Calcolo delle Probabilità II

Gregorio Baldi

April 22, 2014

Contents

1 Teorema Centrale

L'obiettivo di questa prima sezione è sviluppare gli strumenti necessari per dimostrare il **Teorema Centrale** con il metodo delle funzioni caratteristiche. Enunciamolo ...

Teorema 1.1 (Teorema Centrale)**.** *Sia $X_1, X_2, \dots, X_n, \dots$ una successione di variabili casuali che siano*

1. *equidistribuite*
2. *(globalmente) indipendenti*
3. *$X_i \in L^2$ $\forall i = 1, 2, \dots$ con media m e varianza σ^2.*

Allora, definita

$$Y_n = \frac{X_1 + X_2 + \cdots - nm}{\sqrt{n\sigma^2}}$$

si ha che la funzione caratteristica associata a Y_n, che indichiamo con $\varphi_{Y_n}(t)$, per $n \to \infty$ tende alla funzione caratteristica associata a una gaussiana, ovvero

$$\varphi_{Y_n}(t) \to e^{\frac{-t^2}{2}}$$

Dimostreremo che, il Teorema Centrale enunciato come qui sopra, è equivalente alla formulazione più nota:

$$\mathbb{P}\left(a \leqslant \frac{X_1 + X_2 \cdots + X_n - nm}{\sqrt{n\sigma^2}} \leqslant b\right) \to \int_a^b \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}}\, dx.$$

In maniera analoga dimostreremo anche la **Legge dei Grandi Numeri**.

Osservazione 1.1. Dimostreremo e enunceremo questi due teoremi non nella loro massima generalità.

1.1 Preliminari vari

Situazione di partenza, con la quale bisogna prendere più confidenza possibile:

$$\begin{array}{ccc} \Omega & \xi & \mathbb{P} \\ \downarrow X & \uparrow X^{-1} & \downarrow \\ \mathbb{R} & \mathcal{B} & \mu \end{array}$$

- abbiamo una terna $(\Omega, \xi, \mathbb{P})$ dove Ω è un insieme, che assumiamo non vuoto per evitare banalità, ξ è una $\sigma - algebra$ definita su Ω e per concludere $\mathbb{P}$ è una misura di probabilità
- uno spazio di misura $(\mathbb{R}, \mathcal{B}, ?)$ dove $\mathcal{B}$ indica la $\sigma - algebra\ di\ Borel$ ovvero la più piccola $\sigma - algebra$ che contiene la topologia standard con cui sono equipaggiati i reali

- nel punto precedente abbiamo lasciato un ? al posto di una misura. Data una *variabile casuale* $X : \Omega \to \mathbb{R}$, $X^{-1} : \xi \to \mathcal{B}$ vogliamo definire una misura μ sui Boreliani, semplice: poniamo $\mu(I) \doteq \mathbb{P}(X^{-1}(I))\ \forall I \in \mathcal{B}$. Osserviamo che la μ così definita è una misura, in particolare è una misura di probabilità (la verifica è semplice, ad esempio $\mu(\mathbb{R}) = \mathbb{P}(X^{-1}(\mathbb{R})) = 1$). La freccia tratteggiata sta ad indicare che la definizione della μ si ricava dal resto del diagramma.

Teorema 1.2 (Formula di Astratto-Concreto)**.** *Data una funzione $F : \mathbb{R} \to \mathbb{R}$ boreliana (i.e. $F^{-1}(B) \in \mathcal{B}\ \forall B \in \mathcal{B}$)allora vale la seguente relazione*

$$\int_\Omega F(X(\omega))d\mathbb{P}(\omega) = \int_\mathbb{R} F(x)\mu(dx).$$

Proof. Diamo solo una rapida idea della dimostrazione: per prima cosa si dimostra la relazione precedente per le funzioni indicatrici o, nel linguaggio della analisi, per le funzioni caratteristiche, degli intervalli reali. Poi per le funzioni semplici (combinazioni lineari di funzioni indicatrici). Poi per una funzione $F \geq 0$, e per concludere per F. La dimostrazione ricorda molto i passaggi che abbiamo fatto per definire l'integrale astratto.

□

Dato un spazio di misura $(\mathbb{R}, \mathcal{B}, \mathbb{P})$ poniamo

$$F(x) = \mathbb{P}\left((-\infty, x]\right),\ x \in \mathbb{R}$$

questa funzione soddisfa le seguenti proprietà:

1. F è non decrescente
2. $F(-\infty) = 0,\ F(+\infty) = 1$
3. F è continua a destra e ammette limite a sinistra $\forall x \in \mathbb{R}$

Definizione 1.1. Ogni funzione $F = F(x)$ che soddisfa le condizioni 1. 2. e 3. è chiamata *funzione di ripartizione* (*distribution function*).

Teorema 1.3. *Sia $F = F(x)$ una funzione di distribuzioine sulla retta reala $\mathbb{R}$. Allora esiste un'unica misura di probabilità $\mathbb{P}$ su $(\mathbb{R}, \mathcal{B})$ tale che*

$$\mathbb{P}\left((a, b]\right) = F(b) - F(a)\ \forall a, b \in [-\infty, \infty)\ .$$

Altri fatti generali:

1. Esistono i Boreliani come σ-algebra di un generico spazio topologico?

   Teorema 1.4. *Sia $\mathcal{F}$ una qualsiasi collezioni di sottoinsiemi di X, allora esiste la più piccola σ-algebra $\mathcal{F}^*$ di X tale che $\mathcal{F} \in \mathcal{F}^*$.*

   Tale $\mathcal{F}^*$ è detta la σ algebra *generata* da $\mathcal{F}$.

2. Useremo più volte il seguente risultato.

Teorema 1.5 (Teorema di Convergenza Dominata)**.** *Dato uno spazio di misura* $(X, \mathcal{F}, \mu)$ *e una successione* (f_n) *di funzioni misurabili su* X *tali che esiste il limite*

$$\lim_n f_n(x) = f(x) \ \forall x \in X.$$

Se esiste una funzione $g \in L^1$ *che domina le* f_n*, i.e. tale che*

$$|f_n(x)| \leqslant g(x),$$

allora f_n *converge a* f *in* L^1*.*

3. I boreliani hanno la potenza di $\mathbb{R}$: $|\mathcal{B}| = |\mathbb{R}|$

4. Una misura μ si dice completa se ogni sottoinsieme di un insieme di misura nulla è misurabile. Ad esempio la misura di Lebesgue sulla retta $\mathbb{R}$: per provarlo basta far vedere che ogni insieme con misura esterna nulla soddisfa l'uguaglianza di Carathéodory; si ottiene con delle semplici disuguaglianze

5. *L'insieme di Cantor* è un sottoinsieme dell'intervallo $[0, 1]$, chiuso che ha la stessa cardinalità di $[0, 1]$

6. Sia $\mathcal{M}$ l'insieme dei Lebesgue-misurabili: $|\mathcal{M}| = 2^{|\mathbb{R}|}$ (si dimostra usando i punti precedenti)

7. Un esempio di misura non completa è la misura di Lebesgue definita sulla σ algebra dei Borelian $\mathcal{B}$: se per assurdo fosse completa in particolare ogni sottoinsieme dell'insieme di Cantor appartiene ai Boreliani (grazie al punto 5.), quindi i Boreliani devono avere almeno la potenza di $2^{|\mathbb{R}|}$. Ma questa conclusione è una contraddizione con il punto 3.

8. Ogni spazio di misura si può *completare*. Vale, infatti il seguente risultato.

   Teorema 1.6. *Sia* (Ω, ξ, μ) *un qualunque spazio di misura. Esistono una* σ*-algebra* ξ^* *e una misura* μ^* *su* Ω *tali che*

   - *estendono* ξ *e* μ*, i.e.* $\xi \subset \xi^*$ *e* $\mu^*_{|\xi} = \mu$
   - *lo spazio di misura* (Ω, ξ^*, μ^*) *è completo*
   - *se* (Ω, ξ', μ') *è un altro spazio con le due proprietà precedenti, allora*

   $$\xi^* \subset \xi' \text{ e } \mu'_{|\xi^*} = \mu^*.$$

   Lo spazio di misura (Ω, ξ^*, μ^*) si dice il *completamento* dello spazio (Ω, ξ, μ). Grazie a questo risultato, ogni volta che sarà conveniente, potremo assumere che un dato spazio di misura si completo. Ad esempio la σ-algebra dei misurabili secondo Lebesgue e la misura di Lebesgue sono il completamento della σ-algebra dei Boreliani (*su* $\mathbb{R}^n$) con la solita misura.
   (Come corollario a questo risultato si avrebbe anche il seguente: *ogni insieme misurabile secondo Lebesgue è unione disgiunta di un boreliano e di un insieme misurabile secondo Lebesgue, avente misura nulla*).

