# CAPITOLO 1. LE SPERANZE CONDIZIONATE: PROPRIETA'

Le Speranze Condizionate, che d'ora in poi saranno indicate con SC, furono introdotte da Kolmrogorov nel suo trattato [24], facendo seguito alla definitiva messa a punto del teorema di Radon-Nikodým che è circa del 1930. (Per la validità di tale teorema si vedano [47] e [54]).

## 1.1. LA DEFINIZIONE ELEMENTARE.

L'esame del caso elementare dà un'idea intuitiva del concetto di condizionamento.

Sia, ora e nel seguito, salvo esplicita menzione, $(\Omega, \mathscr{F}, \mu)$ uno spazio di probabilità e sia $B \in \mathscr{F}$ un evento tale che $\mu(B)>0$; se $A$ è misurabile si chiama probabilità di A condizionata da B la quantità $\mu_B(A) := \mu(A \cap B)/\mu(B)$. Si osservi che $\mu_B(A)$ non è definita se $\mu(B)=0$. La funzione $A \to {}_B(A)$ è essa stessa una probabilità su $\mathscr{F}$, sicché, fissato $B \in \mathscr{F}$ con $\mu(B)>0$, è naturale considerare accanto a quello dato lo spazio di probabilità $(\Omega, \mathscr{F}, \mu_B)$. Data una v.a. $f:(\Omega, \mathscr{F}) \to (\bar{\mathbb{R}}, \mathscr{B})$ si chiamerà SC di f dato B la speranza di f mediante la misura $\mu_B$ se tale speranza esiste, cioè

$$(1.1) \qquad E_B(f) := \int f d\mu_B = \int_B f d\mu_B + \int_{B'} f d\mu_B = \int_B f d\mu_B = \frac{1}{\mu(B)} \int_B f d\mu$$

sicché $E_B(f) = E(f 1_B)/\mu(B)$. In particolare la (1.1) dà, per $f = 1_A$ $(A \in \mathscr{F})$,

$$(1.2) \qquad E_B(1_A) = \mu_B(A)$$

di modo che per studiare le proprietà della probabilità condizionata

$\mu_B$ basterà studiare quelle della SC $E_B$.

Il vero significato delle SC scaturisce quando esse siano interpretate come valori di funzioni. Dati B e f come sopra, si può definire in $\Omega$ una funzione misurabile a due valori, $E_B(f)1_B+E_{B'}(f)1_{B'}$. In generale sia $\{B_n\}$ una partizione misurabile e, al più, numerabile di $\Omega$ (cioè $B_n\in\mathcal{F}$, $\cup_n B_n=\Omega$ , $B_r \cap B_s=\emptyset$ se $r\neq s$) e sia $\mathcal{G}$ la tribù generata da tale partizione, $\mathcal{G}=\sigma(\{B_n\})$; evidentemente $\mathcal{G}$ è contenuta in $\mathcal{F}$. Si consideri la funzione misurabile definita q.c. da

$$E(f/\mathcal{G}) := \Sigma_n \frac{1_{B_n}}{\mu(B_n)} \int_{B_n} f d\mu \tag{1.3}$$

ove f è una v.a. tale che E(f) esista (e quindi non necessariamente in $L^1(\mathcal{F})=L^1(\Omega,\mathcal{F},\mu)$). La (1.3) definisce la SC di f data la tribù $\mathcal{G}$. Il condizionamento è dato dalla tribù $\mathcal{G}$ piuttosto che dagli insiemi della partizione $\{B_n\}$. Infatti ogni insieme B di $\mathcal{G}$ si scrive nella forma $B = \cup_k B_{n(k)}$ essendo $\{n(k)\}$ una successione, eventualmente finita, estratta da $\mathbb{N}$. ALlora in accordo con la (1.1) e la (1.3)

$$\mu(B)E_B(f) = \int_B f d\mu=\Sigma_k \int_{B_{n(k)}} f d\mu=\Sigma_k E_{B_{n(k)}}(f)\mu(B_{n(k)})= \int_B E(f/\mathcal{G})d\mu_{\mathcal{G}}$$

ove $\mu_{\mathcal{G}}$ indica la restrizione di $\mu$ a $\mathcal{G}$ . Dall'ultima relazione scritta conviene isolare l'eguaglianza

$$\int_B f d\mu = \int_B E(f/\mathcal{G})d\mu_{\mathcal{G}} \tag{1.4}$$

che costituirà la definizione nel caso generale, perché la (1.3) non è più utilizzabile se la tribù $\mathcal{G}$ non è generata da una partizione finita o numerabile.

## 1.2. LA DEFINIZIONE GENERALE

(2.1) TEOREMA. Sia $\mathscr{G}$ una tribù contenuta in $\mathscr{F}$ e sia $f:(\Omega,\mathscr{F})\to(\bar{\mathbb{R}},\mathscr{B})$ una v.a. di $L^1(\mathscr{F})$; esiste allora una funzione g di $L^1(\mathscr{G})$, unica a meno di equivalenza, tale che valga

$$\int_B f d\mu = \int_B g d\mu_{\mathscr{G}} \qquad \text{se} \quad B\in\mathscr{G} .$$

DIM. L'applicazione $B \to E(1_B f)$ definisce su $\mathscr{G}$ una misura reale assolutamente continua rispetto a $\mu_{\mathscr{G}}$. L'asserto è così una conseguenza del teorema di Radon-Nikodym.//

(2.2) DEFINIZIONE. Si dice SC della v.a. $f\in L^1(\mathscr{F})$ data la tribù $\mathscr{G}\subset\mathscr{F}$ l'unica funzione $\mathscr{G}$-misurabile $E(f/\mathscr{G})$ tale che

$$\int_B f \, d\mu = \int_B E(f/\mathscr{G}) \, d\mu_{\mathscr{G}} \qquad \text{se } B\in\mathscr{G} \; ; \tag{2.3}$$

l'unicità s'intende a meno di equivalenze.

Usualmente, e quando ciò non generi confusione, si confonderanno la misura $\mu$ e la sua restrizione $\mu_{\mathscr{G}}$ alla tribù $\mathscr{G}$. Quando convenga mettere in evidenza la natura di operatore della SC si scriverà, talvolta, $E^{\mathscr{G}}$ in luogo di $E(./\mathscr{G})$.

