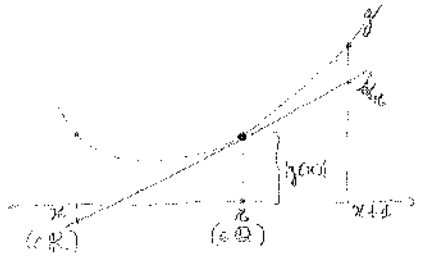
Approssimazione di una funzione convessa

Teorema. *Sia g una funzione reale convessa su $\mathbb{R}$. Allora g è l'inviluppo superiore di una famiglia numerabile di funzioni lineari affini.*

Dimostrazione. Per ogni numero razionale r, denotiamo con f_r la funzione che coincide con g in r e la cui derivata è costantemente eguale alla derivata destra di g calcolata in r. Essendo g convessa, f_r è maggiorata da g. Perciò, posto $h = \sup_{r\in\mathbb{Q}} f_r$, ha luogo la diseguaglianza $h \le g$. Basta dunque provare la diseguaglianza opposta. A questo scopo, fissato un numero reale x, osserviamo che, sull'intervallo $[x, x+1]$, la funzione g è lipschitziana, con costante di Lipschitz L eguale al massimo tra il modulo della derivata sinistra di g in x e il modulo della derivata destra di g in $x+1$. Anche ciascuna delle funzioni f_r, con $r \in [x, x+1] \cap \mathbb{Q}$, è lipschitziana su $[x, x+1]$ con la stessa costante L. Di conseguenza, per ogni elemento r di $[x, x+1]\cap\mathbb{Q}$, le due funzioni f_r e g, che coincidono in r, non possono differire, in x, per più di $2L(r-x)$. Si ha dunque

$$\begin{aligned} g(x) &\le f_r(x) + 2L(r-x) \\ &\le h(x) + 2L(r-x). \end{aligned}$$

Di qui, facendo tendere r a x, si ricava $g(x) \le h(x)$, e ciò conclude la dimostrazione. □



✡ (3) Una *tribù*, o σ-algebra, su un insieme E è un insieme $\mathcal{G} \subseteq \mathcal{P}(E)$ tale che $E \in \mathcal{G}$, $A \in \mathcal{G} \Rightarrow A^c \in \mathcal{G}$ ($\Rightarrow \emptyset \in \mathcal{G}$), e $(A_m)_{m\geq 1}$ in $\mathcal{G} \Rightarrow \bigcup_{m\geq 1} A_m \in \mathcal{G}$, $\Rightarrow \bigcap_{m\geq 1} A_m \in \mathcal{G}$ ($\Rightarrow \forall A, B \in \mathcal{G}$, $B \setminus A \in \mathcal{G}$); in tal caso $(E, \mathcal{G})$ è uno *spazio misurabile*. Un esempio: $\mathcal{G} = \mathcal{P}(E)$. Allora, $\forall \mathcal{G} \subseteq \mathcal{P}(E)$ arbitrario, $\sigma(\mathcal{G}) := \bigcap_{(\sigma\text{-alg. } \supseteq \mathcal{G})} \mathcal{G}$ è la più piccola tribù su E contenente $\mathcal{G}$, fra cioè chiamata "la tribù (su E) generata da $\mathcal{G}$". Nel caso speciale che (E, τ) sia uno spazio topologico, $\sigma(\tau) =: \mathcal{B}(E)$ prende il nome di "tribù di Borel su E". Per $E = \mathbb{R}$ consideriamo sempre $\mathcal{G} = \mathcal{B}(\mathbb{R})$. Una *misura* (positiva) μ sulla tribù $\mathcal{G}$ è una $\mu \in [0,\infty]^{\mathcal{G}}$ tale che $\exists A \in \mathcal{G}$. $\mu(A) < \infty$, e $(A_m)_{m\geq 1}$ in $\mathcal{G}$ e $\forall$ due a due disjunti $\Rightarrow \mu\left(\bigcup_{m\geq 1} A_m\right) = \sum_{m\geq 1} \mu(A_m)$, $\Rightarrow \forall A, B \in \mathcal{G}$, $\mu(B) = \mu(A \cap B) + \mu(A^c \cap B)$, $\Rightarrow \mu(\emptyset) = 0$; $\Rightarrow \forall A, B \in \mathcal{G}$, $A \subseteq B \Rightarrow \mu(B \setminus A) = \mu(B) - \mu(A)$ $\Rightarrow \mu(A) \leq \mu(B)$ (monotonia) e $\mu(A^c) = \mu(E) - \mu(A)$, per cui $\mu(B) = 0 \Rightarrow \mu(A) = 0$, $\Rightarrow \mu(A^c) = \mu(E)$; inoltre $(A_m)_{m\geq 1} \uparrow$, risp. $\downarrow$, in $\mathcal{G} \Rightarrow (\mu(A_m))_{m\geq 1} \uparrow \mu\left(\bigcup_{m\geq 1} A_m\right)$, risp. $\downarrow \mu\left(\bigcap_{m\geq 1} A_m\right)$, in $[0,\infty]$. Un esempio: $\forall A \in \mathcal{G}$, $\mu(A) := 0$; $\forall A \in \mathcal{G}$, $\mu(A) := \begin{cases} \infty & \text{se } |A| \geq \aleph_0 \\ |A| & \text{se } |A| \leq \aleph_0 \end{cases}$; più in generale $E \neq \emptyset$ e $x_0 \in E$ fissato: $\forall A \in \mathcal{G}$, $\mu(A) := \begin{cases} 1 & \text{se } x_0 \in A \\ 0 & \text{se } x_0 \notin A \end{cases}$, $=: \varepsilon_{x_0}(A)$. $(= \mathbb{1}_A(x_0))$

Comunque, in tal caso $(E, \mathcal{G}, \mu)$ è uno *spazio misurato*, o *misurabile*. Se in realtà $\mu(E) \in (0,\infty)$, allora $\nu := \mu(E)^{-1} \cdot \mu$ è una misura su $\mathcal{G}$ tale che $\nu(E) = 1$, $\Rightarrow$ e valori in $[0,1]$; $\Rightarrow \forall A \in \mathcal{G}$, $\nu(A^c) = 1 - \nu(A)$, per cui $\nu(A) = 0 \Leftrightarrow \nu(A^c) = 1$; inoltre, $\forall A, B \in \mathcal{G}$, $A \subseteq B$ e $\nu(A) = 1 \Rightarrow \nu(B) = 1$. Una misura ν su $\mathcal{G}$ tale che $\nu(E) = 1$ è una misura normalizzata, o misura di probabilità, o semplicemente probabilità su $\mathcal{G}$. Un esempio: $\nu = \varepsilon_{x_0}$ di prima.

$(E, \mathcal{G}, \mu)$; se $\mathcal{P}(x)$ è una proprietà che abbia senso $\forall x \in E$, diremo allora che "$\mathcal{P}(x)$ quasi ovunque su E (rispetto a μ)", ovvero "$\mathcal{P}(x)$ μ-q.o.", $\Leftrightarrow \exists N \in \mathcal{G} : E \setminus N \subseteq \{x \in E \mid \mathcal{P}(x)\}$ e $\mu(N) = 0$; se μ fosse in realtà una probabilità, allora diremmo "quasi certamente", anziché "q.c.".

$(F, \mathcal{G})$; ricordiamo che $f : (E, \mathcal{G}) \longrightarrow (F, \mathcal{G})$ è misurabile $\Leftrightarrow \{f^{-1}(A) \mid A \in \mathcal{G}\} \subseteq \mathcal{G}$, per cui la funzione caratteristica di un insieme misurabile è misurabile, e che $A \in \mathcal{G} \Leftrightarrow \mathbb{1}_A : x \mapsto \begin{cases} 1 & \text{se } x \in A \\ 0 & \text{se } x \notin A \end{cases}$ ($\forall x \in E$) è misurabile $(E, \mathcal{G}) \longrightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$, dove $f, g \in \mathbb{R}^E$ misurabili: $f \sim g$ (mod μ), ovvero $f \overset{\mu}{=} g$, $\Leftrightarrow f = g$ μ-q.o., cioè $\mu(\{x \in E \mid f(x) \neq g(x)\}) = 0$, è una relazione di equivalenza sull'insieme di tutte le funzioni misurabili $E \to \mathbb{R}$; si noti infatti che, $\forall A, B \in \mathcal{G}$, $\mu(A \cup B) = \mu(A) + \mu(B) - \mu(A \cap B)$, per cui $\mu(A) = \mu(B) = 0 \Rightarrow \mu(A \cup B) = 0$; se μ fosse in realtà una probabilità, allora $\mu(A) = \mu(B) = 1 \Rightarrow \mu(A \cap B) = 1$, e quindi in tal caso $f = g$ μ-q.c. $\Leftrightarrow \mu(\{x \in E \mid f(x) = g(x)\}) = 1$.