9. Siano date due misure μ e λ sulla stessa $\sigma - algebra\ \xi$, diciamo che μ è *assolutamente continua* rispetto a λ , e scriviamo $\mu \ll \lambda$ se per ogni $A \in \xi$ si ha $\lambda(A) = 0 \Rightarrow \mu(A) = 0$

10. Una funzione f definita sull'intervallo [a,b] a valori in $\mathbb{R}$ diciamo che è *assolutamente continua* se possiede una derivata f' definita quasi ovunque e integrabile secondo Lebesgue tale che:

$$f(x) = f(a) + \int_a^x f'(t)\ dt \quad \forall x \in [a,b]$$

Un esempio di funzione continua ma non assolutamente continua si costruisce grazie all'insieme di Cantor ed è nota come *Funzione di Cantor* o *Scala del Diavolo*

1.2 Funzione Caratterisitica

Definizione 1.2. Data una misura di probabilità μ su $\mathbb{R}$ definiamo la seguente $\varphi : \mathbb{R} \to \mathbb{C}$, $\varphi(t) = \int_{\mathbb{R}} e^{ixt}\mu(dx)$ come la *funzione caratteristica di* μ .

Osservazione 1.2. Esiste l'integrale che abbiamo appena scritto? Ovvero, vale la seguente relazione $\int_{\mathbb{R}} |e^{ixt}|\mu(dx) < +\infty$? Naturalmente, perchè $|e^{ixt}| = 1$, dunque risulta $\int_{\mathbb{R}} |e^{ixt}|\mu(dx) = \int_{\mathbb{R}} 1\mu(dx) = \mu(\mathbb{R}) = 1$

Osservazione 1.3. Notiamo che $|\varphi(t)| \leqq 1$, cioè φ è un cammino contenuto nella palla unitaria complessa.
Inoltre vale $\varphi(0) = 1$.

Proposizione 1.1. *Assegnate due misure μ_1, μ_2 sui reali, tali che le funzioni caratteristiche associate a μ_1 e μ_2 coincidono, allora si ha $\mu_1 = \mu_2$*

Questa proposizione giustifica il nome *funzione caratteristica*, nel senso che c'è iniettività: le informazioni che contiene μ sono legate alle informazioni contenute nella sua funzione caratteristica e viceversa. A breve vedremo alcuni esempi di questo legame.

Esempio 1.1. Sia μ la distribuzione di probabilità uniforme su $[0, 1]$. Vogliamo calcolarne la funzione caratteristica:

$$\varphi(t) = \int_0^1 e^{itx}dx = \left[\frac{e^{itx}}{it}\right]_0^1 = \frac{e^{it}-1}{it} = \frac{e^{i\frac{t}{2}}}{it}\frac{sin\frac{t}{2}}{\frac{t}{2}}$$

A questo punto è ragionevole chiedersi quando una generica funzione $\varphi : \mathbb{R} \to \mathbb{C}$ è una funzione caratteristica. Ovvero ci si chiede se esiste una misura μ tale che $\varphi(t) = \int_{\mathbb{R}} e^{ixt}\mu(dx)$.

Teorema 1.7 (Teorema di Bochner)**.** *Una $\varphi : \mathbb{R} \to \mathbb{C}$ è caratteristica se e solo se valgono le seguenti:*

1. $\varphi(0) = 1$
2. *φ è continua nello 0*
3. $\forall n \geq 1 \forall t_1, t_2, .., t_n \in \mathbb{R} \forall z_1, z_2, .., z_n \in \mathbb{C} \Rightarrow \sum_{j,k=1}^n \varphi(t_j - t_k)z_j\overline{z_k} \geq 0$.

Osservazione 1.4. Risulterà utile più volte la seguente relazione: siano $\alpha_i \in \mathbb{C}\ \forall i = 1, \ldots, n$

$$\sum_{j,k=1}^n \alpha_j\overline{\alpha_k} = \left|\sum_{i=1}^n \alpha_i\right|^2 .$$

La dimostrazione è semplice, basta notare le seguenti relazioni: $\alpha\overline{\alpha} = |\alpha|^2$ e $(x_1 + iy_1)(x_2 - iy_2) + (x_2 + iy_2)(x_1 - iy_1) = 2x_1x_2 + 2y_1y_2$.
La relazione del punto 3. deve valere, in particolare, per $n = 1$. In questo caso si tratta solo di dimostrare che $\forall t \in \mathbb{R}\ \forall z \in \mathbb{C}$ si ha $\varphi(0)|z|^2 \geq 0$ ed è banalmente vera.

Teorema 1.8. *Se $X \in L^1(\Omega)$ allora φ, la funzione caratteristica associata alla misura μ associata alla variabile casuale X (d'ora in avanti ometteremo così tanta precisione nello specificare a cosa è associata una funzione caratteristica), è derivabile e*

$$\varphi'(t) = i \int_{\mathbb{R}} xe^{itx} d\mu.$$

Se $X \in L^2(\Omega)$ allora φ' è derivabile e

$$\varphi''(t) = -\int_{\mathbb{R}} x^2 e^{itx} d\mu.$$

Banalmente si osserva che

$$\varphi'(0) = i \int_{\mathbb{R}} xd\mu = im = i\mathbb{E}(X)$$

e che

$$\varphi''(0) = -\int_{\mathbb{R}} x^2 d\mu = -\mathbb{E}(X^2)$$

grazie alla formula di astratto-concreto; si ha infine

$$Var(X) = \mathbb{E}(X^2) - \mathbb{E}(X) = -\varphi''(0) + \varphi'(0)^2 \ .$$

Proof. Per $\delta > 0$ valgono le seguenti uguaglianze:

$$\frac{\varphi(t+\delta) - \varphi(t)}{\delta} = \int_{\mathbb{R}} \frac{e^{ix(t+\delta)} - e^{ixt}}{\delta}\ d\mu = \int_{\mathbb{R}} e^{itx} \frac{\left(e^{i\delta x} - 1\right)}{\delta}\ d\mu.$$

Dato che

$$\frac{\left(e^{i\delta x} - 1\right)}{\delta} \to ix \text{ per } \delta \to 0,$$

se dimostriamo che si può portare il limite sotto il segno di integrale, abbiamo concluso. Per farlo basta verificare che siano soddisfatte le ipotesi del Teorema di Convergenza Dominata (Teorema 1.5). Si tratta di trovare una g tale che

$$\left| e^{itx} \frac{\left(e^{i\delta x} - 1\right)}{\delta} \right| = \left| \frac{\left(e^{i\delta x} - 1\right)}{\delta} \right| \leqslant g.$$

Per farlo notiamo che valgono le seguenti:

$$\left|e^{i\delta x} - 1\right| = |\cos(\delta x) + i\sin(\delta x) - 1| = \sqrt{\cos^2(\delta x) + 1 - 2\cos(\delta x) + \sin^2(\delta x)} =$$

$$\sqrt{2\left(1 - \cos(\delta x)\right)} = \sqrt{4\sin^2\left(\frac{\delta x}{2}\right)} = \left|2\sin\left(\frac{\delta x}{2}\right)\right|.$$

Dunque, ricordando la relazione

$$\left|\frac{\sin\alpha}{\alpha}\right| \leqslant 1,\ \forall\alpha\in\mathbb{R},$$

abbiamo che

$$\left|\frac{\left(e^{i\delta x}-1\right)}{\delta}\right| = \frac{\left|\sin\left(\frac{\delta x}{2}\right)\right|}{\left|\frac{\delta x}{2}\right|}\,|x| \leqslant |x|\,.$$

Visto che $X\in L^1$ segue che $g=|X|\in L^1$, quindi possiamo applicare il teorema e concludere che

$$\frac{\varphi(t+\delta)-\varphi(t)}{\delta} \to i\int_{\mathbb{R}} xe^{itx}d\mu,\ \textit{per}\ \delta\to 0.$$

Per la derivata seconda il conto è del tutto analogo. □

Vediamo alcune proprietà della funzione caratteristica $\varphi(t)$ associata alla funzione di ripartizione di una variabile casuale X.

Proposizione 1.2. *Sia X una variabile casuale con funzione di ripartizione F=F(x) e funzione caratteristica*

$$\varphi(t)=\mathbb{E}\left[e^{itX}\right]\ .$$

Allora φ ha le seguenti proprietà:

1. $|\varphi(t)|\leqslant\varphi(0)=1$
2. $\varphi(t)$ *è uniforme continua* $\forall t\in\mathbb{R}$
3. $\varphi(t)=\overline{\varphi(-t)}$
4. $\varphi(t)$ *è a valori reali se e solo se F è simmetrica, ovvero* $\int_B dF(x) = \int_{-B} dF(x)\ , B\in\mathcal{B}$.

Alcune Funzioni Caratteristiche ...