In qualche caso è possibile estendere la definizione di SC a v.a. che non ammettono speranza finita, quindi non di $L^1(\mathscr{F})$, ma per le quali sia $|E(f)| = \infty$ .

Si vedrà nel seguito (corollario (3.8)) che la definizione (2.3) coincide con quella della sezione precedente quando $\mathscr{G}$ sia generata da una partizione al più numerabile.

Il teorema che segue consente di definire la probabilità condizionata data una tribù $\mathcal{G}$ e di stabilire il legame con le SC.

**(2.4) TEOREMA.** Esiste un'unica funzione $\mu(./\mathcal{G}) \in L^1(\mathcal{G})$ detta probabilità condizionata data la tribù $\mathcal{G} \subset \mathcal{F}$, tale che

$$(2.5) \qquad \mu(A \cap B) = \int_B \mu(A/\mathcal{G})\, d\mu_{\mathcal{G}} \qquad \text{se } B \in \mathcal{G},\ A \in \mathcal{F}.$$

Vale inoltre

$$(2.6) \qquad \mu(A/\mathcal{G}) = E(1_A/\mathcal{G})$$

DIM. Ponendo $\lambda(B) := \mu(A \cap B)$ si definisce su $\mathcal{G}$ una misura assolutamente continua rispetto a $\mu_{\mathcal{G}}$. Il teorema di Radon-Nikym assicura l'esistenza e l'unicità (a meno di equivalenze) di $\mu(./\mathcal{G})$. Ponendo $f = 1_A$ nella (2.3) si ottiene la (2.6) in virtù dell'unicità q.c. sia di $E(./\mathcal{G})$ sia di $\mu(./\mathcal{G})$.//

Talvolta si scriverà $\mu^{\mathcal{G}}$ in luogo di $\mu(./\mathcal{G})$.

In maniera analoga si dimostrano i due teoremi seguenti che consentono di definire le speranze e le probabilità condizionate da un evento di probabilità anche nulla. Per le dimostrazioni si può consultare [2] 6.3.1, 6.3.3, 6.3.4..

**(2.7) TEOREMA.** Sia g una v.a. di $L^1(\mathcal{F})$ e $f : (\Omega, \mathcal{F}) \to (\Omega', \mathcal{F}')$ una v.a. a valori in $\Omega'$. Esiste allora un'unica funzione $\varphi : (\Omega', \mathcal{F}') \to (\bar{\mathbb{R}}, \mathcal{B})$ tale che

$$\int_{f^{-1}(A)} g\, d\mu = \int_A g(\omega')\, d\mu_f(\omega') \qquad \text{se } A \in \mathcal{F}'$$

ove $\mu_f := \mu \circ f^{-1}$ è la legge della v.a. f. Si pone $g(\omega') = E(g/f = \omega')$.

(2.8) TEOREMA. Sia f come nel teorema (2.7) e sia A un insieme di $\mathscr{F}$. Esiste allora un'unica funzione $k:(\Omega',\mathscr{F}')\to(\mathbb{R},\mathscr{B})$ tale che

$$\mu\{f^{-1}(B)\cap A\} = \int_B k(\omega')d\mu_f(\omega') \qquad \text{se } B\in\mathscr{F}'.$$

Risulta inoltre, se $\mu(A|f=\omega'):=k(\omega')$, $\mu(A|f=\omega') = E(1_A|f=\omega')$.

Le proprietà di $E(g/f=.)$ e di $\mu(A/f=.)$ sono del tutto analoghe a quelle di $E(g/\mathscr{G})$ e di $\mu(A/\mathscr{G})$ e non saranno perciò date esplicitamente.

La notazione $E(g/f)$, che si incontra spesso, indica la SC $E(g/\mathscr{G}$ ove $\mathscr{G}$ è la tribù indotta da f; se $f:(\Omega,\mathscr{F})\to(\Omega',\mathscr{F}')$ è una v.a. riesce $\mathscr{G}=\sigma(f):=\{B\subset\Omega:B=f^{-1}(A),\ A\in\mathscr{F}'\}$.

Nel seguito servirà il seguente teorema, di natura tecnica, sulle tribù indotte.

(2.9) TEOREMA. Se $f:(\Omega,\mathscr{F})\to(\Omega',\mathscr{F}')$ è una v.a., per una funzione $\varphi:\Omega\to\mathbb{R}$, sono equivalenti le proprietà:

(a) $\varphi$ è misurabile rispetto a $\sigma(f)$ e a $\mathscr{B}$;

(b) esiste una funzione $g:(\Omega',\mathscr{F}')\to(\mathbb{R},\mathscr{B})$ tale che $\varphi=g\circ f$.

DIM. Basta dimostrare l'implicazione (a) $\Longrightarrow$ (b). Se $\varphi=1_B$ con $B\in\sigma(f)$ allora $B=f^{-1}(A)$ con $A\in\mathscr{F}'$, sicché, se $g:=1_A$, riesce $g\circ f=1_{f^{-1}(A)}=1_B=\varphi$. L'asserto è dunque vero se $\varphi$ è un indicatore. Se poi $\varphi$ è una funzione semplice $\varphi=\Sigma_{i=1}^n\lambda_i 1_{B_i}$ con $\lambda_i\in\mathbb{R}$ e $B_i\in\sigma(f)$ $(i=1,2,\dots,n)$ allora $1_{B_i}=g_i\circ f$ come sopra, con $g_i:=1_{f^{-1}(A_i)}$ se $B_i=f^{-1}(A_i)$ e $A_i\in\mathscr{F}'$; quindi $\varphi=\Sigma_{i=1}^n\lambda_i g_i$.

Se $\varphi\geq 0$ sia $\{\varphi_n\}$ una successione di funzioni semplici con

$\varphi_n \uparrow \varphi$. Per quanto appena visto, è $\varphi_n = g_n \circ f$. Si definisca $g := \lim_n g_n$ ove tale limite esiste, $g = 0$ altrove; perciò $\varphi = \lim_n \varphi_n = \lim_n g_n \circ f = g \circ f$.

Il caso generale segue ponendo $\varphi = \varphi^+ - \varphi^-$.//

## 1.3. PROPRIETA' DELLE SPERANZE CONDIZIONATE.

Nei teoremi di questa sezione f denoterà sempre una v.a. di $L^1(\mathscr{F})$ e $\mathscr{G}$ una tribù contenuta in $\mathscr{F}$.