Sia $(\Omega, \mathcal{A}, \mathbb{P})$ probabilizzato; designato $H \in \mathcal{A}$ con $\mathbb{P}(H) \neq 0$ e definito, $\forall A \in \mathcal{A}$, $Q(A) := \mathbb{P}(H)^{-1} \mathbb{P}(A \cap H)$, ottenemmo un'altra probabilità su $\mathcal{A}$, detta "condizionale relativo a H" o "condizionale dato H" e indicata $Q =: \mathbb{P}(\cdot \mid H) =: \mathbb{P}_H$; allora $\mathbb{P}_\Omega = \mathbb{P}$, e $\mathbb{P}_H = \mathbb{P} \Rightarrow \mathbb{P}(H) = 1$; inoltre, $\forall A \in \mathcal{A}$, $\mathbb{P}(A) = 0 \Rightarrow \mathbb{P}_H(A) = 0$ (cioè $\mathbb{P}(A) = 1 \Rightarrow \mathbb{P}_H(A) = 1$).

$\forall A, B \in \mathcal{A}$, A, B sono (tra loro) *indipendenti* (stocasticamente) $\overset{\text{def}}{\Leftrightarrow} \mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$, e ciò equivale all'indipendenza sia di A^c e B, sia di A e B^c, sia di A^c e B^c; poi, $\forall A \in \mathcal{A}$, $\mathbb{P}(A) = 0 \Rightarrow A, B$ indipendenti $\forall B \in \mathcal{A}$, ovvero lo stesso se $\mathbb{P}(A) = 1$; si noti che $A \in \mathcal{A}$ è indipendente da sé $\Leftrightarrow \mathbb{P}(A) = \mathbb{P}(A)^2$, $\Leftrightarrow$ o $\mathbb{P}(A) = 0$ o $\mathbb{P}(A) = 1$, $\Rightarrow A$ è indipendente da ogni $B \in \mathcal{A}$. Infine, $\forall A, B \in \mathcal{A}$, A, B indipendenti e $A \cap B = \emptyset \Rightarrow \mathbb{P}(A) = 0$ o $\mathbb{P}(B) = 0$.

$(A_i)_{i \in I}$ in $\mathcal{A}$ ($m \geq 2$ nella); gli A_i sono tra loro indipendenti $\overset{\text{def}}{\Leftrightarrow} \forall J \in \mathcal{P}(I, 2, \dots, m)$, $|J| \geq 2$, $\mathbb{P}\left(\bigcap_{j \in J} A_j\right) = \prod_{j \in J} \mathbb{P}(A_j)$, $\Leftrightarrow \forall (B_i)_{i \in I}$ in $\mathcal{A}$ con $\forall i$, $B_i = A_i$ o $B_i = A_i^c$, $\mathbb{P}\left(\bigcap_i B_i\right) = \prod_i \mathbb{P}(B_i)$.

$(\mathcal{A}_i)_{i \in I}$ tribù in $\mathcal{P}(\Omega)$ ($m \geq 2$); le $\mathcal{A}_i$ sono tra loro indipendenti $\overset{\text{def}}{\Leftrightarrow} \forall (A_i)_{i \in I} \in \prod_i \mathcal{A}_i$, $\mathbb{P}\left(\bigcap_i A_i\right) = \prod_i \mathbb{P}(A_i)$; si noti che, $\forall A \in \mathcal{A}$, $\mathcal{P}(A) = \{\emptyset, A, A^c, \Omega\}$, e che $\mathcal{A}_i$ è $\mathbb{P}$-banale $\Leftrightarrow \forall A \in \mathcal{A}_i$, $\mathbb{P}(A) = 0$ o $\mathbb{P}(A) = 1$.

Sulla nozione di completezza

1. Famiglie filtranti di Cauchy

(1.1) Definizione. Sia dato uno spazio pseudo-metrico (E, d). Una famiglia $(x_t)_{t\in I}$ di elementi di E si dice *filtrante* se si suppone di aver fissato, sull'insieme I degli indici, una relazione di preordine (ossia riflessiva e transitiva) la quale sia per giunta *filtrante* (cioè tale che, per ogni coppia di elementi di I, esista un elemento di I che li segua entrambi). Data una siffatta famiglia $(x_t)_{t\in I}$, e indicando con $\prec$ la relazione di preordine fissata su I, si dice che

- $(x_t)_{t\in I}$ è *di Cauchy in E* se, per ogni numero reale ϵ maggiore di zero, esiste un elemento s di I tale che si abbia $\sup_{t\in I,\, t\succ s} d(x_s, x_t) \leq \epsilon$;
- $(x_t)_{t\in I}$ *converge* verso un elemento y di E se, per ogni numero reale ϵ maggiore di zero, esiste un elemento s di I tale che si abbia $\sup_{t\in I,\, t\succ s} d(y, x_t) \leq \epsilon$.

Inoltre si dice che la famiglia $(x_t)_{t\in I}$ è *convergente in E* se esiste un elemento di E verso il quale essa converga.

La proposizione seguente fornisce una condizione sufficiente ad assicurare che un'assegnata famiglia filtrante di elementi di uno spazio pseudo-metrico sia di Cauchy.

(1.2) Proposizione. *Nelle ipotesi e con le notazioni della Definizione (1.1), si supponga che la famiglia filtrante $(x_t)_{t\in I}$ possieda la proprietà seguente: per ogni successione crescente $(t_n)_{n\geq 0}$ di elementi di I, la successione $(x_{t_n})_{n\geq 0}$ è di Cauchy in E. Allora la famiglia filtrante $(x_t)_{t\in I}$ è di Cauchy in E.*

Dimostrazione. Ragionando per assurdo, si supponga che la famiglia $(x_t)_{t\in I}$ non sia di Cauchy in E, cioè che esista un numero reale ϵ maggiore di zero, tale che, per ogni elemento s di I, si possa trovare un elemento t di I, con $t \succ s$, per il quale risulti $d(x_s, x_t) > \epsilon$. Impiegando l'assioma della scelta, è allora possibile costruire una successione crescente $(t_n)_{n\geq 0}$ di elementi di I, tale che, per ogni n, si abbia $d(x_{t_n}, x_{t_{n+1}}) > \epsilon$. Poiché la successione $(x_{t_n})_{n\geq 0}$ non è di Cauchy in E, ciò contraddice l'ipotesi. La proposizione è dunque dimostrata. □

2. La nozione di completezza per uno spazio pseudo-metrico

Abitualmente, lo spazio pseudo-metrico (E, d) si dice *completo* se, in esso, ogni *successione* di Cauchy è convergente. In realtà, è lecito sostituire, in questa definizione, le successioni di Cauchy con le *famiglie filtranti* di Cauchy. Sussiste infatti la proposizione seguente.

(2.1) Proposizione. *Per lo spazio pseudo-metrico (E, d), le due condizioni che seguono sono tra loro equivalenti:*

(a) Ogni successione di Cauchy è convergente.

(b) Ogni famiglia filtrante di Cauchy è convergente.

Dimostrazione. Naturalmente, basta provare l'implicazione (a) $\Rightarrow$ (b). Supposta dunque verificata la condizione (a), consideriamo una famiglia filtrante $(x_t)_{t\in I}$ di elementi di E che sia di Cauchy. Ciò vuol dire che la funzione numerica δ definita su I dalla relazione

$$\delta(s) = \sup_{(\geq 0)} \{d(x_s, x_t) : \ t \in I, t \succ s\} \quad \text{per} \quad s \in I$$

ha estremo inferiore nullo. Esiste allora, grazie all'assioma della scelta, una successione $(s_n)_{n\geq 1}$ di elementi di I, tale che, per ogni n, si abbia $\delta(s_n) \leq 1/n$. La successione $(x_{s_n})_{n\geq 1}$ è di Cauchy. Infatti, per ogni coppia (n, h) d'interi strettamente positivi, detto t un elemento di I che segua entrambi gli elementi s_n e s_{n+h}, si ha

$$d\left(x_{s_n}, x_{s_{n+h}}\right) \leq d\left(x_{s_n}, x_t\right) + d\left(x_{s_{n+h}}, x_t\right) \leq 1/n + 1/(n+h) < 2/n.$$

Grazie all'ipotesi (a), la successione $(x_{s_n})_{n\geq 1}$ converge dunque verso un elemento y di E. Fissiamo un intero n. Allora, per ogni elemento t di I, con $t \succ s_n$, si ha

$$\begin{aligned} d(y, x_t) &\leq d(y, x_{s_n}) + d(x_{s_n}, x_t) \\ &\leq d(y, x_{s_n}) + \delta(s_n). \end{aligned}$$

Ne segue

$$\limsup_t d(y, x_t) \leq d(y, x_{s_n}) + \delta(s_n).$$

Poiché il secondo membro di quest'ultima relazione converge verso zero al tendere di n all'infinito, ciò prova che il primo membro è nullo, ossia che la famiglia filtrante $(x_t)_{t\in I}$ converge verso y. □

Classi monotone

1. Spazi di Riesz e classi monotone

Sia E un insieme. Lo spazio $\mathbb{R}^E$ costituito da tutte le funzioni reali definite in E sarà sempre considerato come munito della sua abituale struttura di spazio vettoriale ordinato. Con questa struttura, esso è uno spazio di Riesz completamente reticolato: infatti ogni famiglia non vuota di elementi di $\mathbb{R}^E$, maggiorati da un medesimo elemento di $\mathbb{R}^E$, ammette come estremo superiore il proprio inviluppo superiore.

