \frac{1}{n^2(n-1)} = 1$$

Da notare che, per a2) e a3), media e varianza di X non dipendono da n. Nel punto b) si calcola la distribuzione limite di X per $n \to \infty$. Qual è una legge nota che ha media e varianza uguali a 1?

b1) Si calcola immediatamente, $G'(1) = \mathrm{E}(X) = 1$ e

$$G''(1) = \mathrm{E}(X(X-1)) = \mathrm{Var}(X) + \mathrm{E}(X)^2 - \mathrm{E}(X) = 1.$$

b2) Per far vedere che vale la (4.5) si può procedere per induzione. L'affermazione è ovvia per $k = 1$. Se la ammettiamo al livello k e ricordando che $X = \sum_{j=1}^n 1_{A_j}$, si ha

$$\begin{aligned} X(X-1)\dots(X-k+1)(X-k) &= \sum 1_{A_{i_1}\cap\dots\cap A_{i_k}} \Big(\sum_{j=1}^n 1_{A_j} - k\Big) = \\ &= \sum \Big(\sum_{j=1}^n 1_{A_{i_1}\cap\dots\cap A_{i_k}} 1_{A_j} - k 1_{A_{i_1}\cap\dots\cap A_{i_k}}\Big) \end{aligned} \tag{4.12}$$

Ora nella somma

$$\sum_{j=1}^n 1_{A_{i_1}\cap\dots\cap A_{i_k}} 1_{A_j}$$

per $i_1, \dots, i_k$ fissati, vi sono k termini uguali a $1_{A_{i_1}\cap\dots\cap A_{i_k}}$, corrispondenti ai valori $j = i_1, \dots, i_k$. Gli altri termini invece sono della forma $1_{A_{i_1}\cap\dots\cap A_{i_{k+1}}}$, con $i_1, \dots, i_{k+1}$ distinti tra loro. Quindi riprendendo il calcolo

$$\begin{aligned} X(X-1)\dots(X-k+1)(X-k) &= \sum_{i_1\dots i_k} \sum_{j\neq i_1,\dots,j\neq i_k} 1_{A_{i_1}\cap\dots\cap A_{i_k}\cap A_j} = \\ &= \sum_{i_1\dots i_{k+1}} 1_{A_{i_1}\cap\dots\cap A_{i_{k+1}}}. \end{aligned}$$

Un altro modo, più intuitivo, per provare la (4.5) è il seguente. Il termine di sinistra si annulla se $X \le k-1$. Quello di destra pure perché, se vi sono meno di $k-1$ matching, gli eventi $A_{i_1} \cap \dots \cap A_{i_k}$ sono tutti vuoti. Viceversa, se $X = j \ge k$, allora nella somma di destra vi sono $\frac{j!}{(j-k)!} = j(j-1)\dots(j-k+1)$ termini uguali a 1, e ancora i due membri sono uguali. b3) Poichè, per $k \le n$, le k-uple ordinate di indici distinti dall'insieme $\{1\dots n\}$ hanno cardinalità $n(n-1)\dots(n-k+1)$ e $G^{(k)}(1) = \mathrm{E}[X(X-1)\dots(X-k+1)]$, da a3) si ottiene subito che $G^{(k)}(1) = 1$, per ogni $k \le n$. Per $k > n$, si ha subito $G^{(k)} = 0$. Infatti la v.a. X può prendere al più il valore n e, dunque la sua funzione generatrice è un polinomio di grado n.

b4) Dal punto precedente si ricava che la funzione generatrice di X vale

$$G(z) = \sum_{k=1}^n \frac{(z-1)^k}{k!} \tag{4.13}$$

e quindi per $n \to \infty$ si ha

$$G(z) \to \sum_{k=1}^{\infty} \frac{(z-1)^k}{k!} = \mathrm{e}^{z-1}$$

che è la funzione generatrice di una v.a. di Poisson di parametro 1. Ciò fa pensare che X converga in legge verso questa distribuzione. Anzi questo fatto sarebbe provato se si sapesse che la convergenza delle funzioni generatrici implica la convergenza delle leggi. Questo fatto, certamente vero, non è però facile da trovare in letteratura. Per concludere rigorosamente si può sviluppare la (4.13) per ottenere una espressione esplicita della probabilità $\mathrm{P}(X=i)$ e poi farne il limite per $n \to \infty$. Si ha $(z-1)^k = \sum_{i=1}^{k} \binom{k}{i} z^i (-1)^{k-i}$. Dunque