(3.1) TEOREMA. (a) $E\left[E(f|\mathscr{G})\right] = E(f)$;

(b) f $\mathscr{G}$-misurabile $\Longrightarrow E(f|\mathscr{G}) = f$;

(c) $f = a$ q.c. $\Longrightarrow E(f/\mathscr{G}) = a$ (in particolare $E(1/\mathscr{G})=1$);

(d) $f \geq 0 \Longrightarrow E(f/\mathscr{G}) \geq 0$;

(e) $f_1 \geq f_2 \Longrightarrow E(f_1/\mathscr{G}) \geq E(f_2/\mathscr{G})$;

(f) $|E(f/\mathscr{G})| \leq E(|f|/\mathscr{G})$;

(g) $\mathscr{N} = \{\emptyset, \Omega\} \Longrightarrow E(f|\mathscr{N}) = E(f)$;

(h) $c_1, c_2 \in \mathbb{R}$, $f_1, f_2 \in L^1(\mathscr{F}) \Longrightarrow E(c_1 f_1 + c_2 f_2/\mathscr{G}) = c_1 \; E(f_1/\mathscr{G}) + c_2 E(f_2/\mathscr{G})$.

DIM. (a) Basta prendere $B=\Omega$ nella (2.3). Le proprietà (b) e (c) scendono dall'unicità delle derivate di Radon-Nikodým.

(d) Per ogni $A \in \mathscr{G}$ riesce $\int_A f d\mu \geq 0$; (c) Per ogni $A \in \mathscr{G}$ riesce $\int_A f_1 \, d\mu \geq \int_A f_2 d\mu$; (f) basta applicare la (c) all'ovvia diseguaglianza $-|f| \leq f \leq |f|$. (g) $\int_A f \, d\mu = \int_A E(f) d\mu$ se $A=\emptyset$ o se $A=\Omega$ ; l'asserto segue ora dall'unicità di $E(f/\mathscr{N})$.

(h) Per ogni $A \in \mathscr{G}$ si ha

$$\int_A E(c_1 f_1 + c_2 f_2/\mathscr{G}) d\mu = \int_A (c_1 f_1 + c_2 f_2) d\mu = c_1 \int_A f_1 d\mu + c_2 \int_A f_2 d\mu =$$

$$= c_1 \int_A E(f_1/\mathscr{G}) d\mu + c_2 \int_A E(f_2/\mathscr{G}) d\mu = \int_A \{c_1 E(f_1/\mathscr{G}) + c_2 \; E(f_2/\mathscr{G})\} d\mu. //$$

(3.2) OSSERVAZIONE. Le proprietà (3.1)(a) e (3.1)(f) mostrano che l'operatore $E^{\mathscr{G}}: f \to E(f/\mathscr{G})$ è a valori in $L^1(\mathscr{G})$, cioè $E^{\mathscr{G}}: L^1(\mathscr{F}) \to L^1(\mathscr{G})$; esso è anzi lineare per la (3.1)(h), e continuo. E' anzi una contrazione, come si vede immediatamente dalla (3.1)(f) integrando: $\|E(f/\mathscr{G})\|_1 \leq \|f\|_1$. Ciò rende interessante e naturale lo studio di $E^{\mathscr{G}}$ come operatore su $L^1(\mathscr{F})$. Tale studio sarà intrapreso più avanti.

(3.3) **TEOREMA**(a) Sia $f_n \geq 0$ $(n\in\mathbb{N})$ e $f_n \uparrow f$; allora $E(f_n/\mathscr{G}) \uparrow E(f/\mathscr{G})$
(b) Se $f_n \geq 0$ $(n\in\mathbb{N})$, allora $E(\Sigma_{n\in\mathbb{N}} f_n/\mathscr{G}) = \Sigma_{n\in\mathbb{N}} E(f_n/\mathscr{G})$
(c) Se $\{B_n \in\mathbb{N}\} \subset \mathscr{F}$ e $B_i \cap B_j = \emptyset$ $(i\neq j)$ allora $\mu(\cup_{n\in\mathbb{N}} B_n(\mathscr{G}) = \Sigma_{n\in\mathbb{N}} \mu(B_n/\mathscr{G})$.

DIM. (a) Per ogni $A\in\mathscr{G}$ riesce $\int_A f_n d\mu = \int_A E(f_n/\mathscr{G})d\mu$. Grazie al teorema (3.1)(c) la successione $\{E(f_n/\mathscr{G})\}$ tende crescendo ad una funzione $\varphi$ necessariamente $\mathscr{G}$-misurabile. Per il teorema di convergenza monotona si ha perciò

$$\int_A f d\mu = \int_A \varphi d\mu$$

onde $\varphi = E(f/\mathscr{G})$.

(b) E' immediata conseguenza di (a) e di (3.1)(h). La (c) si ottiene dalla (b) ponendo $f_n = 1_{B_n}$.//

(3.4) **TEOREMA** (di convergenza dominata). Se $|f_n| \leq g$ $(n\in\mathbb{N})$ con $g\in L^1(\mathscr{F})$ e $f_n \to f$ allora $E(f_n/\mathscr{G}) \to E(f/\mathscr{G})$.

DIM. Posto $g_n := \sup\{|f_n - f| : k \geq n\}$ si ha $g_n \downarrow 0$. Si osservi che $E(f/\mathscr{G})$ e $E(f_n/\mathscr{G})$ appartengono a $L^1(\mathscr{G})$ e sono q.c. finite per ogni $n\in N$. Inoltre $|E(f_n/\mathscr{G}) - E(f/\mathscr{G})| \leq E(|f_n - f|/\mathscr{G}) \leq E(g_n/\mathscr{F})$ sicché basta mostrare che $E(g_n/\mathscr{G}) \to 0$. Esiste una funzione $\varphi$, $\mathscr{G}$-misurabile, tale che $E(g_n/\mathscr{G}) \downarrow \varphi$. Poiché $0 \leq g_n \leq 2g$ il teorema di convergenza dominata dà

$$0 \leq \int \varphi d\mu \leq \int E(g_n/\mathscr{G}) d\mu = \int g_n \, d\mu \to 0$$

onde $\varphi = 0$.//

Valgono per le SC gli analoghi del teorema di convergenza monotona e dei lemmi di Fatou.