Ricordiamo le formule seguenti (che sono valide in $\mathbb{R}^E$, come in ogni spazio di Riesz):

$$f = f^+ - f^-, \qquad |f| = f^+ + f^-, \tag{1.1}$$

$$f \vee g = f + (g-f)^+ = \tfrac{1}{2}\big(f + g + |g-f|\big), \tag{1.2}$$

$$f \wedge g = -\big[(-f) \vee (-g)\big], \tag{1.3}$$

Ogni volta che si parlerà, nel seguito, di uno spazio vettoriale di funzioni reali definite in E, sarà sempre da intendere che si tratti di un sottospazio vettoriale di $\mathbb{R}^E$. Sia $\mathcal{H}$ un tal sottospazio. Affinché $\mathcal{H}$ sia uno spazio di Riesz (o, più precisamente, un sottospazio di Riesz di $\mathbb{R}^E$), cioè affinché $\mathcal{H}$ sia stabile per le due operazioni "reticolari"

$$(f,g) \mapsto f \wedge g, \qquad (f,g) \mapsto f \vee g, \tag{1.4}$$

(occorre e) basta che $\mathcal{H}$ sia stabile per una qualsiasi delle tre operazioni seguenti:

$$f \mapsto |f|, \qquad f \mapsto f^+, \qquad f \mapsto f^-.$$

Ciò risulta dalle relazioni (1.1), (1.2), (1.3).

(1.5) Definizione. Sia $\mathcal{M}$ una parte di $\mathbb{R}^E$. Diremo che $\mathcal{M}$ è *monotòna* (o che $\mathcal{M}$ è una *classe monotòna*) se essa è "stabile per la convergenza monotona dominata delle successioni", nel senso che possiede le due proprietà seguenti:

(a) Per ogni successione crescente $(f_n)_{n\geq 1}$ di elementi di $\mathcal{M}$, maggiorati da un medesimo elemento di $\mathcal{M}$, si ha $\sup_n f_n \in \mathcal{M}$.

(b) Per ogni successione decrescente $(f_n)_{n\geq 1}$ di elementi di $\mathcal{M}$, minorati da un medesimo elemento di $\mathcal{M}$, si ha $\inf_n f_n \in \mathcal{M}$.

(1.6) Osservazione. Se la classe monotona $\mathcal{M}$ è stabile per le operazioni reticolari (1.4), allora essa è *numerabilmente reticolata*, nel senso che contiene l'inviluppo superiore (risp. inferiore) di ogni successione $(f_n)_{n\geq 1}$, non necessariamente monotona, di elementi di $\mathcal{M}$, tutti maggiorati (risp. minorati) da un medesimo elemento di $\mathcal{M}$. Ciò risulta dalle relazioni

$$\sup_n f_n = \sup_n (f_1 \vee \ldots \vee f_n), \qquad \inf_n f_n = \inf_n (f_1 \wedge \ldots \wedge f_n).$$

(1.7) Esempi. (a) Se $(E,\mathcal{E})$ è uno spazio misurabile, la classe $M(\mathcal{E})$ costituita da tutte le funzioni reali misurabili su $(E,\mathcal{E})$ e la classe $M_b(\mathcal{E})$ costituita dalle funzioni limitate e misurabili su $(E,\mathcal{E})$ sono due esempi di spazi di Riesz monotoni.

(b) Se $(E,\mathcal{E},\mu)$ è uno spazio misurato, lo spazio $\mathcal{L}^1(\mu)$ costituito da tutte le funzioni reali integrabili secondo μ è un ulteriore esempio di spazio di Riesz monotono.

(1.8) Definizione. Data una classe $\mathcal{H}$ di funzioni reali definite in E, l'intersezione di tutte le parti monotone di $\mathbb{R}^E$ contenenti $\mathcal{H}$ è la *minima* di esse (rispetto alla relazione d'inclusione): la si chiama la *classe monotona generata* da $\mathcal{H}$, e la si indica $\mathcal{M}(\mathcal{H})$.

(1.9) Osservazione. Nelle ipotesi della definizione precedente, conveniamo di chiamare *$\mathcal{H}$-inquadrabile* ogni funzione che sia compresa tra due funzioni della classe $\mathcal{H}$. Allora, se $\mathcal{M}$ è la classe monotona generata da $\mathcal{H}$, ogni elemento di $\mathcal{M}$ è $\mathcal{H}$-inquadrabile. Infatti la classe costituita dalle funzioni $\mathcal{H}$-inquadrabili è monotona e contiene $\mathcal{H}$, dunque $\mathcal{M}$.

(1.10) Proposizione. *Se $\mathcal{H}$ è un sottospazio vettoriale (risp. di Riesz) di $\mathbb{R}^E$, tale è la classe monotona $\mathcal{M}$ generata da $\mathcal{H}$, cioè $\mathcal{M}(\mathcal{H})$.*

Dimostrazione. Supponiamo che $\mathcal{H}$ sia un sottospazio vettoriale di $\mathbb{R}^E$, e proviamo che lo stesso accade per $\mathcal{M}$.

(a) Fissato uno scalare α, mostriamo che $\mathcal{M}$ è stabile per l'operazione di moltiplicazione per α. Basta per questo osservare che la classe

$$\{f \in \mathbb{R}^E \,:\, \alpha f \in \mathcal{M}\}$$

è monotona e contiene $\mathcal{H}$, dunque $\mathcal{M}$.

(b) Mostriamo ora che, per ogni elemento f di $\mathcal{H}$, la classe $\mathcal{M}$ è stabile per l'operazione $g \mapsto f+g$. Ciò risulta dal fatto che la classe

$$\{g \in \mathbb{R}^E \,:\, f+g \in \mathcal{M}\}$$

è monotona e contiene $\mathcal{H}$, dunque $\mathcal{M}$.

(c) Partendo dal risultato appena ottenuto, e ripetendo lo stesso ragionamento, si prova che, per ogni elemento g di $\mathcal{M}$, la classe $\mathcal{M}$ è stabile per l'operazione $f \mapsto f+g$.

Si vede così che $\mathcal{M}$ è uno spazio vettoriale. Se poi si suppone che $\mathcal{H}$ sia uno spazio di Riesz, allora, per provare che lo spazio vettoriale $\mathcal{M}$ è a sua volta uno spazio di Riesz, basta verificare, con un ragionamento simile a quello impiegato nel punto (a), che $\mathcal{M}$ è stabile per l'operazione $f \mapsto f^+$. □

(1.11) Definizione. Data una classe $\mathcal{H}$ di funzioni reali definite in E, si chiama *tribù generata da $\mathcal{H}$*, e si denota con $\mathcal{T}(\mathcal{H})$, la minima fra tutte le tribù $\mathcal{F}$ su E che "rendano misurabili le funzioni della classe $\mathcal{H}$", cioè che verifichino (con le notazioni di (1.7)) l'inclusione $M(\mathcal{F}) \supset \mathcal{H}$.

2. I teoremi delle classi monotone

(2.1) Teorema. *Siano dati un insieme E e uno spazio di Riesz monotono $\mathcal{M}$, costituito da funzioni limitate su E e contenente le costanti. Si ponga*

$$\mathcal{E} = \{A \in \mathcal{P}(E) : I_A \in \mathcal{M}\}. \tag{2.2}$$

Valgono allora (con le notazioni di (1.7) e (1.11)) le affermazioni seguenti:

(a) $\mathcal{E}$ è una tribù su E.

(b) $\mathcal{M} = M_b(\mathcal{E})$.

(c) $\mathcal{E} = \mathcal{T}(\mathcal{M})$.

Dimostrazione. Dalla relazione (2.2) che definisce $\mathcal{E}$ risulta che l'intero insieme E appartiene a $\mathcal{E}$ perché la sua funzione indicatrice, ossia la costante 1, appartiene, per ipotesi, a $\mathcal{M}$. Così pure, se A è un elemento di $\mathcal{E}$, allora anche l'insieme A^c (complementare di A rispetto a E) appartiene a $\mathcal{E}$: infatti la sua funzione indicatrice, essendo eguale a $1 - I_A$, appartiene a $\mathcal{M}$. Infine, se A è l'unione di una successione $(A_n)_{n\geq 1}$ di elementi di $\mathcal{E}$, allora A appartiene a $\mathcal{E}$ grazie all'eguaglianza $I_A = \sup_n I_{A_n}$ e al fatto che lo spazio di Riesz $\mathcal{M}$ è numerabilmente reticolato (si veda (1.6)). Si è così provato che $\mathcal{E}$ è una tribù su E.