Esempio 1.2. Consideriamo una variabile casuale di tipo gaussiano $N(0,1)$, i.e. distribuita secondo

$$f(x)=e^{itx}\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}.$$

La sua funzione caratteristica risulta essere

$$\psi(t)=\int_{\mathbb{R}} e^{itx}\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}\ dx=\cdots=e^{-\frac{t^2}{2}}.$$

Per una generica variabile casuale distribuita secondo $N(m,\sigma^2)$, tramite semplici passaggi algebrici, si ha che vale la seguente relazione tra la sua funzione caratteristica φ e quella della gaussiana $N(0,1)$:

$$\varphi(t)=\int_{\mathbb{R}} e^{itx}\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x-m)^2}{2\sigma^2}}\ dx=\cdots=e^{imt}\varphi(\sigma t).$$

Esempio 1.3. Consideriamo una variabile casuale con distribuzione di tipo esponenziale $Exp(\lambda)$, i.e. con funzione di ripartizione sui reali positivi

$$f(x) = \lambda e^{-\lambda x},$$

la funzione caratteristica associata è

$$\varphi(t) = \int_0^{+\infty} e^{itx} \lambda e^{-\lambda x} \, dx = \frac{\lambda}{\lambda - it}.$$

Per ottenere l'ultima uguaglianza basta risolvere i due (semplici) seguenti integrali:

$$\int_0^{+\infty} \cos(xt) e^{-\lambda x} \, dx, \quad \int_0^{+\infty} \sin(xt) e^{-\lambda x} \, dx.$$

1.3 Convergenza in Legge

Sia $\mathcal{M}$ l'insieme di tutte le misure di probabilità sullo spazio di misura $(\mathbb{R}, \mathcal{B})$ i.e. $\mu \in \mathcal{M} \Rightarrow \mu \geqslant 0, \mu(\mathbb{R}) = 1$, equipaggiato della somma $+$:

$$(\mu + \lambda)(I) = \mu(I) + \lambda(I) \ \forall \mu, \lambda \in \mathcal{M}, \ \forall I \in \mathbb{R}.$$

Banalmente si osserva che $\mathcal{M}$ è stabile per combinazioni convesse (solo convesse perchè devono sempre essere misure di probabilità).

Definizione 1.3 (Convergenza in Legge)**.** Sia $(\mu_n) \in \mathcal{M}$ diciamo che (μ_n) *converge in legge* (o *in distribuzione* o *debole*) a μ e scriviamo $\mu_n \xrightarrow{(L)} \mu$ se e solo se

$$\forall f \in C_b(\mathbb{R}), \int_{\mathbb{R}} f(x) \mu_n(dx) \to \int_{\mathbb{R}} f(x) \mu(dx).$$

Esercizio 1. Se $\mu_n \xrightarrow{(L)} \mu$ allora, in particolare, si ha

$$f(x) = cos(xt) \rightsquigarrow \int_{\mathbb{R}} cos(xt) d\mu_n \to \int_{\mathbb{R}} cos(xt) d\mu$$

analogamente per

$$f(x) = sin(xt) \rightsquigarrow \int_{\mathbb{R}} sin(xt) d\mu_n \to \int_{\mathbb{R}} sin(xt) d\mu.$$

Equivalentemente

$$\int_{\mathbb{R}} e^{itx} d\mu_n \to \int_{\mathbb{R}} e^{itx} d\mu,$$

ovvero $\varphi_{\mu_n}(t) \to \varphi_\mu(t) \forall t \in \mathbb{R}$ (puntualmente).

Abbiamo dimostrato che la convergenza in legge implica la convergenza puntuale delle funzioni caratteristiche. La buona notizia è che vale anche il viceversa.

Proposizione 1.3. *La convergenza in legge delle misure è equivalente alla convergenza delle rispettive funzioni caratteristiche.*

Proposizione 1.4. *Sia $(\mu_n) \in \mathcal{M}$ una successione di misure su $(\mathbb{R}, \mathcal{B})$ con rispettive funzioni caratteristiche associate $F_n(x) = \mu_n(]-\infty, x])$. Sono equivalenti*

1. $\mu_n \xrightarrow{(L)} \mu$

2. $\forall I \in \mathcal{B}$ *t.c.* $\mu(\partial I) = 0,\ \mu_n(I) \to \mu(I)$

3. $F_n(x) \to F(x) \forall x \in \mathbb{R}$ *t.c.* F *è continua.*

1.4 Dimostrazione del Teorema Centrale

Definizione 1.4. Siano $X_1, X_2, \dots, X_n$ n variabili casuali, diciamo che sono *globalmente indipendenti* se

$$\mathbb{P}\left(X_1 \in I_1, X_2 \in I_2, \dots, X_n \in I_n\right) = \mathbb{P}\left(X_1 \in I_1\right) \dots \left(X_n \in I_n\right)$$

per ogni $I_k \in \mathcal{B}$.

Proposizione 1.5. *$X_1, X_2, \dots, X_n$ n variabili casuali sono globalmente indipendenti (d'ora in poi ometteremo globalmente) se e solo se per ogni scelta di $f_1, f_2, \dots, f_n$ boreliane e limitate, $f_i : \mathbb{R} \to \mathbb{R}$, si ha*

$$\mathbb{E}[f_1(X_1) f_2(X_2) \dots f_n(X_n)] = \mathbb{E}[f_1(X_1)] \dots \mathbb{E}[f_n(X_n)]$$

Esercizio 2. $X_1, X_2, \dots, X_n$ n variabili casuali sono indipendenti se e solo se per ogni scelta di $g_1, g_2, \dots, g_n$ boreliane e limitate, $g_i : \mathbb{R} \to \mathbb{C}$, si ha

$$\mathbb{E}[g_1(X_1) g_2(X_2) \dots g_n(X_n)] = \mathbb{E}[g_1(X_1)] \dots \mathbb{E}[g_n(X_n)].$$

Una implicazione è ovvia. Per l'altra basta scrivere ogni $g_i(t) = \mathcal{R}_i(t) + i\mathcal{I}_i(t)$, riscrivere le relazioni precedenti e applicare la proposizione precedente alla parte reale e immaginaria di g_i.

Definizione 1.5. Una successione di variabili casuali si dice *(globalmente) indipendente* se ogni suo sottoinsieme finito di variabili casuali è composto di variabili casuali globalmente indipendenti (nel senso della definizione precedente).

Ricapitolando stavamo dicendo: sia X una variabile casuale, $X \in L^2(\Omega)$, possiamo dunque calcolare $\mathbb{R}(X) = m$ e $Var(X) = \sigma^2$. In modo canonico associamo a X la misura immagine μ, e a μ associamo la funzione caratteristica $\varphi(t)$. Possiamo anche calcolare direttamente φ da X mediante la relazione

$$\varphi(t) = \int_{\mathbb{R}} e^{ixt} \mu(dx) = \mathbb{E}(e^{itx}).$$

Esercizio 3. Consideriamo la variabile casuale $Y = \frac{X-m}{\sigma}$. Y è tale che

$$\mathbb{E}(Y) = 0 \text{ e } Var(Y) = 1.$$

Per questo Y si dice *normalizzata* di X. Le relazioni precedenti seguono dalla linearità dell'integrale che definisce l'aspettazione e dalla relazione $Var(Z) = -\mathbb{E}(Z)^2 + \mathbb{E}(Z^2)$.

Siano X, Y come prima, sia φ la funzione caratteristica associata a X e calcoliamo la funzione caratteristica associata a Y:

$$\varphi_Y(t) = \mathbb{E}\left(e^{itY}\right) = \mathbb{E}\left(e^{it\frac{x-m}{\sigma}}\right) = \mathbb{E}\left(e^{-it\frac{m}{\sigma}} e^{ix\frac{t}{\sigma}}\right) =$$

$$= e^{-it\frac{m}{\sigma}}\mathbb{E}\left(e^{ix\frac{t}{\sigma}}\right) = e^{-im\frac{t}{\sigma}}\varphi\left(\frac{t}{\sigma}\right).$$

Osserviamo inoltre che

$$\varphi_Y''(0) = -\mathbb{E}\left(Y^2\right) = -Var\left(Y\right) - \mathbb{E}\left(Y\right)^2 = -1 - 0 = -1$$

Ora siamo pronti per dimostrare il Teorema 1.1

Proof. Consideriamo $X_1, X_2, \ldots, X_n$ n variabili casuale equidistribuite, indipendenti e appartenenti a L^2 con media m e varianza σ^2.
Definiamo

$$Y_n \doteq \frac{X_1 + X_2 + \cdots + X_n - nm}{\sqrt{n\sigma^2}}.$$

banalmente si ottiene

$$Y_n = \frac{1}{\sqrt{n}}\left(\frac{X_1 - m}{\sigma} + \cdots + \frac{X_n - m}{\sigma}\right).$$