(3.5) TEOREMA. Sia $f_n \geq g \in L^1(\mathscr{F})$ $(n \in \mathbb{N})$ con $E(g) > -\infty$;

(a) se $f_n \uparrow f$ allora $E(f_n/\mathscr{G}) \uparrow E(f/\mathscr{G})$;

(b) $\liminf_n E(f_n/\mathscr{G}) \geq E(\liminf_n f_n/\mathscr{G})$.

Sia $f_n \leq g \in L^1(\mathscr{F})$ $(n \in \mathbb{N})$ con $E(g) < +\infty$;

(c) se $f_n \downarrow f$ allora $E(f_n/\mathscr{G}) \downarrow E(f/\mathscr{G})$;

(d) $\limsup_n E(f_n/\mathscr{G}) \leq E(\limsup_n f_n/\mathscr{G})$.



DIM. Basta dimostrare le proprietà (a) e (b) perché la (c) e la (d) scendono da quelle cambiando il segno a tutte le v.a. che intervengono.

(a) Poiché $0 \leq f_n^+ \uparrow f^+$ si ha $E(f_n^+/\mathscr{G}) \uparrow E(f/\mathscr{G})$ per il teorema (3.3)(a). Inoltre da $0 \leq f_n^- \leq g^- \in L^1(\mathscr{F})$ e dal teorema (3.4) scende $E(f_n^-/\mathscr{G}) \to E(f^-/\mathscr{G})$ sicché $E(f_n/\mathscr{G}) = E(f_n^+/\mathscr{G}) - E(f_n^-/\mathscr{G}) \to E(f^+/\mathscr{G}) - E(f^-/\mathscr{G}) = E(f/\mathscr{G})$.

(b) Posto $f_n' := \inf\{f_k : k \geq n\}$ si ha $f_n' \uparrow f' := \liminf f_n$ onde per la prima parte del teorema $E(f'/\mathscr{G}) = E(\liminf_n f_n/\mathscr{G}) = \liminf_n E(f_n'/\mathscr{G}) \leq \leq \liminf_n E(f_n/\mathscr{G})$.//

Le proprietà seguenti si riferiscono tutte a condizioni di regolarizzazione. Ad esse occorre premettere il seguente lemma di carattere tecnico.

(3.6) **LEMMA.** Siano $\mathcal{G}$ una tribù di sottoinsiemi di $\Omega$, $\nu$ una misura su $\mathcal{G}$, $f : \Omega \to \bar{\mathbb{R}}$ una funzione $\mathcal{G}$-misurabile e B un atomo di $\mathcal{G}$ rispetto a $\nu$, cioè $B \in \mathcal{G}$, $\nu(B) > 0$ e se $A \subset B$ con $A \in \mathcal{G}$ allora risulta $\nu(A)=0$ oppure $\nu(B-A)=0$. La funzione f è costante q.c. su B.

DIM. Sia $x \in \bar{\mathbb{R}}$ tale che $\nu(B \cap \{f < x\})=0$; di conseguenza risulta $\nu(B \cap \{f < y\}) = 0$ per ogni $y \leq x$. Posto $k := \sup\{x \in \bar{\mathbb{R}} : \nu(B \cap \{f < x\})=0\}$ si ha:

$$\nu(B \cap \{f<k\}) = \nu[\bigcup_{\substack{r \in \mathbb{Q} \\ r<k}} (B \cap \{f<r\})] = 0.$$

Se $x > k$ risulta $\nu(B \cap \{f < x\}) > 0$ onde, poiché B è un atomo, $\nu(B \cap \{f \geq x\}) = 0$. Perciò

$$\nu(B \cap \{f>k\}) = \nu[\bigcup_{\substack{r \in \mathbb{Q} \\ r \ k}} (B \cap f\{\geq r\})] = 0.$$

Pertanto $f=k$ $\nu$-q.c. in B.//

Siamo ora in grado di mostrare, come già anticipato, che quando la tribù $\mathcal{G}$ sia generata da una partizione al più numerabile di atomi, $E(f/\mathcal{G})$ coincide con la costruzione della sezione 1.

(3.7) **TEOREMA.** Sia B un atomo di $\mathcal{G}$ rispetto a $\mu$ e sia $f \in L^1(\mathcal{F})$. Allora

$$E(f/\mathcal{G}) = \frac{1}{\mu(B)} \int_B f d\mu \qquad \text{q.c. in B.}$$

DIM. Per il lemma (3.6) $E(f/\mathscr{G})$ è q.c. costante in B. Perciò la (2.3) dà $E(f/\mathscr{G})\,\mu(B) = \int_B f\,d\mu$.//

(3.8) COROLLARIO. Se $\pi = \{B_n\}$ è una partizione $\mathscr{F}$-misurabile di $\Omega$ con $\mu(B_n)>0$ $(n\in\mathbb{N})$ e se $\mathscr{G}$ è la tribù generata da $\pi$, $\mathscr{G}=\sigma(\pi)$, vale la (1.3).

(3.9) TEOREMA. Siano f e g due v.a. su $(\Omega,\mathscr{F},\mu)$ tali che $f\in L^1(\mathscr{F})$, $fg\in L^1(\mathscr{F})$ e g sia $\mathscr{G}$-misurabile. Allora

$$E(fg/\mathscr{G}) = g\,E(f/\mathscr{G}). \tag{3.10}$$

DIM. Si supponga dapprima che g sia una funzione indicatrice, $g = 1_B$ con $B\in\mathscr{G}$: $\int_A E(fg/\mathscr{G})d\mu = \int_A fg\,d\mu = \int_{A\cap B} f d\mu = \int_{A\cap B} E(f/\mathscr{G})d\mu = \int_A 1_B E(f/\mathscr{G})d\mu$

$$= \int_A g\,E(f/\mathscr{G})d\mu$$

per ogni insieme A di $\mathscr{G}$, sicché la (3.10) vale per le funzioni indicatrici e quindi, per linearità, anche per le funzioni semplici. Ricorrendo al teorema di convergenza monotona (3.5)(a) si dimostra la (3.10) per le funzioni $\mathscr{G}$-misurabili positive e infine, mediante la decomposizione, $g=g^+-g^-$ per tutte le funzioni $\mathscr{G}$-misurabili.//