Sia f un elemento di $\mathcal{M}$. Allora f è misurabile su $(E, \mathcal{E})$ grazie al fatto che, per ogni numero reale c, si ha

$$I_{\{f>c\}} = \sup_{n\geq 1}\left[1 \wedge n(f-c)^+\right] \in \mathcal{M}.$$

Inversamente, ogni funzione semplice su $(E, \mathcal{E})$ appartiene a $\mathcal{M}$ come combinazione lineare finita di funzioni indicatrici di elementi di $\mathcal{E}$ (e dunque come combinazione lineare finita di speciali elementi di $\mathcal{M}$). Di conseguenza, anche ogni funzione limitata e misurabile su $(E, \mathcal{E})$ appartiene a $\mathcal{M}$ in quanto inviluppo superiore di una successione crescente di funzioni semplici su $(E, \mathcal{E})$. L'eguaglianza (b) è così dimostrata.

Infine, per provare l'eguaglianza (c), basta osservare che, per ogni tribù $\mathcal{F}$ su E, vale (grazie alla già dimostrata eguaglianza (b)) la seguente catena di equivalenze :

$$\begin{aligned} \mathcal{F} \supset \mathcal{T}(\mathcal{M}) &\Leftrightarrow M_b(\mathcal{F}) \supset \mathcal{M} \\ &\Leftrightarrow M_b(\mathcal{F}) \supset M_b(\mathcal{E}) \\ &\Leftarrow \mathcal{F} \supset \mathcal{E}. \quad \square \end{aligned}$$

(2.3) Definizione. Dato uno spazio misurabile $(E, \mathcal{E})$, una *base* per la tribù $\mathcal{E}$ è una parte $\mathcal{I}$ di $\mathcal{E}$ che possieda le proprietà seguenti:

(a) La tribù $\mathcal{T}(\mathcal{I})$ generata da $\mathcal{I}$ coincide con $\mathcal{E}$: $\mathcal{T}(\mathcal{I}) = \mathcal{E}$.

(b) $\mathcal{I}$ è stabile per l'operazione d'intersezione di due insiemi: $A, B \in \mathcal{I} \Rightarrow A \cap B \in \mathcal{I}$.

(c) L'intero insieme E è unione di una successione di elementi di $\mathcal{I}$: $\exists (A_m)_{m\geq 1}$ in $\mathcal{I}$: $E = \bigcup_{m\geq 1} A_m$.

Una base $\mathcal{I}$ per la tribù $\mathcal{E}$ si dice poi *speciale* se per essa vale il seguente rafforzamento della proprietà (c):

(c′) L'intero insieme E appartiene a $\mathcal{I}$.

Il teorema seguente si rivela spesso più utile del Teorema (2.1):

(2.4) Teorema. *Siano dati uno spazio misurabile $(E, \mathcal{E})$, una base speciale $\mathcal{I}$ per la tribù $\mathcal{E}$, e uno spazio vettoriale monotono $\mathcal{L}$ di funzioni reali su E, con*

$$\mathbb{R}^E \supset \mathcal{L} \supset \{I_A : A \in \mathcal{I}\} \tag{2.5}$$

Si ha allora $\mathcal{L} \supset M_b(\mathcal{E})$.

Dimostrazione. Denotiamo con $\mathcal{R}$ lo spazio vettoriale generato da $\{I_A : A \in \mathcal{I}\}$, e proviamo che esso è uno spazio di Riesz. Osserviamo innanzitutto che, essendo l'insieme $\{I_A : A \in \mathcal{I}\}$ stabile per la moltiplicazione, tale è anche $\mathcal{R}$. Sia f un elemento qualsiasi di $\mathcal{R}$. Allora f si può mettere nella forma $f = \sum_{i\in I} a_i I_{A_i}$, dove $(A_i)_{i\in I}$ è un'opportuna famiglia finita di elementi di $\mathcal{I}$ e gli a_i sono opportuni numeri reali. Per ogni parte H di I, poniamo

$$B_H = \left(\bigcap_{i\in H} A_i\right) \cap \left(\bigcap_{j\in I\setminus H} A_j^c\right).$$

Si ha allora

$$I_{B_H} = \left(\prod_{i\in H} I_{A_i}\right) \cdot \left(\prod_{j\in I\setminus H} \left(1 - I_{A_j}\right)\right) \in \mathcal{R}.$$

Inoltre gli insiemi B_H sono a due a due disgiunti, e si ha

$$f = \sum_{H\in\mathcal{P}(I)} b_H I_{B_H}, \quad \text{con} \quad b_H = \sum_{i\in H} a_i.$$

Ne segue

$$f^+ = \sum_{H\in\mathcal{P}(I),\, b_H>0} b_H I_{B_H} \in \mathcal{R}.$$

È così provato che $\mathcal{R}$ è uno spazio di Riesz. Denotiamo ora con $\mathcal{M}$ la classe monotona generata da $\mathcal{R}$. Allora $\mathcal{M}$ è il minimo spazio vettoriale monotono contenente l'insieme $\{I_A : A \in \mathcal{I}\}$. Tenendo conto di ciò, è facile riconoscere che valgono le due relazioni seguenti:

$$\mathcal{L} \supset \mathcal{M}, \qquad \mathcal{T}(\mathcal{M}) = \mathcal{E}. \tag{2.6}$$

Per provare la prima, basta ricordare che, per ipotesi, $\mathcal{L}$ è uno spazio vettoriale monotono contenente l'insieme $\{I_A : A \in \mathcal{I}\}$. Per provare la seconda, basta osservare che, per ogni tribù $\mathcal{F}$ su E, poiché $M(\mathcal{F})$ è uno spazio vettoriale monotono, vale la seguente catena di equivalenze:

$$\begin{aligned} \mathcal{F} \supset \mathcal{T}(\mathcal{M}) &\Leftrightarrow M(\mathcal{F}) \supset \mathcal{M} \\ &\Leftrightarrow M(\mathcal{F}) \supset \{I_A : A \in \mathcal{I}\} \\ &\Leftrightarrow \mathcal{F} \supset \mathcal{I} \\ &\Leftrightarrow \mathcal{F} \supset \mathcal{E} \end{aligned}$$

(dove l'equivalenza finale è dovuta al fatto che, grazie alla definizione di base, la classe d'insiemi $\mathcal{I}$ è, in particolare, un sistema di generatori per la tribù $\mathcal{E}$).

D'altra parte, $\mathcal{M}$ è uno spazio di Riesz monotono (si veda (1.10)) contenente le costanti e costituito da funzioni limitate, in quanto $\mathcal{R}$-inquadrabili (si veda (1.9)). Grazie al Teorema (2.1), si ha dunque

$$\mathcal{M} = M_b(\mathcal{T}(\mathcal{M})). \tag{2.7}$$

Per concludere la dimostrazione, basta osservare che da (2.6) e (2.7) discende

$$\mathcal{L} \supset \mathcal{M} = M_b(\mathcal{T}(\mathcal{M})) = M_b(\mathcal{E}). \qquad \square$$

Come esempio di applicazione del teorema appena dimostrato, proviamo ora un importante criterio per la coincidenza di due misure.

(2.8) Teorema. *Siano μ, ν due misure finite sullo spazio misurabile $(E, \mathcal{E})$ e sia $\mathcal{I}$ una base speciale per la tribù $\mathcal{E}$. Si supponga che le due misure μ, ν coincidano su $\mathcal{I}$. Allora esse sono identiche.*

Dimostrazione. Poniamo $\mathcal{L} = \left\{f \in \mathcal{L}^1(\mu) \cap \mathcal{L}^1(\nu) : \int f\,d\mu = \int f\,d\nu\right\}$. La coincidenza delle due misure μ, ν su $\mathcal{I}$ si traduce allora mediante l'inclusione (2.5). Grazie al Teorema (2.4), ne segue

$$\mathcal{L} \supset M_b(\mathcal{E}) \supset \{I_A : A \in \mathcal{E}\}$$

e ciò significa che le due misure μ, ν sono identiche. $\square$

Il criterio appena dimostrato può essere esteso nel modo seguente.