Per alleggerire la notazione, definiamo

$$\widehat{X_i} \doteq \frac{X_i - m}{\sigma},\ \forall i = 1, 2, \ldots, n.$$

Possiamo dunque calcolare la funzione caratteristica φ_{Y_n} associata a Y_n:

$$\varphi_{Y_n}(t) = \mathbb{E}\left(e^{itY_n}\right) = \mathbb{E}\left(e^{i\frac{t}{\sqrt{n}}\left(\widehat{X_1}+\cdots+\widehat{X_n}\right)}\right) =$$

$$= \mathbb{E}\left(e^{i\frac{t}{\sqrt{n}}\widehat{X_1}}\ldots e^{i\frac{t}{\sqrt{n}}\widehat{X_n}}\right) = \mathbb{E}\left(f\left(\widehat{X_1}\right)\ldots f\left(\widehat{X_n}\right)\right)$$

dove si è posto

$$f\left(X\right) = e^{i\frac{t}{\sqrt{n}}X}.$$

Data l'indipendenza delle variabili casuali, e ricordando uno degli esercizi precedenti, possiamo fattorizzare:

$$\varphi_{Y_n}(t) = \mathbb{E}\left(e^{i\frac{t}{\sqrt{n}}\widehat{X_1}}\right)\ldots\mathbb{E}\left(e^{i\frac{t}{\sqrt{n}}\widehat{X_n}}\right) = \widehat{\varphi}\left(\frac{t}{\sqrt{n}}\right)^n$$

dove l'ultima uguaglianza segue dal fatto che le X_i siano equidistribuite, ovvero ad ogni X_i viene associata la stessa misura μ, alla quale viene associata la stessa funzione caratteristica $\varphi(t)$; inoltre abbiamo visto prima che la funzione caratteristica associata alle normalizzate $\widehat{X_k}$ è:

$$\widehat{\varphi}(t) = e^{-i\frac{t}{\sigma}m}\varphi\left(\frac{t}{\sigma}\right)$$

e non dipende da k, quindi l'uguaglianza.
Dobbiamo mostrare che

$$\varphi_{Y_n}(t) = \widehat{\varphi}\left(\frac{t}{\sqrt{n}}\right)^n \to e^{\frac{-t^2}{2}}.$$

Per farlo sviluppiamo in serie di Taylor la funzione $\widehat{\varphi}(s)$ in un intorno dell'origine:

$$\widehat{\varphi}(s) = \widehat{\varphi}(0) + \widehat{\varphi}'(0)s + \frac{1}{2}\widehat{\varphi}''(0)s^2 + o\left(s^2\right) = 1 - \frac{s^2}{2} + o\left(s^2\right)$$

dove l'ultima uguaglianza segue dalle relazioni precedenti tra una generica φ e le sue derivate, e naturalmente dal fatto che le variabili casuali appartengono a L^2.
Ora si ha:

$$\varphi_{Y_n}(t) = \widehat{\varphi}\left(\frac{t}{\sqrt{n}}\right)^n = \left(1 - \frac{t^2}{2n} + o\left(\frac{t^2}{n}\right)\right)^n = \left(1 - \frac{t^2}{2n}\right)^n \left(1 + \frac{o\left(\frac{t^2}{n}\right)}{1 - \frac{t^2}{2n}}\right)^n .$$

Visto che vale la relazione

$$\left(1 - \frac{t^2}{2n}\right)^n \to e^{-\frac{t^2}{2}}$$

ci rimane da mostrare che il secondo fattore dell'uguaglianza precedente tende a 1. Sappiamo che

$$no\left(\frac{t^2}{n}\right) \to 0 \ per \ n \to \infty$$

o anche

$$n\frac{o\left(\frac{t^2}{n}\right)}{1 - \frac{t^2}{2n}} \to 0 \ per \ n \to \infty.$$

Se definiamo

$$z_n \doteq \frac{o\left(\frac{t^2}{n}\right)}{1 - \frac{t^2}{2n}}$$

abbiamo che $nz_n \to 0$. Per mostrare che

$$\left(1 + \frac{o\left(\frac{t^2}{n}\right)}{1 - \frac{t^2}{2n}}\right)^n - 1 = (1 + z_n)^n - 1 \to 0 \ per \ n \to \infty$$

e quindi concludere, notiamo che valgono le seguenti relazioni:

$$|(1 + z_n)^n - 1| = \left|1 + \sum_{k=1}^{n} \binom{n}{k} z_n^k - 1\right| \leqslant \sum_{k=1}^{n} \binom{n}{k} |z_n|^k + 1 - 1 =$$

$$= (1 + |z_n|)^n - 1 \leqslant e^{n|z_n|} - 1 \to 0$$

dove le ultime uguaglianze seguono da

$$|z_n| \geq 0 \Rightarrow 1 + |z_n| \leqslant e^{|z_n|} \ \left(e^x = 1 + x + \frac{x^2}{2} + \dots\right)$$

e da

$$nz_n \to 0 \Rightarrow e^{n|z_n|} \to 1.$$

□

Proposizione 1.6. *Il Teorema Centrale dimostrato nella forma precedente è equivalente a*

$$\mathbb{P}\left(a \leqslant \frac{X_1 + X_2 \dots + X_n - nm}{\sqrt{n\sigma^2}} \leqslant b\right) \to \int_a^b \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}} \ dx$$

Proof. Dall'Esempio 1.2 sappiamo che

$$\varphi(t) = e^{-\frac{t^2}{2}}$$

è la funzione caratteristica associata ad una variabile casuale ti tipo $N(0,1)$. Nella sezione "Convergenza in Legge" abbiamo visto l'equivalenza tra la convergenza in legge delle misure associate $\mu_n \xrightarrow{(L)} \mu$ e la convergenza delle funzioni caratteristiche $\varphi_n(t) \to \varphi(t)$.
Grazie al teorema appena dimostrato abbiamo che

$$\mu_n \xrightarrow{(L)} N(0,1).$$

Equivalentemente, grazie alla Proposizione 1.4,

$$\forall I \in \mathcal{B} \ t.c. \ \mu(\partial I) = 0, \ \mu_n(I) \to \mu(I)$$

in particolare deve valere per $I = [a,b]$, di conseguenza si ha

$$\mathbb{P}\left(a \leqslant \frac{X_1 + X_2 \cdots + X_n - nm}{\sqrt{n\sigma^2}} \leqslant b\right) = \mu_n\left([a,b]\right) \to$$

$$\to \int_a^b \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}} \, dx = \mu\left([a,b]\right)$$

□

In maniera analoga dimostriamo l'altro risultato notevole citato all'inizio.

Teorema 1.9 (Legge dei Grandi Numeri). *Sia $X_1, X_2, \ldots, X_n, \ldots$ una successione di variabili casuali che siano*

1. *equidistribuite*
2. *(globalmente) indipendenti*
3. $X_i \in L^1 \ \forall i = 1, 2, \ldots$

Allora, per $n \to \infty$, si ha

$$\forall \epsilon > 0, \mathbb{P}\left(\left|\frac{X_1 + X_2 \cdots + X_n}{n} - m\right| > \epsilon\right) \to 0$$

Proof. Consideriamo la variabile casuale

$$Y_n \doteq \frac{X_1 + X_2 \cdots + X_n}{n}$$

la funzione caratteristica φ_{Y_n} associata a Y_n è

$$\varphi_{Y_n}(t) = \mathbb{E}\left(e^{itY_n}\right) = \mathbb{E}\left(e^{it\frac{(X_1 + \cdots + X_n)}{n}}\right) =$$

$$= \mathbb{E}\left(e^{it\frac{X_1}{n}} \ldots e^{it\frac{X_n}{n}}\right) = \mathbb{E}\left(e^{i\frac{t}{n}X_1}\right)^n = \psi\left(\frac{t}{n}\right)^n$$

dove si è posto

$$\psi(s) = \mathbb{E}\left(e^{isX_1}\right).$$

Sviluppiamo la funzione $\psi(s)$ ma, questa volta, solo fino al primo ordine. Considerare $\psi''(s)$ non avrebbe senso visto che le variabili casuali sono, per ipotesi, elementi di L^1 e non di L^2 come era, invece, nelle ipotesi del Teorema Centrale. Abbiamo dunque:

$$\psi(s) = \psi'(0) + \psi'(0)s + o(s) = 1 + ims + o(s), \ \textit{con } so(s) \to 0 \textit{ se } s \to 0.$$

Quindi

$$\varphi_{Y_n}(t) = \mathbb{E}\left(e^{i\frac{t}{n}X_1}\right)^n = \psi\left(\frac{t}{n}\right)^n = \left(1 + im\frac{t}{n} + o\left(\frac{t}{n}\right)\right)^n =$$

$$= \left(1 + im\frac{t}{n}\right)^n \left(1 + \frac{o\left(\frac{t}{n}\right)}{1 + im\frac{t}{n}}\right)^n \xrightarrow{n\to\infty} e^{imt}$$

dove l'ultimo limite si ottiene con considerazioni simili a quelle fatte per l'ultimo limite del Teorema Centrale. Sappiamo che

$$\varphi(t) = e^{imt}$$

è la funzione caratteristica associata ad una *delta di Dirac* δ_m concentrata in m (è un caso particolare di normale di media m e varianza 0). Ovvero è la funzione caratteristica associata alla misura

$$\mu(I) = \delta_m(I) = \{1 \textit{ se } m \in I, \ 0 \textit{ altrimenti}.$$

Ora, grazie al fatto che le funzioni caratteristiche convergono se e solo se convergono in legge le misure, si ha

$$\mathbb{P}\left(\left|\frac{X_1 + X_2 \cdots + X_n}{n} - m\right| > \epsilon\right) = \mathbb{P}\left(Y_n \in [m-\epsilon, m+\epsilon]^c\right) =$$

$$\mu_n\left([m-\epsilon, m+\epsilon]^c\right) \to \delta_m\left([m-\epsilon, m+\epsilon]^c\right) = 0$$

dove le ultime uguaglianze seguono dalle seguenti

$$\partial\left([m-\epsilon, m+\epsilon]^c\right) = \{m-\epsilon, m+\epsilon\}$$

$$\delta_m\left(m-\epsilon, m+\epsilon\right) = 0$$

$$m \notin [m-\epsilon, m+\epsilon]^c.$$

Si ha dunque la tesi. □

Osservazione 1.5. Nel Teorema Centrale e nella Legge dei Grandi Numeri è stata fatta, implicitamente, un'altra ipotesi: che esista una una successione di variabili casuali con le proprietà richieste. Ci torneremo ...