$$\begin{aligned} G(z) &= \sum_{k=1}^{n} \frac{1}{k!} \sum_{i=1}^{k} \binom{k}{i} z^i (-1)^{k-i} = \sum_{i=1}^{n} z^i \sum_{k=i}^{n} \frac{1}{k!} \binom{k}{i} (-1)^{k-i} = \\ &= \sum_{i=1}^{n} \frac{z^i}{i!} \sum_{k=i}^{n} \frac{1}{(k-i)!} (-1)^{k-i} = \sum_{i=1}^{n} \frac{z^i}{i!} \sum_{j=1}^{n-i} \frac{1}{j!} (-1)^j \end{aligned}$$

da cui si ricava

$$\mathrm{P}(X=i) = \frac{1}{i!} \sum_{j=1}^{n-i} \frac{1}{j!} (-1)^j \underset{n\to\infty}{\to} \mathrm{e}^{-1} \frac{1}{i!}$$

che conclude rigorosamente il calcolo della legge limite.

Indice analitico

σ-additività, 4
σ-algebra di Borel, 1
σ-algebre, 1
  indipendenti, 17

algebre, 1

Borel-Cantelli, lemma, 60

Carathéodory, teorema, 5
Cauchy, legge, 55
Chebyshev, disuguaglianza, 24
classi monotone, 1
  teorema, 2
Cochran, teorema, 42
convergenza
  in L^p, 62
  in legge, 73
  in probabilità, 62
  quasi certa, 62
  stretta, 66
covarianza, 24
  matrice, 26

densità, 12
distribuzioni condizionali, 46
  per le gaussiane multivariate, 48
disuguaglianza
  di Chebyshev, 24
  di Hölder, 9, 22
  di Jensen, 21
  di Markov, 24
  di Minkowski, 9, 22
  di Schwartz, 9, 22

eventi, 15

Fisher, approssimazione, 89
Fubini, teorema, 14
funzione di ripartizione, 15
funzioni
  integrabili, 7
  semi-integrabili, 7
funzioni caratteristiche, 26

Hölder, disuguaglianza, 9, 22

indipendenza
  di σ-algebre, 17
  di v.a., 18

Jensen, disuguaglianza, 21

Kolmogorov
  legge 0-1, 20
  legge forte, 66

Lebesgue
  misura, 6, 14
  teorema, 8
legge forte
  di Kolmogorov, 66
  di Rajchmann, 65
leggi
  beta, 87
  condizionali, 46
  dei grandi numeri, 65
  di Cauchy, 55
  di Rayleigh, 53
  di Student, 43
  lognormali, 85
  normali multivariate, 38
lemma
  di Borel-Cantelli, 60
  di Slutsky, 81
lemma di Fatou, 8

Markov, disuguaglianza, 24
masse di Dirac, 11
matching, 85, 96
matrice di covarianza, 26
metodo delta, 81
Minkowski, disuguaglianza, 9, 22
misura di Lebesgue, 6, 14
misure, 4
  σ-finite, 4
  definite da una densità, 12
  di Borel, 5
  di Dirac, 11
  di probabilità, 4
  finite, 4
  immagine, 11
  su un'algebra, 4
momenti di una v.a., 23

ordinamenti stocastici, 52

Pearson, teorema, 76

Radon-Nikodym, teorema, 13
Rajchmann, legge forte, 65
Rayleigh, legge, 53
retta di regressione, 25

Scheffé, teorema, 90
Schwartz, disuguaglianza, 9, 22
Slutsky, lemma, 81
spazi L^p, 9
spazi di misura, 4
  completi, 6
speranza matematica, 15

teorema
  d'inversione, 33, 91
  di Beppo-Levi, 7
  di Carathéodory, 5
  di Cochran, 42
  di derivazione sotto il segno, 8
  di Fubini, 14
  di Lebesgue, 8
  di Pearson, 76
  di Radon-Nikodym, 13
  di Scheffé, 90
  limite centrale, 75
trascurabile
  insieme, 6
trasformata di Laplace complessa, 34

variabili aleatorie, 15
  centrate, 16
  indipendenti, 18
  non correlate, 25
varianza di una v.a., 23