Un caso particolare della (3.10) avrà particolare importanza nelle caratterizzazioni delle SC: se $f_1,f_2\in L^1(\mathscr{F})$ e se anche $f_1f_2\in L^1(\mathscr{F})$ allora

$$E[f_1E(f_2/\mathscr{G})/\mathscr{G}] = E(f_1/\mathscr{G})\;E\,(f_2/\mathscr{G}). \tag{3.11}$$

(3.12) OSSERVAZIONE. Se $\mathscr{G}_1$ e $\mathscr{G}_2$ sono due sottotribù distinte di $\mathscr{F}$ , in generale gli operatori $E^{\mathscr{G}_1}$ e $E^{\mathscr{G}_2}$ non commutano come mostra il seguente esempio, tratto da [36]. Sia $\{B_1,B_2,B_3\}$ una partizione misurabile di $\Omega$ e siano $\mathscr{G}_1 = \sigma(\{B_1 \cup B_2, B_3\})$, $\mathscr{G}_2 = \sigma(\{B_1, B_2 \cup B_3\})$ e $f = 1_{B_3}$. Allora $E^{\mathscr{G}_2} E^{\mathscr{G}_1} f = E^{\mathscr{G}_2} f$ che non è $\mathscr{G}_1$-misurabile e non può perciò coincidere con $E^{\mathscr{G}_1} E^{\mathscr{G}_2} f$. I due teoremi che seguono danno invece due importanti esempî di situazioni nelle quali $E^{\mathscr{G}_1}$ e $E^{\mathscr{G}_2}$ commutano.

(3.13) **TEOREMA.** Siano $\mathscr{G}_1$ e $\mathscr{G}_2$ due sottotribù di $\mathscr{F}$ con $\mathscr{G}_1 \subset \mathscr{G}_2$ e sia $f \in L^1(\mathscr{F})$. Allora gli operatori $E^{\mathscr{G}_1}$ e $E^{\mathscr{G}_2}$ commutano e vale

$$E^{\mathscr{G}_2} E^{\mathscr{G}_1} f = E^{\mathscr{G}_1} E^{\mathscr{G}_2} f = E^{\mathscr{G}_1} f. \tag{3.14}$$

DIM. Poiché $E^{\mathscr{G}_1} f$ è misurabile rispetto a $\mathscr{G}_1$ e quindi a $\mathscr{G}_2$ si ha per la (3.10),

$$E^{\mathscr{G}_2} E^{\mathscr{G}_1} f = (E^{\mathscr{G}_1} f)(E^{\mathscr{G}_2} 1) = E^{\mathscr{G}_1} f.$$

Inoltre per ogni $A \in \mathscr{G}_1$ si ha

$$\int_A E^{\mathscr{G}_1} f \, d\mu = \int_A f d\mu = \int_A E^{\mathscr{G}_2} f \, d\mu = \int_A E^{\mathscr{G}_1} E^{\mathscr{G}_2} f \, d\mu \quad .//$$

Si osservi che se $\mathscr{G}_1 = \mathscr{N}$ e $\mathscr{G}_2 = \mathscr{G}$ la (3.10) dà, poiché $E^{\mathscr{N}} = E$ (teorema (3.1)(g)), la (3.1)(a).

Se $\mathscr{G}_1 = \mathscr{G}_2 = \mathscr{G}$ la (3.14) dà $E^{\mathscr{G}} E^{\mathscr{G}} = E^{\mathscr{G}}$ sicché $E^{\mathscr{G}}$ è una proiezione da $L^1(\mathscr{F})$ in $L^1(\mathscr{G})$.

Due tribù $\mathscr{F}_1$ e $\mathscr{F}_2$ entrambe contenute in $\mathscr{F}$ si dicono <u>indipendenti</u> (rispetto alla probabilità $\mu$) se risulta $\mu(A \cap B) = \mu(A)\ \mu(B)$ per ogni scelta di A in $\mathscr{F}_1$ e di B in $\mathscr{F}_2$. Una tribù $\mathscr{G} \subset \mathscr{F}$ e una v.a. f su $(\Omega,\mathscr{F},\mu)$ si dicono indipendenti se sono tali $\mathscr{G}$ e la tribù $\sigma(f)$ generata da f. Infine due v.a. f e g su $(\Omega,\mathscr{F},\mu)$ si dicono indipendenti se sono tali le tribù che generano, $\sigma(f)$ e $\sigma(g)$.

(3.15) **TEOREMA.** Se la v.a. f di $L^1(\mathscr{F})$ è indipendente dalla tribù $\mathscr{G}$ si ha $E(f/\mathscr{G}) = E(f)$.

DIM. Poiché $E(f) \in \mathbb{R}$, riesce, in virtù dell'indipendenza di f e di $\mathscr{G}$, $\int_A E(f)d\mu = E(f)\,\mu(A) = E(f)E(1_A) = E(f1_A) = \int_A f d\mu = \int_A E(f/\mathscr{G})d\mu$ quale che sia $A \in \mathscr{G}$; perciò $E(f/\mathscr{G}) = E(f)$.//

(3.16) **TEOREMA.** Se $\mathscr{G} = \sigma(g)$ ove g è una v.a. esiste una funzione borelliana $\varphi_g : \mathbb{R} \to \mathbb{R}$ tale che $E(f/\mathscr{G}) = \varphi_g \circ f$.

DIM. Per definizione $E(f/\mathscr{G})$ è misurabile rispetto a $\sigma(g)$ e quindi, per il teorema (2.9) esiste la funzione della tesi.//

Altre proprietà delle SC sono legate alla diseguaglianza di Jensen e quindi alle proprietà delle funzioni convesse che sono qui succintamente richiamate.