1.5 Altri tipi di Convergenza

Come in Analisi, anche in Teoria della Probabilità, abbiamo bisogno di vari tipi di convergenza delle variabili casuali. Tra questi i più importanti sono i seguenti:

Definizione 1.6 (Convergenza in Probabilità)**.** Diciamo che una successione di variabili casuali (X_n) converge *in probabilità* alla variabile casuale X

$$X_n \xrightarrow{(P)} X \Leftrightarrow \forall \epsilon > 0,\ \mathbb{P}\left(|X_n - X| > \epsilon\right) \to 0 \ per\ n \to \infty.$$

Abbiamo già visto un'espressione di questa forma nella Legge dei Grandi Numeri; tale legge, alla luce della definizione appena data, dice che *la media aritmetica di una successione di variabili casuali converge in probabilità alla media teorica*.
In analisi questa convergenza è nota come *convergenza in misura*.

Definizione 1.7 (Convergenza quasi sicuramente)**.** Diciamo che una successione di variabili casuali (X_n) converge *quasi sicuramente* alla variabile casuale X

$$X_n \to X \ q.s. \Leftrightarrow X_n(w) \to X(w) \ \forall w \in N^c, N \in \xi \ t.c. \ \mathbb{P}\left(N\right) = 0$$

i.e. se l'insieme dei punti $w \in \Omega$ per i quali $X_n(w)$ non converge a X ha probabilità 0.

Definizione 1.8 (Convergenza nella norma di L^p)**.** Diciamo che una successione di variabili casuali $(X_n) \in L^p$ converge *nel senso di* L^p alla variabile casuale $X \in L^p$

$$X_n \xrightarrow{L^p} X \Leftrightarrow ||X_n - X||_{L^p} \to 0 \ per\ n \to \infty.$$

Teorema 1.10. *Valgono le seguenti implicazioni:*

$$X_n \xrightarrow{(P)} X \Rightarrow \mu_n \xrightarrow{(L)} \mu$$

$$X_n \to X \ q.s. \Rightarrow X_n \xrightarrow{(P)} X$$

$$X_n \xrightarrow{L^p} X \Rightarrow X_n \xrightarrow{(P)} X$$

Proof. La prima implicazione non è banale, la dimostrazione si trova qui
Per la terza implicazione ci serviamo della *disuguaglianza di Chebyshev*:

$$\mathbb{P}\left(|X| > \epsilon\right) = \int_{|X|>\epsilon} \mathbb{P}(dw) \leqslant \int_{|X|>\epsilon} \frac{|X|^p}{\epsilon^p}\mathbb{P}(dw) \leqslant$$

$$\frac{1}{\epsilon^p}\int_{\Omega} |X|^p d\mathbb{P} = \mathbb{E}\left(|X|^p\right) = \frac{1}{\epsilon^p}||X||^p_{L^p}.$$

Dunque

$$X_n \xrightarrow{L^p} X \Rightarrow \mathbb{P}\left(|X_n - X| > \epsilon\right) \leqslant \frac{1}{\epsilon^p}||X_n - X||^p_{L^p} \to 0 \Rightarrow X_n \xrightarrow{(P)} X.$$

□

2 Caso n-dimensionale

In questa sezione generalizzeremo quanto visto fin ora al caso n-dimensionale. Nella prima parte daremo la definizione di funzione caratteristica associata a misure definite sui Boreliani di $\mathbb{R}^n$ e vedremo che continuano a valere i teoremi che abbiamo enunciato nel caso 1-dimensionale. Nella seconda ci occuperemo di cosa succede alle variabili casuali del tipo $N(m, \sigma^2)$ in più dimensioni.

2.1 Nuovo linguaggio

Siano X_i con i=1,2, ..., n delle variabili casuali tali che

$$\begin{tikzcd} \Omega \arrow[d, "X_i"] & \xi & \mathbb{P} \arrow[d, dashed] \\ \mathbb{R} & \mathcal{B} \arrow[u, "X_i^{-1}"'] & \mu_i \end{tikzcd}$$

Possiamo dunque considerare il *vettore casuale*

$$\mathbf{X} = (X_1, X_2, \dots, X_n)$$

e il seguente diagramma:

$$\begin{tikzcd} \Omega \arrow[d, "\mathbf{X}"] & \xi & \mathbb{P} \arrow[d, dashed] \\ \mathbb{R}^n & \mathcal{B}^n \arrow[u, "\mathbf{X}^{-1}"'] & \mu \end{tikzcd}$$

dove abbiamo indicato con $\mathcal{B}^n$ i *boreliani di* $\mathbb{R}^n$.

Dato un qualsiasi vettore casuale

$$\mathbf{X} : \Omega \to \mathbb{R}^n$$

possiamo considerare

$$X_i = \pi_i \circ \mathbf{X} : \Omega \to \mathbb{R}$$

dove π_i è la proiezione sull'i-esimo fattore da $\mathbb{R}^n$ in $\mathbb{R}$. Si ha che ciascuna X_i così definita è una variabile casuale.
Dal fatto che

$$\mathcal{B}(\mathbb{R}^n) = \mathcal{B}_1 \otimes \mathcal{B}_1 \otimes \cdots \otimes \mathcal{B}_1\text{[1]}$$

possiamo osservare l'equivalenza tra le seguenti:

1. siano $X_1, \dots, X_n$ v.c. allora formano un vettore casuale $\mathbf{X} = (X_1, \dots, X_n)$
2. ogni proiezione del vettore casuale $\mathbf{X}$, indicata con X_i, è una variabile casuale.

[1] con il simbolo $\otimes$ si è indicata la σ-algebra prodotto, per i dettagli si veda la sezione 2.3.1. sulle Misure Prodotto.

La misura immagine tramite il vettore casule $\mathbf{X}$, indicata nel diagramma precedente con μ, è una misura di probabilità sulla σ-algebra dei Boreliani di $\mathbb{R}^n$. L'obiettivo della prossima sezione è trovare un'altra misura *di probabilità* μ_g, assolutamente continua rispetto alla misura di Lebesgue, sugli stessi Boreliani. In analogia al caso 1-dimensionale, saremo interessati a calcolare la *funzione caratteristica* associata a questa misura μ_g. Prima di questi passi vediamo la generalizzazione della definizione di f.c. e enunciamo alcuni risultati che continuano a valere.

Definizione 2.1. Sia μ una misura di probabilità definita sullo spazio di misura $(\mathbb{R}^n, \mathcal{B}^n)$. La *funzione caratteristica associata a* μ è

$$\varphi : \mathbb{R}^n \to \mathbb{C}$$

$$\varphi(t) = \int_{\mathbb{R}^n} e^{i<t,x>} \, d\mu$$

Osservazione 2.1. Esiste l'integrale che abbiamo appena scritto? Naturalmente: come nel caso 1-dimensionale si ha $\left|e^{i<t,x>}\right| = 1$.

Proposizione 2.1. *Assegnate due misure μ_1, μ_2 su $(\mathbb{R}^n, \mathcal{B}^n)$, tali che le funzioni caratteristiche associate a μ_1 e μ_2 coincidono, allora si ha $\mu_1 = \mu_2$.*

Teorema 2.1 (Teorema di Bochner n-dim)**.** *Una funzione $\varphi : \mathbb{R}^n \to \mathbb{C}$ è caratteristica se e solo se valgono le seguenti:*

1. $\varphi(0) = 1$
2. φ *è continua nello 0*
3. $\forall n \geq 1 \forall t_1, t_2, .., t_n \in \mathbb{R}^n \forall z_1, z_2, .., z_n \in \mathbb{C} \Rightarrow \sum_{j,k=1}^{n} \varphi(t_j - t_k) z_j \overline{z_k} \geq 0$.

2.2 Gaussiana Generalizzata

Consideriamo una matrice quadrata n-dimensionale

$$A = \begin{bmatrix} a_{1,1} & \cdots & a_{1,n} \\ \vdots & \ddots & \vdots \\ a_{n,1} & \cdots & a_{n,n} \end{bmatrix}$$

tale che

- A sia simmetrica (i.e. $a_{i,j} = a_{j,i} \ \forall i, j = 1, \ldots, n$)
- la forma quadratica associata ad A sia definita positiva, i.e.

$$\langle Ax, x \rangle^{2} = \sum_{i,j=1}^{n} a_{i,j} x_i x_j \geq 0 \text{ e } = 0 \Leftrightarrow x = 0$$

[2] da qui in avanti indicheremo sempre il prodotto scalare standard con il simbolo $< \cdot \, , \, \cdot >$.