(2.9) Teorema. *Siano μ, ν due misure (non necessariamente finite) sullo spazio misurabile $(E, \mathcal{E})$ e sia $\mathcal{I}$ una base per la tribù $\mathcal{E}$. Si supponga che le due misure μ, ν coincidano, e siano finite, su ciascun elemento di $\mathcal{I}$. Allora esse sono identiche.*

Dimostrazione. Sia $(E_n)_{n\geq 1}$ una successione di elementi di $\mathcal{I}$ la cui unione coincida con l'intero insieme E. Poniamo, per ogni n,

$$\mathcal{E}_n = \{A \cap E_n : A \in \mathcal{E}\}, \qquad \mathcal{I}_n = \{A \cap E_n : A \in \mathcal{I}\}.$$

Allora $\mathcal{E}_n$ è una tribù su E_n, ammettente $\mathcal{I}_n$ come base speciale. Denotiamo inoltre con μ_n, ν_n le restrizioni delle due misure μ, ν alla tribù $\mathcal{E}_n$. Per ogni n, poiché le due misure μ_n, ν_n sono finite, e coincidono su $\mathcal{I}_n$, esse sono identiche, grazie a (2.8). È facile dedurne che le due misure μ, ν sono identiche. A questo scopo, assegnato un elemento A di $\mathcal{E}$, poniamo, per ogni n, $A_n = A \cap E_n$. L'insieme A è l'unione degli insiemi A_n, ma questi non sono necessariamente disgiunti. Tuttavia, se si pone

$$B_n = A_n \cap \Big(\bigcup_{1\leq k<n} A_k\Big)^c,$$

si ottiene una successione $(B_n)_{n\geq 1}$ d'insiemi a due a due disgiunti, con $B_n \in \mathcal{E}_n$ per ogni n, la cui unione coincide con A. Poiché le due misure μ, ν coincidono su ciascuno degli insiemi B_n, esse coincidono su A, e ciò conclude la dimostrazione. $\square$

(
$(E,\mathcal{E}) \xrightarrow{\varphi} (\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}}))$
$X \uparrow \quad \nearrow Y$
$(\Omega, \sigma(X))$
)

Teorema: criterio di misurabilità di Doob.

$(E,\mathcal{E})$, $(\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}}))$; $E, \Omega \neq \emptyset$; $X \in E^{\Omega}$; $(\Omega, \sigma(X))$.

$Y \in \overline{\mathbb{R}}^{\Omega}$ $\sigma(X)$-misurabile ($\in M(\sigma(X))$) $\Rightarrow \exists\, \varphi \in \overline{\mathbb{R}}^{E}$ $\mathcal{E}$-misurabile ($\in M(\mathcal{E})$) : $Y = \varphi \circ X$

Dim. Possiamo farlo solo per $Y \geq 0$, in quanto poi dal momento $Y = Y^+ - Y^- = \varphi_1 \circ X - \varphi_2 \circ X = (\varphi_1 - \varphi_2) \circ X$ per Y qualsiasi. Allora, $\forall \alpha \in \mathbb{Q}_+$, $\{Y \geq \alpha\} \in \sigma(X)$, cioè, $\forall \alpha \in \mathbb{Q}_+$, $\exists\, A_\alpha \in \mathcal{E} : \{Y \geq \alpha\} = \{X \in A_\alpha\}$, per cui

$$Y = \sup_{\alpha \in \mathbb{Q}_+} \alpha I_{\{Y \geq \alpha\}} = \sup_{\alpha \in \mathbb{Q}_+} \alpha I_{\{X \in A_\alpha\}} = \sup_{\alpha \in \mathbb{Q}_+} \alpha (I_{A_\alpha} \circ X) = \sup_{\alpha \in \mathbb{Q}_+} \alpha I_{A_\alpha} \circ X$$

$$= \Big(\sup_{\alpha \in \mathbb{Q}_+} \alpha I_{A_\alpha}\Big) \circ X \quad ; \quad \varphi := \sup_{\alpha \in \mathbb{Q}_+} \alpha I_{A_\alpha} \text{ [illegible]} \qquad \square$$

(Se $\mathcal{G}$ è base per $\mathcal{E}$, allora X v.a. $\Leftrightarrow \forall A \in \mathcal{G}$, $X^{-1}(A) \in \mathcal{A}$)

Se $(\Omega, \mathcal{A}, P)$ è uno spazio probabilizzato, un'applicazione $X : (\Omega, \mathcal{A}) \to (E, \mathcal{E})$ misurabile viene sempre chiamata (piuttosto) variabile aleatoria, o v.a., su $(\Omega, \mathcal{A})$ a valori in $(E, \mathcal{E})$. Nel caso particolare $(E,\mathcal{E}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$, X è v.a.r. Come noto, definendo $\forall A \in \mathcal{E}$, $Q(A) := P(X^{-1}(A))$ $(= P(\{X \in A\}) =: P[X \in A])$, otteniamo una probabilità su $\mathcal{E}$, detta "probabilità immagine di P mediante X", e indicata $Q =: P_X =: X(P)$. Si noti la sua associatività:

$(E, \mathcal{E}, X(P)) \xrightarrow{\varphi} (F, \mathcal{F}, \cdot)$
$X \uparrow \quad \nearrow \varphi \circ X$
$(\Omega, \mathcal{A}, P)$

$\forall \varphi : (E, \mathcal{E}, X(P)) \to (F, \mathcal{F})$ v.a., è (su $\mathcal{F}$)
$(\varphi \circ X)(P) = \varphi(X(P))$ $\quad ((\varphi \circ X)^{-1}(A) = X^{-1}(\varphi^{-1}(A)))$.

$X(P)$ viene anche chiamata "la legge (di probabilità)" della v.a. X.
Sia $(E, \mathcal{E}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$, e ricordiamo (riformuliamo) il fondamentale teorema d'integrazione rispetto ad una probabilità immagine (per v.a.r.):

$(\mathbb{R}, \mathcal{B}(\mathbb{R}), X(P)) \xrightarrow{\varphi} (\mathbb{R}, \mathcal{B}(\mathbb{R}))$
$X \uparrow \quad \nearrow \varphi \circ X$
$(\Omega, \mathcal{A}, P)$

$\forall \varphi \in \mathbb{R}^{\mathbb{R}}$ boreliana, $\varphi \circ X \in \mathcal{L}^1(P) \Leftrightarrow \varphi \in \mathcal{L}^1(X(P))$
e in tal caso $\displaystyle\int_\Omega \varphi(X(\omega))\, dP(\omega) = \int_{\mathbb{R}} \varphi(x)\, dX(P)(x)$.

In particolare
$\displaystyle\int_\Omega X(\omega)\, dP(\omega) = \int_{\mathbb{R}} x\, dX(P)(x)$; per $X \in \mathcal{L}^1(P)$, o $X \geq 0$, indichiamo $\displaystyle\int_\Omega X\, dP =: E[X] =: P[X]$. Per $X \geq 0$ (cioè $X = |Z|$, "Z come X"), $\forall \varepsilon > 0$ vale, $\varepsilon \cdot I_{\{X \geq \varepsilon\}} \leq X \Rightarrow \varepsilon \cdot P[X \geq \varepsilon] \leq P[X]$ ("dis. di Markov"). Per $X \in \mathcal{L}^2(P)$, indichiamo $Var[X] := P[(X - P[X])^2] = P[X^2] - P[X]^2$ $(\forall\, 1 \leq p < q < \infty$, $X \in \mathcal{L}^q(P) \Rightarrow X \in \mathcal{L}^p(P))$; in questo $\forall \varepsilon > 0$ vale, $\varepsilon^2 \cdot P[|X - P[X]| \geq \varepsilon] \leq Var[X]$ ("dis. di Chebyshev"), $Var[X] = 0 \Leftrightarrow X \overset{P}{=} P[X]$.
(cioè X è costante [illegible])

Famiglie di variabili aleatorie

1. Blocco di una famiglia di variabili aleatorie

(1.1) Definizione. Sia data una famiglia $((E_i, \mathcal{E}_i))_{i\in I}$ di spazi misurabili.

(a) Il *prodotto cartesiano* della famiglia d'insiemi $(E_i)_{i\in I}$ è l'insieme costituito da tutte le famiglie della forma $(x_i)_{i\in I}$ tali che, per ciascun indice i, il termine x_i sia un elemento di E_i. Esso si denota con $\prod_{i\in I} E_i$.

(b) Per ciascun indice j, la *proiezione canonica* di indice j, relativa alla famiglia d'insiemi $(E_i)_{i\in I}$, è l'applicazione ξ_j che, ad ogni elemento $(x_i)_{i\in I}$ del prodotto cartesiano $\prod_{i\in I} E_i$, associa x_j:

(c) La *tribù prodotto* della famiglia di tribù $(\mathcal{E}_i)_{i\in I}$ è la minima tribù su $\prod_{i\in I} E_i$ che renda misurabili tutte le proiezioni canoniche o, in modo equivalente, è la minima tribù su $\prod_{i\in I} E_i$ che contenga tutti gli insiemi della forma $\{\xi_j \in A_j\}$, con $j \in I$ e $A_j \in \mathcal{E}_j$. Essa si denota con $\bigotimes_{i\in I} \mathcal{E}_i$.

(1.2) Definizione. Nelle ipotesi della definizione precedente, siano dati uno spazio probabilizzato $(\Omega, \mathcal{A}, P)$ e una famiglia $(X_i)_{i\in I}$ di applicazioni, tale che, per ciascun indice i, l'applicazione X_i abbia Ω come insieme di partenza e E_i come insieme d'arrivo.