Sia $\varphi: I \to \mathbb{R}$ (I è un intervallo di $\mathbb{R}$) <u>convessa</u>, soddisfaccia cioè, per ogni $\lambda \in [0,1]$ e per ogni coppia $x,y \in I$ alla diseguaglianza

(3.17) $$\lambda\varphi(x) + (1-\lambda)\varphi(y) \geq \varphi(\lambda x+(1-\lambda)y);$$

si dice che $\varphi$ è <u>strettamente convessa</u> se nella (3.17) vale la diseguaglianza stretta se $\lambda \in ]0,1[$. E' noto che una funzione

$\varphi$ è convessa se, e solo se, il suo rapporto incrementale relativo ad un qualsiasi punto $x_0$ di I, $x \to [\varphi(x)-\varphi(x_0)]/(x-x_0)$ è crescente. Perciò $\varphi$ è continua ed ammette derivata a destra $\varphi'_d$ e derivata a sinistra $\varphi'_s$ in ogni punto interno di I. Inoltre $\varphi$ è derivabile in I, tranne, al più, in un'infinità munerabile di punti. Per tutte queste (e altre) proprietà si veda, ad esempio [21]. La funzione $\varphi$ è convessa in $[a,b]$ se, e solo se, esiste una funzione $\varphi$ crescente $g : [a,b] \to \mathbb{R}$ tale che

$$\varphi(x) = \varphi(a) + \int_a^x g(t)dt. \tag{3.18}$$

Si può prendere $g = \varphi'_d$; g è strettamente crescente se $\varphi$ è strettamente convessa.

Se $\varphi$ è convessa nell'intervallo aperto $]a,b[$ $(a,b \in \bar{\mathbb{R}})$ esistono due successioni di numeri reali $\{a_n\}$ e $\{b_n\}$ tali che, se $x \in ]a,b[$, riesca

$$\varphi(x) = \sup\{a_n x + b_n : n \in \mathbb{N}\}. \tag{3.19}$$

Quest'ultimo risultato è noto come <u>teorema della linea di supporto</u> (si veda, per esempio, [2] pp. 286-287).

**(3.20) TEOREMA.** Sia $\varphi : I \to \bar{\mathbb{R}}$ con I intervallo aperto (limitato o illimitato) di $\mathbb{R}$ e sia f una v.a. di $L^1(\mathscr{F})$ a valori in I. Allora vale la diseguaglianza di Jensen

$$E(\varphi \circ f/\mathscr{G}) \geq \varphi \circ E(f/\mathscr{G}) \tag{3.21}$$

Se, inoltre, $\varphi$ è strettamente convessa nella (3.21) vale il segno d'eguaglianza se, e solo se, f è $\mathscr{G}$-misurabile.

DIM. In primo luogo risulta $E(f/\mathscr{G}) \in I$. Infatti, se, per esempio, è f>a si ha

$$0 \geq \int_{\{E(f/\mathscr{G}) \leq a\}} [E(f/\mathscr{G})-a]\,d\mu = \int_{\{E(f/\mathscr{G}) \leq a\}} (f-a)\,d\mu \geq 0$$

sicché f=a q.c. in $\{E(f/\mathscr{G}) \leq a\}$ ; perciò $\mu\{E(f/\mathscr{G}) \leq a\} = 0$ cioè $E(f/\mathscr{G}) > a$ q.c.. Scende dalla (3.19) che $\varphi\circ f \geq a_n f+b_n$ $(n\in\mathbb{N})$ onde $E(\varphi\cdot f/\mathscr{G}) \geq a_n\, E(f/\mathscr{G})+b_n$ $(n\in\mathbb{N})$. La (3.21) si ottine prendendo l'estremo superiore in quest'ultima diseguaglianza.

Si supponga ora che $\varphi$ sia strettamente convessa e che valga

$$E(\varphi\circ f/\mathscr{G}) = \varphi\circ E(f/\mathscr{G}) \quad . \tag{3.22}$$

Segue dalla (3.18) che $\varphi(x)-\varphi(a) \geq (x-a)\ \varphi'_d(a)$ relazione nella quale vale, anzi, la diseguaglianza stretta se $x\neq a$; perciò

$$\varphi\circ f - \varphi\circ E(f/\mathscr{G}) \geq \{f-E(f/\mathscr{G})\}[\varphi'_d\circ E(f/\mathscr{G})] \tag{3.23}$$

onde integrando, e tenendo presente la (3.22),

$$\begin{aligned}\int \varphi\circ f\,d\mu &\geq \int\varphi\circ E(f/\mathscr{G})d\mu + \int \varphi'_d\circ E(f/\mathscr{G})\,\{f-E(f/\mathscr{G})\}\,d\mu = \\ &= \int E(\varphi\circ f/\mathscr{G})d\mu + \int E^{\mathscr{G}}\,[\varphi'_d\circ E(f/\mathscr{G})\,\{f-E(f/\mathscr{G})\}]\,d\mu = \\ &= \int \varphi\circ f\,d\mu + \int\{\varphi'_d\circ E(f/\mathscr{G})\}\,E^{\mathscr{G}}[f-E(f/\mathscr{G})]\,d\mu = \int\varphi\circ f\,d\mu.\end{aligned}$$

Nell'ultima relazione valgono allora solo segni d'eguaglianza e pertanto, tenendo conto della (3.23) si ottiene

$$0 \leq \int\{\varphi\circ f - \varphi\circ E(f/\mathscr{G}) - [\varphi'_d\circ E(f/\mathscr{G})]\,[f-E(f/\mathscr{G})]\}\,d\mu = 0.$$

Poiché, in virtù della (3.23), l'integrando è positivo, essa si annulla q.c.. Ma $\varphi$ è strettamente convessa e perciò, nella (3.23), vale il segno di diseguaglianza stretta, a meno che non sia $f=E(f/\mathcal{G})$, cioè a meno che f non sia $\mathcal{G}$-misurabile.//

La seguente conseguenza della diseguaglianza di Jensen è particolarmente importante ed estende al caso $p > 1$, una proprietà già vista per p=1 (osservazione (3.2)).

(3.24) **TEOREMA.** Per ogni tribù $\mathcal{G} \subset \mathcal{F}$, l'operatore $E^{\mathcal{G}}$ su $L^p(\mathcal{F})$ con $p\in[1,+\infty]$ è una proiezione contrattiva e positiva di immagine $L^p(\mathcal{G})$.