Sotto queste ipotesi A è invertibile ($\det A \neq 0$), possiamo dunque considerare la seguente funzione:

$$f(x) = e^{-\frac{1}{2}\langle A^{-1}x,x\rangle}$$

Notiamo subito che vale

$$f(x) > 0 \; \forall x \in \mathbb{R}^n$$

A questo punto ha senso considerare il seguente integrale:

$$\int_{\mathbb{R}^n} f(x)\; dx$$

chiedersi se risulta minore di infinito e, in questo caso, quanto vale.

Sappiamo che per $n = 1$ si ha $A = \{a\}$ e

$$\int_{\mathbb{R}} e^{-\frac{1}{2}\frac{x^2}{a}}\; dx = \sqrt{2\pi a}$$

Il caso n=2 è stato trattato nel corso di CdP I, ma per un generico n come possiamo fare? Possiamo diagonalizzare la matrice A (è simmetrica e definita positiva per ipotesi), questa idea guiderà la dimostrazione del prossimo risultato.

Teorema 2.2. *Sia A una matrice come sopra, allora vale*

$$\int_{\mathbb{R}^n} e^{-\frac{1}{2}\langle A^{-1}x,x\rangle} = (2\pi)^{\frac{n}{2}}\sqrt{\det A}$$

Proof. Per ipotesi sulla matrice A, esiste una matrice unitaria (o ortogonale, i.e. tale che $U^{-1} = U^*$. O a parole: tale che l'inversa è uguale alla trasposta) per la quale si ha la seguente uguaglianza

$$U^{-1}AU = D = \begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_n \end{bmatrix}$$

con $\lambda_1, \dots, \lambda_n \in \mathbb{R}$ grazie all'ipotesi di simmetria della matrice A (altrimenti, a priori, sono dei numeri complessi). Da questo ricaviamo che

$$\begin{bmatrix} \frac{1}{\lambda_1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{1}{\lambda_n} \end{bmatrix} = D^{-1} = \left(U^{-1}AU\right)^{-1} = UAU^{-1}$$

Dalla quali si ricava anche

$$A^{-1} = UD^{-1}U^{-1}$$

Ora che abbiamo scritto la matrice inversa di A in una forma più comoda possiamo calcolare l'integrale iniziale:

$$\int_{\mathbb{R}^n} e^{-\frac{1}{2}\langle A^{-1}x,x\rangle}\; dx = \int_{\mathbb{R}^n} e^{-\frac{1}{2}\langle UD^{-1}U^{-1}x,x\rangle}\; dx$$

ricordando la relazione che vale per ogni $t, z \in \mathbb{R}^n$

$$\langle Bt, z\rangle = \langle t, B^* z\rangle$$

possiamo procedere con la sostituzione $y = U^{-1}x$, e ricordando l'invarianza per rotazione della misura di Lebesgue (o meglio perchè il determinante dello Jacobiano del cambio di coordinate, in modulo, è 1), abbiamo che l'integrale precedente è uguale a

$$\int_{\mathbb{R}^n} e^{-\frac{1}{2}\left\langle D^{-1}y,y\right\rangle}\ dy = \int_{\mathbb{R}^n} e^{-\frac{1}{2}\sum_{j=1}^{n}\frac{y_j^2}{\lambda_j}}\ dy =$$

$$= \left(\int_{\mathbb{R}} e^{-\frac{y_1^2}{2\lambda_1}}\ dy_1\right)\left(\int_{\mathbb{R}} e^{-\frac{y_2^2}{2\lambda_2}}\ dy_2\right)\dots\left(\int_{\mathbb{R}} e^{-\frac{y_n^2}{2\lambda_n}}\ dy_n\right) =$$

$$= \prod_{j=1}^{n}\sqrt{2\pi\lambda_j} = (2\pi)^{\frac{n}{2}}\sqrt{\lambda_1\lambda_2\dots\lambda_n} = (2\pi)^{\frac{n}{2}}\sqrt{\det A}$$

□

Ora possiamo considerare la funzione

$$g(x) = (2\pi)^{-\frac{n}{2}}\frac{1}{\sqrt{\det A}}e^{-\frac{1}{2}\left\langle A^{-1}x,x\right\rangle}$$

come una densità di probabilità rispetto alla misura n-dimensionale di Lebesgue; perchè abbiamo appena dimostrato che

$$\int_{\mathbb{R}^n} g(x)\ dx = 1$$

In maniera naturale, da g densità di probabilità, ricaviamo una misura μ_g sui boreliani di $\mathbb{R}^n$ ponendo

$$\mu_g\left(I\right) = \int_I g(x)\ dx$$

per ogni I boreliano di $\mathbb{R}^n$.

Proposizione 2.2. *La funzione caratteristica associata a μ_g è*

$$\varphi_g(t) = e^{-\frac{1}{2}\langle At,t\rangle}$$

Proof. Per comodità nei prossimi passaggi chiamiamo C la seguente costante

$$C \doteq (2\pi)^{-\frac{n}{2}}\frac{1}{\sqrt{\det A}}$$

Dalla definizione di funzione caratteristica segue

$$\varphi_g(t) = C\int_{\mathbb{R}^n} e^{i<t,x>}e^{-\frac{1}{2}\left\langle D^{-1}U^{-1}x,U^{-1}x\right\rangle}\ dx$$

come prima procediamo con la sostituzione $y = U^{-1}x$ e, come prima, abbiamo che l'integrale precedente è uguale a

$$C\int_{\mathbb{R}^n} e^{i<t,Uy>}e^{-\frac{1}{2}\left\langle D^{-1}y,y\right\rangle}\ dy = C\int_{\mathbb{R}^n} e^{i\left\langle U^{-1}t,y\right\rangle}e^{-\frac{1}{2}\sum_{j=1}^{n}\frac{y_j^2}{\lambda_j}}\ dy$$

Possiamo procedere con un'ulteriore sostituzione ponendo $s = U^{-1}t$, di conseguenza abbiamo le seguenti uguaglianze

$$\left\langle U^{-1}t, y\right\rangle = \langle s, y\rangle = \sum_{j=1}^{n} s_j y_j$$

Quindi l'integrale precedente risulta uguale a

$$C\int_{\mathbb{R}^n} e^{is_1y_1-\frac{y_1^2}{2\lambda_1}}\dots e^{is_ny_n-\frac{y_n^2}{2\lambda_n}}\; dy_1\dots dy_n =$$

$$= \left(\int_{\mathbb{R}} \frac{e^{is_1y_1-\frac{y_1^2}{2\lambda_1}}}{\sqrt{2\pi\lambda_1}}\; dy_1\right)\dots\left(\int_{\mathbb{R}} \frac{e^{is_ny_n-\frac{y_n^2}{2\lambda_n}}}{\sqrt{2\pi\lambda_n}}\; dy_n\right)$$

ma ogni integrale 1-dimensionale corrisponde alla funzione caratteristica di una gaussiana del tipo $N(o, \sigma^2)$, che sappiamo calcolare. Risulta dunque che l'ultima espressione è uguale a

$$e^{-\frac{\lambda_1 s_1^2}{2}}\dots e^{-\frac{\lambda_n s_n^2}{2}} = e^{-\frac{1}{2}\sum_{j=1}^{n}\lambda_j s_j^2} = e^{-\frac{1}{2}\langle At,t\rangle}$$

dove l'ultima uguaglianza segue da

$$\sum_{j=1}^{n}\lambda_j s_j^2 = \langle Ds, s\rangle = \left\langle U^{-1}AUs, s\right\rangle =$$

$$= \left\langle U^{-1}AUU^{-1}t, t\right\rangle = \left\langle U^{-1}At, U^{-1}t\right\rangle = \langle At, t\rangle\,.$$

□

Osservazione 2.2. La funzione caratteristica è *un'arma a doppio taglio*: mette in luce e nasconde alcune informazioni. In questo caso propone una notevole semplificazione in quanto non compare $\sqrt{\det A}$ e, al posto della sua inversa, si ha proprio la matrice A.

Data una matrice A simmetrica e semidefinita positiva (i.e. $\langle Ax, x\rangle \geq 0$) possiamo considerare la seguente funzione

$$\varphi(t) = e^{-\frac{1}{2}\langle At,t\rangle\geq 0}$$

E' ben definita e risulta essere maggiore di 0 sul dominio. A questo punto ha senso chiedersi se è una funzione caratteristica.
Nel caso in cui A sia definita positiva sappiamo esibire, in maniera esplicita, una misura. In generale?