(a) Il *blocco* della famiglia di applicazioni $(X_i)_{i\in I}$ è l'applicazione di Ω in $\prod_{i\in I} E_i$ che, a ciascun elemento ω di Ω, associa l'elemento $\big(X_i(\omega)\big)_{i\in I}$ di $\prod_{i\in I} E_i$. Esso si denota con $[X_i]_{i\in I}$. Nel caso particolare in cui l'insieme I degli indici sia l'insieme degli interi compresi tra 1 e un fissato intero n, si usa anche la notazione $[X_i]_{1\le i\le n}$ oppure $[X_1, \dots, X_n]$.

(b) Per ciascun indice j, l'applicazione X_j si chiama la *componente* di indice j del blocco $[X_i]_{i\in I}$;

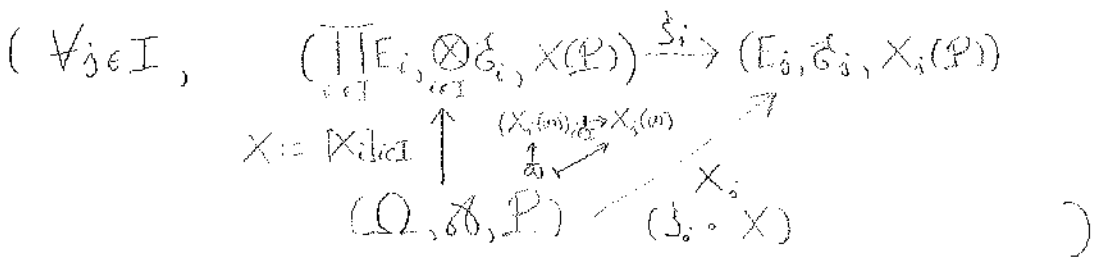
(1.3) Proposizione. *Nelle ipotesi di (1.2), le condizioni che seguono sono tra loro equivalenti:*

(a) Per ciascun indice j, l'applicazione X_j è una variabile aleatoria su $(\Omega, \mathcal{A})$ a valori in $(E_j, \mathcal{E}_j)$.

(b) Il blocco $[X_i]_{i\in I}$ è una variabile aleatoria su $(\Omega, \mathcal{A})$ a valori nello spazio misurabile

$$\Big(\prod_{i\in I} E_i,\ \bigotimes_{i\in I} \mathcal{E}_i\Big).$$

(1.4) Definizione. Nelle ipotesi della proposizione precedente, si suppongano soddisfatte le due condizioni equivalenti (a), (b), e si denoti concisamente con X il blocco $[X_i]_{i\in I}$. Allora la legge di X secondo P, ossia la misura immagine $X(P)$, si chiama anche la *legge congiunta* della famiglia $(X_i)_{i\in I}$ (o, più familiarmente, la *legge congiunta* delle X_i). Essa è una misura di probabilità sulla tribù prodotto $\bigotimes_{i\in I} \mathcal{E}_i$.



(1.5) Osservazione. Nelle ipotesi della definizione precedente, per ciascun indice j, essendo $X_j = \xi_j \circ X$, si ha

$$X_j(P) = \xi_j(X(P)),$$

ossia la legge di X_j è l'immagine, mediante la proiezione canonica ξ_j, della legge congiunta delle X_i.

(1.6) Osservazione. Nelle ipotesi di (1.1), sia data, sulla tribù prodotto $\bigotimes_{i\in I} \mathcal{E}_i$, un'arbitraria misura normalizzata ν. Allora è sempre possibile costruire uno spazio probabilizzato $(\Omega, \mathcal{A}, P)$ e, su di esso, una famiglia di variabili aleatorie la cui legge congiunta sia ν. Per questo, basta prendere Ω eguale al prodotto cartesiano $\prod_{i\in I} E_i$, la tribù $\mathcal{A}$ eguale alla tribù prodotto $\bigotimes_{i\in I} \mathcal{E}_i$, la misura P eguale all'assegnata misura ν, e considerare la famiglia $(\xi_j)_{j\in I}$ delle proiezioni canoniche. Infatti la legge congiunta di questa famiglia, ossia l'immagine di ν mediante il blocco $[\xi_j]_{j\in I}$, coincide con ν per il semplice fatto che il blocco $[\xi_j]_{j\in I}$ non è altro che l'applicazione identica del prodotto cartesiano $\prod_{i\in I} E_i$.

2. Famiglie di variabili aleatorie indipendenti

(2.1) Definizione. Sia $\big((E_i, \mathcal{E}_i, \mu_i)\big)_{i\in I}$ una famiglia di spazi probabilizzati. Una misura ν sulla tribù $\bigotimes_{i\in I} \mathcal{E}_i$ si chiama una *misura prodotto* della famiglia $(\mu_i)_{i\in I}$ se, per ogni parte finita J di I ed ogni famiglia d'insiemi $(A_j)_{j\in J}$, avente J come insieme degli indici e tale che, per ciascun elemento j di J, l'insieme A_j appartenga alla tribù $\mathcal{E}_j$, ha luogo la relazione

$$\nu\Big(\bigcap_{j\in J}\{\xi_j \in A_j\}\Big) = \prod_{j\in J} \mu_j(A_j) \tag{2.2}$$

(dove ξ_j denota la proiezione canonica di indice j, così come definita in (1.1) (b)).

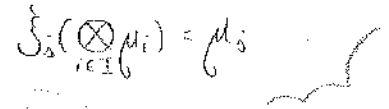
(2.3) Teorema. *Nelle ipotesi della definizione precedente, esiste una e una sola misura prodotto della famiglia* $(\mu_i)_{i\in I}$.

(2.4) Notazione. Nelle ipotesi della Definizione (2.1), l'unica misura prodotto della famiglia $(\mu_i)_{i\in I}$ si denota col simbolo $\bigotimes_{i\in I} \mu_i$. Nel caso particolare in cui l'insieme I degli indici sia l'insieme degli interi compresi tra 1 e un fissato intero n, si usa anche la notazione $\bigotimes_{1\le i\le n} \mu_i$ oppure $\mu_1 \otimes \ldots \otimes \mu_n$.

(2.5) Osservazione. Nelle ipotesi della Definizione (2.1), si denoti con ν la misura prodotto $\bigotimes_{i\in I} \mu_i$. Allora, fissato un elemento j di I, se B è un arbitrario elemento di $\mathcal{E}_j$, la relazione (2.2) nella quale si prenda $J = \{j\}$ e $A_j = B$, si riduce alla forma

$$\nu\{\xi_j \in B\} = \mu_j(B).$$

Per l'arbitrarietà di B, ciò prova che l'immagine della misura prodotto $\nu = \bigotimes_{i\in I} \mu_i$ mediante la proiezione canonica ξ_j coincide con μ_j. In formule: $\xi_j(\nu) = \mu_j$, cioè



(2.6) Definizione. Su uno spazio probabilizzato $(\Omega, \mathcal{A}, P)$ sia $(X_i)_{i\in I}$ una famiglia di variabili aleatorie (non necessariamente a valori in un medesimo spazio misurabile). Si dice che essa è una *famiglia di variabili aleatorie indipendenti* se la sua legge congiunta (nel senso di (1.4)) è una misura prodotto (nel senso di (2.1)).

(2.7) Osservazione. Nelle ipotesi della definizione precedente, se la famiglia $(X_i)_{i\in I}$ è una famiglia di variabili aleatorie indipendenti, allora la sua legge congiunta ν può essere espressa *in un sol modo* come come misura prodotto: precisamente, quest'unico modo consiste nell'esprimere ν come misura prodotto della famiglia $(\mu_i)_{i\in I}$, dove μ_i denota, per ciascun indice i, la legge di X_i. Per provare questa affermazione, denotiamo concisamente con X il blocco $[X_i]_{i\in I}$. Inoltre, per ogni indice i, denotiamo con $(E_i, \mathcal{E}_i)$ lo spazio d'arrivo di X_i. Supponiamo che si abbia $\nu = \bigotimes_{i\in I} \mu_i$, dove μ_i sia (per ciascun indice i) un'opportuna misura di probabilità sulla tribù $\mathcal{E}_i$. Allora, essendo $X_j = \xi_j \circ X$, si ha

$$X_j(P) = \xi_j(\nu) = \mu_j$$

(dove l'eguaglianza finale discende da (2.5)). L'affermazione è così dimostrata.

Una coppia di variabili aleatorie si può vedere come una speciale famiglia di variabili aleatorie (avente come insieme degli indici un insieme costituito da due elementi). Il concetto di famiglia di variabili aleatorie indipendenti contiene dunque, come caso particolare, quello di *coppia di variabili aleatorie indipendenti*. Si riconosce immediatamente che, se (X, Y) è una coppia di variabili aleatorie indipendenti, tale è anche la coppia (Y, X). Si dice allora, in modo più simmetrico, che le due variabili aleatorie X, Y sono *tra loro indipendenti*. Questo concetto si riduce a una banalità nel caso di una coppia di variabili aleatorie tra loro eguali. Sussiste infatti la proposizione seguente (di dimostrazione immediata).