DIM. Si è già visto che $E^{\mathcal{G}}$ è una proiezione positiva; esso è anche una contrazione. Infatti se $\varphi : x \to |x|^p$ con $p\in[1,+\infty[$ il teorema (3.20) dà per $f\in L^p(\mathcal{F})$

$$\|f\|_p^p = \int |f|^p d\mu = \int \varphi\circ f \, d\mu = \int E^{\mathcal{G}}(\varphi\circ f) d\mu \geq \int \varphi\circ E^{\mathcal{G}} f d\mu = \|E^{\mathcal{G}} f\|_p^p \ .$$

Se invece $p=+\infty$ riesce $|f| \leq \|f\|_\infty$ sicché

$$|E^{\mathcal{G}} f| \leq E^{\mathcal{G}}(|f|) \leq E^{\mathcal{G}}(\|f\|_\infty) = \|f\|_\infty$$

e quindi $\|E^{\mathcal{G}} f\|_\infty \leq \|f\|_\infty$ . Ciò mostra che $E^{\mathcal{G}}$ è una contrazione anche in $L^\infty(\mathcal{F})$. Ora, se $p\in[1,+\infty]$, si ha $E^{\mathcal{G}} f\in L^p(\mathcal{G})$ per ogni $f\in L^p(\mathcal{F})$ ma siccome $E^{\mathcal{G}}$ ristretto a $L^p(\mathcal{G})$ coincide con l'identità su $L^p(\mathcal{G})$ si ha $E^{\mathcal{G}} L^p(\mathcal{F}) = L^p(\mathcal{G})$.//

L'ultima delle proprietà delle quali ci occuperemo è legata al comportamento delle SC rispetto alle trasformazioni che conservano la misura.

Si dice che la trasformazione $T : \Omega \to \Omega$ conserva la misura se essa soddisfa alle due condizioni:

(a) è misurabile, cioè $T^{-1}A\in\mathcal{F}$ per ogni $A\in\mathcal{F}$ (in breve $T^{-1}\mathcal{F}\subset\mathcal{F}$);

(b) $\mu(T^{-1}A) = \mu(A)$ per ogni $A\in\mathcal{F}$.

**(3.25) TEOREMA.** Se $T:\Omega\to\Omega$ conserva la misura, per ogni $f\in L^1(\mathcal{F})$ si ha

$$(E^{\mathcal{G}}f)\circ T = E^{T^{-1}\mathcal{G}}(f\circ T).$$

DIM. Sfruttando il teorema del cambio di variabili (per esempio [23] teorema 6.8) si ha per ogni $A\in\mathcal{G}$

$$\int_{T^{-1}A} E^{T^{-1}\mathcal{G}}(f\circ T)d\mu = \int_{T^{-1}A} f\circ T\, d\mu = \int 1_{T^{-1}A}(f\circ T)d\mu = \int (1_A\circ T)(f\circ T)\, d\mu =$$

$$= \int 1_A f\, d(\mu\circ T^{-1}) = \int 1_A f\, d\mu = \int_A f\, d\mu = \int_A E^{\mathcal{G}} f\, d\mu =$$

$$= \int_A E^{\mathcal{G}} f\, d(\mu\circ T^{-1}) = \int_{T^{-1}A}(E^{\mathcal{G}} f)\circ T\, d\mu\ .//$$

## 1.4. UN APPROCCIO DIFFERENTE

La definizione di SC può essere data in maniera differente ma equivalente (teorema (4.3)). E' questo l'approccio di [7].

Se $\mathcal{G}$ è una sottotribù di $\mathcal{F}$, è evidente che $L^2(\mathcal{G})$ è un sottospazio di $L^2(\mathcal{F})$.

**(4.1) DEFINIZIONE.** Sia $E: L^2(\mathcal{G})\to L^2(\mathcal{F})$ l'operatore di immersione di $L^2(\mathcal{G})$ in $L^2(\mathcal{F})$ e sia $E^*: L^{2*}(\mathcal{F}) = L^2(\mathcal{F}) \to L^{2*}(\mathcal{G}) = L^2(\mathcal{G})$ il suo aggiunto. Si definisca allora come operatore di SC rispetto alla tribù $\mathcal{G}$ l'operatore $E^{\mathcal{G}}: L^2(\mathcal{F})\to L^2(\mathcal{G})$ definito da

$$E^{\mathcal{G}} := EE^*. \tag{4.2}$$

Si osservi che, salvo dimostrare che la definizione appena data coincide con quella della seqione 2, $E^{\mathscr{G}}$ è qui definita come un operatore su $L^2(\mathscr{F})$ anziché su $L^1(\mathscr{F})$. Ciò corrisponde a definire $E^{\mathscr{G}}$ <u>in senso ampio</u>; la definizione <u>in senso stretto</u> considera E come un operatore su $L^1(\mathscr{F})$. Tale distinzione è dovuta a Doob. Sui rapporti tra le definizioni in senso ampio e stretto si tornerà nel seguito, sezione 2.3, dove si vedrà che il dominio di $E^{\mathscr{G}}$ si può estendere a $L^1(\mathscr{F})$.

L'operatore definito dalla (4.2) è chiaramente autoaggiunto.

**(4.3) TEOREMA.** Le definizioni (4.1) e (2.2) coincidono in $L^2(\mathscr{F})$.

DIM. Per ogni f di $L^2(\mathscr{F})$, $EE^*f$ appartiene a $L^2(\mathscr{G})$ sicché $E^{\mathscr{G}}f$ è $\mathscr{G}$-misurabile. Inoltre per ogni $A\in\mathscr{G}$ è

$$\int_A EE^*f\,d\mu = \int(EE^*f)1_A d\mu = \langle EE^*f, 1_A\rangle = \langle f, EE^*1_A\rangle = \langle f, 1_A\rangle$$

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

L'operatore $EE^*$ soddisfa dunque alla (2.3); per l'unicità della derivata di Radon-Nikodým si ha $EE^* = E^{\mathscr{G}}$ .//

**(4.4) DEFINIZIONE.** Un operatore lineare $T : L^1(\mathscr{F}) \to L^1(\mathscr{F})$ si dice <u>doppiamente stocastico</u> se soddisfa alle condizioni (a) è positivo ($f \geq 0 \Longrightarrow Tf \geq 0$); (b) lascia invariata la speranza ($E(Tf) = E(f)$); (c) lascia invariate le costanti ($T1=1$).

Evidentemente un operatore di SC è doppiamente stocastico e alcune sue proprietà si possono ottenere per questa via come si farà qui di seguito.