Proposizione 2.3. *Sia $\varphi(t)$ come sopra, allora soddisfa le condizioni del Teorema di Bochner n-dim e quindi è una funzione caratteristica.*

Proof. Che $\varphi(0) = 1$ e la continuità in 0 si verificano facilmente. Per l'ultima condizione consideriamo la seguente matrice

$$A_\epsilon \doteq A + \epsilon I$$

Si ha
$$\langle A_\epsilon x, x\rangle = \langle Ax, x\rangle + \epsilon |x|^2$$
che risulta essere > 0 per ogni $x \neq 0$ e $= 0$ per $x = 0$, i.e. A_ϵ è una matrice definita positiva e, naturalmente, simmetrica. Sappiamo la funzione caratteristica associata alla misura
$$\mu_\epsilon(I) = \int_I g_\epsilon(x)\ dx$$
dove, per la stessa C della Proposizione 2.2, si è posto
$$g_\epsilon(x) \doteq Ce^{-\frac{1}{2}\left\langle A_\epsilon^{-1}x,x\right\rangle}$$
è proprio
$$\varphi_\epsilon(t) = e^{-\frac{1}{2}\langle A_\epsilon t,t\rangle}$$
Segue che $\varphi_\epsilon(t)$ soddisfano la disuguaglianza del Teorema di Bochner per ogni ϵ. Vale inoltre
$$\varphi_\epsilon(t) = e^{-\frac{1}{2}\langle A_\epsilon t,t\rangle} = \varphi(t)e^{-\frac{\epsilon}{2}|t|^2}$$
Visto che le $\varphi_\epsilon(t)$ soddisfano Bochner e che tendono a $\varphi(t)$ per ϵ che tende a zero, segue dunque che $\varphi(t)$ deve soddisfare la disuguaglianza. Quindi la tesi. □

Data una matrice A simmetrica e definita positiva possiamo considerare la funzione
$$g(x) = \left(\frac{1}{2\pi}\right)^n \frac{1}{\sqrt{\det A}} e^{-\frac{1}{2}\left\langle A^{-1}(x-m),(x-m)\right\rangle}$$
con $m \in \mathbb{R}^n$. Tale funzione è la densità di una gaussiana di tipo $N(m, A)$ dove m è detta *media* e A *matrice di covarianza*. In questo caso, tramite un cambio di variabile, abbiamo già dimostrato che la funzione caratteristica risulta
$$\varphi(t) = e^{i<m,t>-\frac{1}{2}\langle At,t\rangle}$$

Definizione 2.2. Una misura μ, non per forza data in termini di densità rispetto alla misura di Lebesgue, si dice *gaussiana generalizzata* di tipo $N(m, A)$ se la sua funzione caratteristica è
$$\varphi(t) = e^{i<m,t>-\frac{1}{2}\langle At,t\rangle}$$
per qualche matrice A simmetrica semidefinita positiva.

Esercizio 4. Sia A definita positiva e $m = (m_1, m_2, \dots, m_n) \in \mathbb{R}^n$ e la funzione g densità del tipo $N(m, A)$. Allora valgono i seguenti

1. $\int_{\mathbb{R}^n} x_j g(x)\ dx = m_j$
2. $\int_{\mathbb{R}^n} (x_j - m_j)^2 g(x)\ dx = Var(j - esima\ componente)$
3. $\int_{\mathbb{R}^n} (x_j - m_j)(x_k - m_k) g(x)\ dx = a_{jk}$.

Sia X un vettore gaussiano di tipo $N(m, A)$ in $\mathbb{R}^n$ e B una generica matrice con m righe e n colonne. Il vettore $Y = BX$ è una variabile casuale a valori in $\mathbb{R}^m$ (la composizione di funzioni boreliane è boreliana); ma come è distribuita?

Proposizione 2.4. *$Y = BX$ è un vettore gaussiano di tipo $N(Bm, BAB^*)$.*

Proof. La funzione caratteristica associata ad X è nota. Sia $t \in \mathbb{R}^m$ la f.c. associata a Y è

$$\varphi_y(t) = \mathbb{E}\left(e^{i<t,Y>}\right) = \mathbb{E}\left(e^{i<t,BX>}\right) = \mathbb{E}\left(e^{i<B^*t,X>}\right) =$$

$$= e^{i<m,B^*t>-\frac{1}{2}<AB^*t,B^*t>} = e^{i<Bm,t>-\frac{1}{2}<BAB^*t,t>}$$

dove si è usato il fatto che $< t, Bx >=< B^*t, x >$ (il primo è un prodotto scalare in $\mathbb{R}^m$ mentre il secondo in $\mathbb{R}^n$). Con dei semplici conti si ha che BAB^* è semidefinita positiva, quindi la tesi. □

Osservazione 2.3. Come casi particolari del risultato precedente si ha che l'i-esima componente di X è guassiana, basta scegliere la matrice

$$B = (0, 0, \ldots, 1, \ldots, 0)$$

Ma lo sono, per lo stesso motivo, anche le sue componenti prese a coppie, a tre a tre, e così via. In particolare ogni combinazione lineare

$$\alpha_1 X_1 + \alpha_2 X_2 + \cdots + \alpha_n X_n$$

di componenti di X è gaussiano.

Osservazione 2.4. Se $\overline{X}$ è un generico vettore le cui componenti sono tutte di tipo gaussiano allora, in generale, non possiamo concludere che $\overline{X}$ lo sia! E' necessaria l'indipendenza delle componenti.

2.3 Misure in più dimensioni

2.3.1 Misure Prodotto

Per definire la misura di Lebesgue in $\mathbb{R}^n$ è necessario ripercorrere la procedura svolta per definire la misura di Lebesgue sulla retta reale. Ecco i passaggi principali (la loro dimostrazione è analoga al caso 1-dimensionale):

- si definisce il *volume* n-dimensionale dei pluri-intervalli $R = \prod_{i=1}^n I_i$ dove gli I_i sono intervalli della retta reale ponendo

  $$v_n(R) = \prod_{i=1}^n l(I_i)$$

  dove si era posto $l(I) = b - a$ se I è della forma $(a, b) \subset I \subset [a, b]$, infinito se I è illimitato; con la convenzione naturale $0 \cdot \infty = 0$

- si introduce la misura esterna n-dimensionale m_n^* su ogni sottoinsieme E di $\mathbb{R}^n$ ponendo

  $$m_n^*(E) = \inf\{\sum_{i\in\mathbb{N}} v_n(R_i) \text{ t.c. } E \subseteq \bigcup_{i\in\mathbb{N}} R_i,\ R_i \text{ pluri-intervalli }\}$$

- m_n^* estende v_n, è monotona, nulla sull'insieme vuoto, numerabilmente subadditiva e invariante per traslazioni

- si introduce la classe $\mathcal{M}_n$ degli insiemi misurabili:

$$\mathcal{M}_n = \{E \subseteq \mathbb{R}^n \text{ t.c. } m_n^*(A) = m_n^*(A \cap E) + m_n^*(A \cap E^c) \, \forall A \subseteq \mathbb{R}^n\}$$

- $\mathcal{M}_n$ è una σ-algebra contenente i pluri-intervalli, gli aperti e i chiusi; in particolare contiene la σ-algebra dei Boreliani di $\mathbb{R}^n$
- si definisce la misura di Lebesgue in $\mathbb{R}^n$ come la restrizione di m_n^* a $\mathcal{M}_n$, e si dimostra che è numerabilmente additiva.

Nonostante questa costruzione non siamo, ancora, capaci di ricondurre il calcolo degli integrali in $\mathbb{R}^n$ a più integrazioni lungo $\mathbb{R}$ e non sappiamo quando è lecito scambiare l'ordine di integrazione. Per permettere questi passaggi dobbiamo vedere che rapporti intercorrono tra la misura di Lebesgue n-dimensionale e quella 1-dimensionale. Per questo è il momento di parlare di di σ-algebre e misure prodotto.

Dati due spazi di misura $(X, \mathcal{F}, \mu)$ e $(Y, \mathcal{G}, \psi)$ consideriamo l'insieme $X \times Y$ e il suo sottoinsieme dei *rettangoli misurabili*

$$\mathcal{R} = \{A \times B \text{ t.c. } A \in \mathcal{F}, B \in \mathcal{G}\}$$

Questo insieme è chiusa soltanto rispetto all'intersezione, ed è una collezione molto ristretta di sottinsiemi di $X \times Y$: nel caso di $X = \mathbb{R} = Y$ vorremmo misurare, oltre ai rettangoli, anche i cerchi, i triangoli e molti altri. D'altra parte, su $\mathcal{R}$ possiamo definire in maniera del tutto naturale una misura λ:

$$\lambda(E) = \lambda(A \times B) = \mu(A) \psi(B) \text{ per ogni } E = A \times B \in \mathcal{R}.$$

Definizione 2.3. Indichiamo con $\mathcal{F} \otimes \mathcal{G}$ la più piccola σ-algebra contenente i rettangoli misurabili, esiste per il Teorema 1.4. si chiama la σ-algebra *prodotto*. La misura λ, definita prima su $\mathcal{R}$, si estende in modo unico[3] ad una misura definita sulla σ-algebra prodotto, la *misura prodotto*, che si indica con $\mu \otimes \psi$.

Per compiere delle integrazioni successive useremo spesso i Teoremi di Fubini e Tonelli.

2.3.2 Marginali

Vediamo un altro modo di costruire misure in più dimensioni.