(2.8) Proposizione. *Su uno spazio probabilizzato $(\Omega, \mathcal{A}, P)$ sia X una variabile aleatoria a valori in uno spazio misurabile $(E, \mathcal{E})$. Le condizioni che seguono sono tra loro equivalenti:*

(a) La coppia (X, X) è una coppia di variabili aleatorie indipendenti

(b) La legge di X è degenere, ossia non prende che i valori 0 e 1.

3. Successioni stazionarie

(3.1) Definizione. Su uno spazio probabilizzato $(\Omega, \mathcal{A}, P)$ sia $(X_n)_{n\geq 1}$ una successione di variabili aleatorie a valori in un medesimo spazio misurabile $(E, \mathcal{E})$. Denoteremo con $E^{\mathbb{N}^*}$ il prodotto cartesiano di una successione d'insiemi tutti eguali a E, e con $\mathcal{E}^{\otimes\mathbb{N}^*}$ la tribù prodotto di una successione di tribù tutte eguali a $\mathcal{E}$. Denoteremo concisamente con X il blocco $[X_n]_{n\geq 1}$, che, grazie a (1.3), potremo considerare come una variabile aleatoria a valori nello spazio misurabile $\left(E^{\mathbb{N}^*}, \mathcal{E}^{\otimes\mathbb{N}^*}\right)$. Denoteremo inoltre con θ l'applicazione (misurabile) di questo spazio misurabile in sé che, ad ogni

elemento $(x_n)_{n\geq 1}$ di $E^{\mathbb{N}^*}$, associa l'elemento $(x_{n+1})_{n\geq 1}$. Chiameremo θ l'operatore di *slittamento*. Ciò posto, diremo che la successione $(X_n)_{n\geq 1}$ è *stazionaria* se il blocco X è isonomo al blocco $\theta \circ X$.

(3.2) Osservazione. Nelle ipotesi della definizione precedente, si denoti con ν la legge congiunta delle X_n, ossia la legge $X(P)$ del blocco X. Allora il blocco $\theta \circ X$ (ossia il blocco $[X_{n+1}]_{n\geq 1}$) ha come legge la misura $\theta(\nu)$ (immagine di ν mediante θ). Si può dunque dire, in modo equivalente, che la successione $(X_n)_{n\geq 1}$ è stazionaria se la sua legge congiunta è invariante per l'operatore θ di slittamento.

(3.3) Proposizione. *Nelle ipotesi della Definizione (3.1), consideriamo le condizioni seguenti:*

(a) La successione $(X_n)_{n\geq 1}$ è una successione di variabili aleatorie indipendenti e isonome.

(b) La successione $(X_n)_{n\geq 1}$ è stazionaria.

(c) Le variabili aleatorie X_n sono tra loro isonome.

Valgono allora le due implicazioni (a) $\Rightarrow$ (b) e (b) $\Rightarrow$ (c). Nessuna di queste può essere invertita.

Dimostrazione. (a) $\Rightarrow$ (b): Tenendo conto dell'Osservazione (2.7), si vede che la condizione (a) equivale al fatto che la legge congiunta ν delle X_n sia il prodotto di una successione di misure, tutte eguali a una medesima misura di probabilità μ sulla tribù $\mathcal{E}$. È immediato verificare che una tale misura prodotto è invariante per l'operatore θ di slittamento.

(b) $\Rightarrow$ (c): Si supponga che la successione $(X_n)_{n\geq 1}$ sia stazionaria. Allora, ragionando per induzione, si vede che, per ciascun intero k strettamente positivo, la legge del blocco $[X_n]_{n\geq 1}$ è eguale a quella del blocco $[X_{n+k}]_{n\geq 1}$. Ne segue che la la variabile aleatoria X_1 (ottenuta componendo il primo di questi due blocchi con la proiezione canonica di indice 1) è isonoma alla variabile aleatoria X_{1+k} (ottenuta componendo l'altro blocco con la stessa proiezione canonica).

Per provare che l'implicazione (b) $\Rightarrow$ (a) non è sempre vera, basta pensare ad una successione i cui termini siano tutti identici a una medesima variabile aleatoria non degenere (cioè la cui legge non sia degenere). Una tal successione è banalmente stazionaria, ma non è una successione di variabili aleatorie indipendenti (si veda (2.8)).

Per provare che l'implicazione (c) $\Rightarrow$ (b) non è sempre vera, basta partire da una successione $(Z_n)_{n\geq 1}$ di variabili aleatorie indipendenti, isonome, non degeneri, e costruire la successione $(X_n)_{n\geq 1}$ nel modo seguente:

$$X_1 = Z_1, \qquad X_{n+1} = Z_n \quad \text{per } n \geq 1. \qquad \square$$

La legge zero-uno di Kolmogorov

1. Definizione. Su uno spazio probabilizzato $(\Omega, \mathcal{A}, P)$ sia data una famiglia $(X_i)_{i\in I}$ di variabili aleatorie (a valori in spazi misurabili non necessariamente eguali tra loro), avente come insieme degli indici un arbitrario insieme infinito I. Per ogni parte J di I, si ponga

$$\mathcal{F}_J = \mathcal{T}(X_i : i \in J).$$

Si denoti inoltre con $\mathcal{H}$ l'insieme delle parti finite di I, e si ponga

$$\boxed{\mathcal{T} = \bigcap_{H\in\mathcal{H}} \mathcal{F}_{I\setminus H}}$$

La tribù $\mathcal{T}$ così definita si chiama la ~~tribù terminale~~ (relativa all'assegnata famiglia di variabili aleatorie). Una variabile aleatoria (a valori in un arbitrario spazio misurabile) si dice poi *terminale* se è misurabile rispetto a $\mathcal{T}$ (cioè rispetto ad ogni tribù della forma $\mathcal{F}_{I\setminus H}$, con $H \in \mathcal{H}$). Infine un evento si dice *terminale* se appartiene a $\mathcal{T}$ (o, ciò ch'è lo stesso, se la sua funzione indicatrice è terminale).

2. Esempio. Sia $(X_i)_{i\in I}$ una famiglia infinita *numerabile* di variabili aleatorie *reali*, e si ponga

$$A = \left\{\sum_{i\in I} |X_i| < \infty\right\}.$$

Allora l'evento A così definito è terminale (relativamente all'assegnata famiglia). Infatti, se H è una parte finita di I, si ha

$$A = \left\{\sum_{i\in I\setminus H} |X_i| < \infty\right\}.$$

3. Esempio. Sia $(X_n)_{n\geq 1}$ una successione di variabili aleatorie *numeriche*, e si ponga

$$Z = \liminf_{n\to\infty} X_n.$$

Allora Z è una variabile aleatoria terminale (relativamente all'assegnata successione). Infatti, se H è una parte finita di $\mathbb{N}^*$, si ha

$$Z = \liminf_{n\to\infty,\, n\notin H} X_n.$$

4. Teorema. (Legge zero-uno di Kolmogorov) *Nelle ipotesi della Definizione 1, si supponga che la famiglia $(X_i)_{i\in I}$ sia una famiglia di variabili aleatorie indipendenti. Allora la tribù terminale ad essa relativa è degenere, ossia ogni evento terminale ha probabilità eguale a 0 o a 1. Di conseguenza, ogni variabile aleatoria terminale è degenere.*

Dimostrazione. Adottiamo le notazioni della Definizione 1. Si ha allora $\mathcal{T} \subset \mathcal{F}_I$. Dunque, per dimostrare che $\mathcal{T}$ è degenere (ossia indipendente da sé stessa), basta provare che $\mathcal{T}$ è indipendente da $\mathcal{F}_I$. Per far ciò, il criterio per l'indipendenza da un blocco (nella sua formulazione in termini di tribù) permette di ridursi a verificare che $\mathcal{T}$ è indipendente da ogni tribù della forma $\mathcal{F}_H$, con H elemento di $\mathcal{H}$. A questo scopo, fissato H, basta osservare che, per la supposta indipendenza delle X_i, le due tribù

$$\mathcal{F}_H, \qquad \mathcal{F}_{I\setminus H}$$

sono tra loro indipendenti e che, per la definizione stessa di $\mathcal{T}$, si ha $\mathcal{T} \subset \mathcal{F}_{I\setminus H}$. □

I lemmi di Borel-Cantelli

1. I due lemmi

(1.1) Primo lemma di Borel-Cantelli. *In uno spazio probabilizzato $(\Omega, \mathcal{A}, P)$ sia data una successione $(A_n)_{n\geq 1}$ di eventi, con*

$$\sum_{n\geq 1} P(A_n) < \infty.$$

Allora l'evento $\limsup A_n$ è trascurabile.