**(4.5) TEOREMA.** Se $T:L^1(\mathscr{F}) \to L^1(\mathscr{F})$ è doppiamente stocastico allora (a) $TL^p(\mathscr{F}) \subset L^p(\mathscr{F})$ per ogni $p\in[1,+\infty]$; (b) $\|T\|_1=\|T\|_\infty=1$; (c) $\|T\|_p \leq 1$ se $p\in]1,+\infty[$; (d) l'operatore autoaggiunto $T^*$ è pure doppiamente stocastico.

DIM. Se $f\in L^\infty(\mathscr{F})$ si scriva $f=f^+-f^-$; per la (4.4)(a) è $Tf^+ \geq 0$ e $Tf^- \geq 0$ onde $|Tf|=|Tf^+-Tf^-| \leq Tf^++Tf^-=T|f|$; dalle (4.4)(b) e (c) scende $T|f| \leq T\|f\|_\infty = \|f\|_\infty$. Perciò $\|Tf\|_\infty \leq \|f\|_\infty$ cioè $Tf\in L^\infty(\mathscr{F})$ e $\|T\|_\infty \leq 1$; ma poiché $T1=1$ si ha $\|T\|_\infty = 1$.

Dalla (4.4)(b) segue, per ogni $f\in L^1(\mathscr{F})$ che $\langle f,1\rangle = \int f d\mu = \int Tf\, d\mu = \langle Tf,1\rangle = \langle f,T^*1\rangle$ onde $T^*1=1$. Come sopra si ha allora che $\|T^*\|_\infty = \|T\|_1 = 1$.

La diseguaglianza $\|T\|_p \leq 1$ per $p\in]1,+\infty[$ scende dal teorema di convessità di Riesz ([19] p.526).

Per concludere la dimostrazione basta far vedere che $T^*L^1(\mathscr{F}) \subset L^1(\mathscr{F})$ e che $E(T^*f) = E(f)$.

Se $f\in L^\infty(\mathscr{F})$ allora $\int T^*f d\mu = \langle T^*f,1\rangle = \langle f,T1\rangle = \langle f,1\rangle = \int f d\mu$. Poiché $L^\infty(\mathscr{F})$ è denso in $L^1(\mathscr{F})$, esiste un'unica estensione continua di $T^*$ a tutto $L^1(\mathscr{F})$ che soddisfaccia alla (4.4)(b).

Per dimostrare che $T^*L^1(\mathscr{F}) \subset L^1(\mathscr{F})$ si integri la diseguaglianza $|T^*f| \leq T^*|f|$ ottenendo $\|T^*f\|_1 \leq \int T^*|f| d\mu = \int |f| d\mu = \|f\|_1$.//

**(4.6) COROLLARIO.** $E^{\mathscr{G}} L^p(\mathscr{F}) \subset L^p(\mathscr{G})$ se $p\in[1,\infty]$; $\|E^{\mathscr{G}}\|_1=\|E^{\mathscr{G}}\|_\infty=1$ mentre $\|E^{\mathscr{G}}\|_p \leq 1$ se $p\in]1,\infty[$.

**(4.7) TEOREMA.** Sia $\{f_n\}$ una successione di $L^p(\mathscr{F})$ $p\in[1,\infty]$ limitata da una v.a. $g$ di $L^p(\mathscr{F})$, $|f_n| \leq g$ $(n\in\mathbb{N})$. Allora, se $f_n \to f$ q.c.

e se T è un operatore doppiamente stocastico, risulta $Tf_n \to Tf$ q.c. e in $L^p(\mathcal{F})$.

DIM. La convergenza in $L^p(\mathcal{F})$ scende dalla (c) del teorema (4.5) e dal teorema di convergenza dominata.

Si supponga ora che $\{f_n\}$ sia una successione monotona, per esempio $f_n \uparrow f$; è perciò monotona anche la successione $\{Tf_n\}$ che dunque converge a una v.a. $\varphi$, $TF_n \uparrow \varphi$, onde, per il teorema di convergenza monotona, $E(\varphi) = \lim_n E(Tf_n) = \lim_n E(f_n) = E(f) = E(Tf)$. In modo analogo, se A è in $\mathcal{G}$ si ottiene $\int_A \varphi \, d\mu = \int_A Tf \, d\mu$ che applicata all'insieme $A := \{Tf \neq \varphi\}$ dà $\mu(A)=0$ onde $Tf = \varphi$ q.c.. Il teorema di convergenza dominata segue poi nella maniera usuale da quello di convergenza monotona.//

**(4.8) TEOREMA** (diseguaglianza di Jensen). Siano $\varphi : [a,b] \to \mathbb{R}$ una funzione convessa e T un operatore doppiamente stocastico. Allora

(4.9) $$\varphi \circ Tf \le T(\varphi \circ f)$$

per ogni v.a. di $L^1(\mathcal{F})$ a valori in $[a,b]$.

DIM. Se f è una v.a. semplice $f = \Sigma_{i=1}^r \lambda_i 1_{A_i}$ ove $\{A_i\}$ è una partizione misurabile di $\Omega$ e $\lambda_i \in \mathbb{R}$ $(i=1,2,\ldots,r)$ allora $\varphi \circ f = \Sigma_{i=1}^r \varphi(\lambda_i) 1_{A_i}$ sicché $T(\varphi \circ f) = \Sigma \varphi(\lambda_i) T1_{A_i} \ge \varphi(\Sigma \lambda_i \, T1_{A_i}) = \varphi(T \Sigma \lambda_i 1_{A_i}) = \varphi \, Tf$.

Data $f \ge 0$ sia $\{f_n\}$ una successione di v.a. semplici tali che $f_n \uparrow f$; allora $Tf_n \uparrow Tf$ e, per la continuità di $\varphi$, $\varphi \circ Tf_n \to \varphi \circ Tf$ q.c.. D'altro canto, $|\varphi \circ f_n| \le \max\{|\varphi(t)| : t \in [a,b]\} =: c < +\infty$

sicché il teorema precedente dà $T(\varphi \circ f_n) \to T(\varphi \circ f)$ q.c. dalla quale scende da (4.9) passando al limite sulla diseguaglianza $T(\varphi \circ f_n) \geq$ $\geq \varphi \circ Tf_n$. //