La nostra situazione iniziale: un vettore casuale $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ e il seguente diagramma:

$$\begin{array}{ccccc} \Omega & & \xi & & \mathbb{P} \\ \downarrow \mathbf{X} & & \mathbf{X}^{-1} \uparrow & & \downarrow \\ \mathbb{R}^n & & \mathcal{B}^n & & \mu \end{array}$$

Partiamo dal caso $n = 2$, chi è $\mu(I \times \mathbb{R})$?

$$\mu(I \times \mathbb{R}) = \mathbb{P}\left(\mathbf{X}^{-1}(I \times \mathbb{R})\right) = \mathbb{P}\left(X_1^{-1}(I)\right) = \mu_1(I)$$

[3] andrebbe dimostrato naturalmente

dove la penultima uguaglianza segue dalla seguente

$$\mathbf{X}^{-1}(I \times \mathbb{R}) = \{w \in \Omega \text{ t.c } \mathbf{X}(w) \in I \times \mathbb{R}\} =$$

$$= \{w \in \Omega \text{ t.c. } X_1(w) \in I, X_2(w) \in \mathbb{R}\} = \{w \in \Omega \text{ t.c. } X_1(w) \in I\}$$

Per concludere, nota la misura 2-dimensionale μ, possiamo definire due sue *marginali 1-dimensionali* μ_1 e μ_2 ponendo

$$\mu_1(I) = \mu(I \times \mathbb{R}) \text{ e } \mu_2(J) = \mu(\mathbb{R} \times J) \text{ per ogni } I, J \in \mathbb{R}.$$

Più in generale, nel caso generico n-dimensionale potremo considerare n misure μ_i 1-dimensionali, le marginali due dimensionali μ_{ij} definite in maniera analoga, fino alle marginali $n-1$ dimensionali.

Osservazione 2.5. Una misura non è univocamente determinata dalle sue Marginali.

Con le Marginali c'è, in un certo senso, sovrabbondanza di informazione "una contenuta nell'altra". Come spiega la prossima proposizione non è necessario conoscere tutte le marginali, ma, da quelle k-dimensionali, si ricavano quelle di dimensione minore in maniera naturale.

Proposizione 2.5. *Data una misura n-dimensionale μ consideriamo la sua marginale 2-dimensionale μ_{ij} e le sue marginali 1-dimensionali μ_k. Possiamo anche considerare le marginali 1-dimensionali di μ_{ij}, indicate con $(\mu_{ij})_i$ e $(\mu_{ij})_j$. Si ha*

$$(\mu_{ij})_i = \mu_i \text{ e } (\mu_{ij})_j = \mu_j$$

Proof. Valgono le seguenti uguaglianze

$$\mu_{ij}(I \times J) = \mathbb{P}(\{w \text{ t.c. } (X_i(w), X_j(w)) \in I \times J\}) =$$

$$= \mathbb{P}(\{w \text{ t.c. } X_i(w) \in I, X_j(w) \in J\}) = \mathbb{P}\left(X_i^{-1}(I) \cap X_j^{-1}(J)\right)$$

Ora si ha

$$(\mu_{ij})_i(I) = \mu_{ij}(I \times \mathbb{R}) = \mathbb{P}\left(X_i^{-1}(I) \cap X_j^{-1}(\mathbb{R})\right) = \mathbb{P}\left(X_i^{-1}(I)\right) = \mu_i(I)$$

L'altra uguaglianza è identica, scambiando gli indici. Quindi la tesi. □

Esempio 2.1. Sia $X_1, X_2, \ldots, X_n, \ldots$ una successione di variabili casuali *indipendenti*. Essa fornisce una successione di misure $\lambda_1, \lambda_2, \ldots, \lambda_n, \ldots$. Se consideriamo il vettore casuale (X_i, X_j) abbiamo anche la misura $\mu_i j$. Possiamo esprimerla in funzione delle misure 1-dimensionali:

$$\mu_{ij}(I \times J) = \mathbb{P}(X_i \in I, X_j \in J) = {}^4\mathbb{P}(X_i \in I)\,\mathbb{P}(X_j \in J) = \mu_i(I)\,\mu_j(J)$$

ovvero

$$\mu_{ij} = \mu_i \otimes \mu_j$$

Quindi, in questo caso, note le misure 1-dimensionali, posso ricavare tutte le altre.

[4] questa uguaglianza è possibile solo grazie all'indipendenza delle v.c.

Sia $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ un vettore casuale e sia μ la misura immagine. Possiamo considerare la funzione caratteristica

$$\varphi(t) = \mathbb{E}\left(e^{i\langle t, \mathbf{X}\rangle}\right)$$

Se si può calcolare e risulta

$$\mathbb{E}(\mathbf{X}) = m = (m_1, m_2, \ldots, m_n)$$

allora si ha

$$\frac{\partial \varphi}{\partial t_j}(0) = i m_j$$

Se poi $\mathbf{X} \in L^2$ si ha

$$\mathbb{E}((X_j - m_j)(X_k - m_k)) = \mathbb{E}(X_j X_k) - m_j m_k =$$

$$= -\frac{\partial \varphi^2}{\partial t_j \partial t_k}(0) + \frac{\partial \varphi}{\partial t_j}(0)\frac{\partial \varphi}{\partial t_k}(0)$$

In particolare se $\mathbf{X}$ è del tipo $N(m, A)$ abbiamo che

$$\mathbb{E}((X_j - m_j)(X_k - m_k)) = a_{jk}$$

Per la dimostrazione si può procedere in maniera simile al Teorema 1.8.

$$\frac{\varphi(t + (0, 0, \ldots, \delta, 0, \ldots, 0)) - \varphi(t)}{\delta} = \int_{\mathbb{R}^n} \frac{e^{i\delta x_j}}{\delta}\, d\mu$$

e l'ultimo integrale tende, grazie al Teorema di Convergenza Dominata, a

$$\int_{\mathbb{R}^n} i x_j\, d\mu = \int_{\mathbb{R}^n} i\pi_j(x)\, d\mu = \int_\Omega i\pi_j(\mathbf{X})\, d\mathbb{P} = \int_\Omega i X_j\, d\mathbb{P} = i m_j$$

grazie alla Formula di Astratto-Concreto applicata alla funzione π_j, i.e. la proiezione sul j-esimo fattore .
Per la seconda parte basta notare che

$$\mathbb{E}((X_j - m_j)(X_k - m_k)) = \mathbb{E}(X_j X_k - m_j X_k - m_k X_j + m_j m_k) =$$

$$= \mathbb{E}(X_j X_k) - m_j m_k - m_j m_k + m_j m_k = \mathbb{E}(X_j X_k) - m_j m_k$$

Per concludere, usando i punti precedenti, rimane da mostrare che

$$\mathbb{E}(X_j X_k) = -\frac{\partial \varphi^2}{\partial t_j \partial t_k}(0)$$

Questo si ha perché

$$\frac{\partial \varphi^2}{\partial t_j \partial t_k}(t)_{|t=0} = \int_{\mathbb{R}^n} i x_j i x_k\, d\mu = -\int_{\mathbb{R}^n} x_i x_k\, d\mu =$$

$$= -\int_\Omega X_j X_k\, d\mathbb{P} = \mathbb{E}(X_j X_k)$$

dove si è utilizzata, come prima, la Formula di Astratto-Concreto.

Definizione 2.4. Sia $X = (X_1, X_2, \dots, X_n)$ un vettore casuale t.c. $X \in L^2$, definiamo

$$a_{jk} := \mathbb{E}\left((X_j - m_j)(X_k - m_k)\right)$$

e di conseguenza la *matrice di covarianza* $A = (a_{ij})$.

Osservazione 2.6. La matrice di covarianza è definita per ogni vettore casuale X che ammette varianza. E' simmetrica perché il prodotto commuta. Inoltre risulta definita positiva notando le seguenti uguaglianze:

$$\sum_{j,k} a_{jk} z_j z_k = \sum_{j,k} \mathbb{E}\left((X_j - m_j)(X_k - m_k)\right) z_j z_k =$$

$$\mathbb{E}\left(\sum_{j,k} (X_j - m_j) z_j (X_k - m_k) z_k\right) = \mathbb{E}\left(\sum_k (X_k - m_k) z_k\right)^2$$

Esercizio 5. Indichiamo con C la circonferenza di raggio r, ovvero

$$C = \{p \in \mathbb{R}^2 \text{ t.c. } ||p|| = r\}$$

Sia $X = (X_1, X_2)$ un vettore casuale e supponiamo che la sua misura immagine μ sia definita, per ogni Boreliano B nel seguente modo:

$$\mu(B) = m_1(B \cap C) / m_1(C)$$

dove m_1 sta ad indicare la misura 1-dimensionale di Lebesgue in $\mathbb{R}^2$. La misura μ è concentrata su un sottoinsieme C di $\mathbb{R}^2$ avente misura (di Lebesgue 2-dimensionale) nulla, quindi non è una misura assolutamente continua rispetto a Lebesgue; equivalentemente non può essere data in termini di densità rispetto a Lebesgue.

- Calcolare le marginali di μ, μ_1 e μ_2.
- Confrontare μ con $\mu_1 \otimes \mu_2$, sono la stessa funzione?
- Chi è la funzione caratteristica associata a μ? E alle marginali?
- Ripetere l'esercizio con la misura μ concentrata sulla circonferenza piena, i.e. concentrata sul disco

$$D = \{p \in \mathbb{R}^2 \text{ t.c. } ||p|| \leq r\}$$

Risulteranno funzioni non integrabili elementarmente.