Dimostrazione. Per ogni n, denotiamo con X_n la funzione indicatrice di A_n. Si ha allora

$$P\left[\sum_{n\geq 1} X_n\right] = \sum_{n\geq 1} P(A_n) < \infty.$$

Dunque la variabile aleatoria numerica $\sum_{n\geq 1} X_n$ è integrabile. Ne segue che essa è quasi certamente finita, ossia che l'evento $\left\{\sum_{n\geq 1} X_n = \infty\right\}$ (identico all'evento $\limsup_n A_n$) è trascurabile. □

(1.2) Secondo lemma di Borel-Cantelli. *In uno spazio probabilizzato $(\Omega, \mathcal{A}, P)$ sia data una successione $(A_n)_{n\geq 1}$ di eventi, con*

$$\sum_{n\geq 1} P(A_n) = \infty. \tag{1.3}$$

Si supponga inoltre che gli eventi A_n siano a due a due indipendenti.

Allora l'evento $\limsup A_n$ è quasi-certo.

Dimostrazione. Per ogni n, denotiamo con X_n la funzione indicatrice di A_n. Poniamo inoltre $S_n = \sum_{j=1}^n X_j$. L'ipotesi (1.3) fornisce allora

$$P[S_n] = \sum_{j=1}^n P[X_j] = \sum_{j=1}^n P(A_j) \uparrow \infty.$$

Si ha quindi, per n abbastanza grande, $P[S_n] > 0$. Senza ledere la generalità, supporremo che ciò accada per ogni n. Poniamo allora $Z_n = S_n/P[S_n]$ e mostriamo che la successione $(Z_n)_{n\geq 1}$ converge in $\mathcal{L}^2(P)$ verso la costante 1. A questo scopo, osserviamo che, essendo $P[Z_n] = 1$, si ha

$$P\left[(Z_n - 1)^2\right] = \mathrm{Var}[Z_n] = (P[S_n])^{-2}\,\mathrm{Var}[S_n]. \tag{1.4}$$

D'altra parte, essendo le X_n bernoulliane e a due a due indipendenti, se si pone $p_j = P(A_j)$, $q_j = 1 - p_j$, si può scrivere

$$\mathrm{Var}[S_n] = \sum_{j=1}^n \mathrm{Var}[X_j] = \sum_{j=1}^n p_j q_j \leq \sum_{j=1}^n p_j = P[S_n].$$

Dalla relazione (1.4) discende dunque

$$P\left[(Z_n - 1)^2\right] \leq (P[S_n])^{-1} \to 0.$$

È così provato che la successione $(Z_n)_{n\geq 1}$ converge verso la costante 1 in $\mathcal{L}^2(P)$, e quindi anche in probabilità. Ne segue che esiste una successione strettamente crescente $(m_k)_k$ d'interi, tale che la successione $(Z_{m_k})_k$ converga verso 1 quasi certamente. Per quasi ogni ω, si ha allora, al tendere di k all'infinito,

$$S_{m_k}(\omega) \sim P[S_{m_k}] \uparrow \infty.$$

Ciò prova che l'evento $\{\sup_n S_n = \infty\}$ è quasi certo. Poiché questo evento è identico all'evento $\left\{\sum_{n\geq 1} X_n = \infty\right\}$, ossia all'evento $\limsup_n A_n$, l'asserzione è dimostrata. □

(1.5) Osservazione. Nella sua formulazione originaria, il secondo lemma di Borel-Cantelli era enunciato per una successione $(A_n)_n$ di eventi indipendenti. La formulazione (1.2) sopra dimostrata, nella quale questa ipotesi è sostituita dall'ipotesi più debole dell'indipendenza *a due a due*, è dovuta a Erdös e Rényi.

(1.6) Osservazione. Nel caso di una successione $(A_n)_n$ di eventi indipendenti, la legge zero-uno di Kolmogorov permette solo di affermare che l'evento $\limsup_n A_n$, essendo terminale rispetto alla successione $(X_n)_n$ delle funzioni indicatrici degli A_n, è degenere, ossia ha probabilità eguale a 0 o a 1. I due lemmi di Borel-Cantelli si possono considerare come un'utile precisazione di questa dicotomia.

2. Due semplici applicazioni

Come esempio di applicazione del primo lemma di Borel-Cantelli, vogliamo ora esporre una dimostrazione del seguente classico risultato (del quale ci siamo sopra serviti per dimostrare il secondo lemma di Borel-Cantelli).

(2.1) Teorema. *Sullo spazio probabilizzato $(\Omega, \mathcal{A}, P)$ sia $(X_n)_{n\geq 1}$ una successione di variabili aleatorie reali, la quale converga in probabilità verso una variabile aleatoria reale X.*

Allora è possibile estrarre da $(X_n)_n$ una sottosuccessione che converga verso X quasi certamente.

Dimostrazione. Sfruttando l'ipotesi di convergenza in probabilità, è possibile costruire (per induzione) una successione strettamente crescente $(m_k)_{k\geq 1}$ d'interi, in modo tale che, per ogni indice k, si abbia

$$P\{|X_{m_k} - X| > 1/k\} < 2^{-k}.$$

Allora, grazie al primo lemma di Borel-Cantelli, l'evento

$$\limsup_k \{|X_{m_k} - X| > 1/k\}$$

è trascurabile. Dunque il suo complementare, ossia l'evento

$$\liminf_k \{|X_{m_k} - X| \leq 1/k\},$$

è quasi certo. In ogni punto ω di quest'ultimo evento, la successione $(|X_{m_k} - X|)_k$, essendo definitivamente maggiorata dalla successione infinitesima $(1/k)_k$, è a sua volta infinitesima. Dunque la successione $(X_{m_k})_k$ converge verso X quasi certamente. □

Come esempio di applicazione del secondo lemma di Borel-Cantelli, impiegheremo ora questo lemma per costruire una successione di variabili aleatorie la quale converga in probabilità, ma non quasi certamente. A questo scopo, osserviamo che, mediante lo schema delle prove indipendenti, è possibile costruire uno spazio probabilizzato $(\Omega, \mathcal{A}, P)$ e, su di esso, una successione $(X_n)_{n\geq 1}$ di variabili aleatorie indipendenti, a valori in $\{0, 1\}$, tale che, per ogni n, la variabile aleatoria X_n abbia legge di Bernoulli di parametro $1/n$. La successione così costruita converge in $\mathcal{L}^1(P)$, e quindi in probabilità, verso la costante 0. Tuttavia essa non converge quasi certamente verso 0. Infatti l'insieme costituito dagli elementi ω di Ω tali che la successione $\big(X_n(\omega)\big)_n$ *non* converga verso 0 è identico all'insieme $\limsup_n \{X_n = 1\}$, il quale, grazie al secondo lemma di Borel-Cantelli, è un evento quasi certo.

La speranza condizionale

1. Definizione di speranza condizionale

Supporremo fissati uno spazio probabilizzato $(\Omega, \mathcal{A}, P)$ e una sottotribù $\mathcal{F}$ di $\mathcal{A}$. Tutte le variabili aleatorie considerate saranno tacitamente intese, se non diversamente specificato, come variabili aleatorie su $(\Omega, \mathcal{A}, P)$. Inoltre le variabili aleatorie misurabili rispetto alla sottotribù $\mathcal{F}$ saranno di preferenza denotate con lettere del tipo U, V, W. Ci sarà utile l'osservazione seguente.

(1.1) Osservazione. Se ν è una misura su $\mathcal{A}$, e se μ denota la sua restrizione a $\mathcal{F}$, è possibile vedere μ come l'immagine di ν mediante l'applicazione identica di Ω, considerata come applicazione (misurabile) dello spazio $(\Omega, \mathcal{A})$ nello spazio $(\Omega, \mathcal{F})$. Pertanto, grazie al teorema sull'integrazione rispetto a una misura immagine, se U è una variabile aleatoria reale misurabile rispetto a $\mathcal{F}$, l'integrabilità di U rispetto a ν equivale all'integrabilità di U rispetto a μ. Inoltre, se queste due condizioni equivalenti d'integrabilità sono soddisfatte, si ha $\int U \, \mathrm{d}\nu = \int U \, \mathrm{d}\mu$.

Ricordiamo che, per la tribù $\mathcal{F}$, una *base* è un sistema di generatori $\mathcal{J}$, stabile per l'operazione d'intersezione di due insiemi, tale che l'intero spazio Ω sia unione di una successione di elementi di $\mathcal{J}$. Ricordiamo inoltre il seguente *criterio per la coincidenza di due misure*: se $\mathcal{J}$ è una base per la tribù $\mathcal{F}$, allora due misure su $\mathcal{F}$, le cui restrizioni a $\mathcal{J}$ coincidano e siano finite, sono identiche.

(1.2) Definizione. Una variabile aleatoria X si dirà *ortogonale* (nello spazio $\mathcal{L}^1(P)$) alla sottotribù $\mathcal{F}$ se appartiene a $\mathcal{L}^1(P)$ e verifica la relazione $\int UX \, \mathrm{d}P = 0$ per ogni variabile aleatoria reale U, misurabile rispetto a $\mathcal{F}$, tale che il prodotto UX sia integrabile.

(1.3) Proposizione. *Sia X un elemento di $\mathcal{L}^1(P)$. Sia inoltre $\mathcal{J}$ una base per la tribù $\mathcal{F}$. Affinché X sia ortogonale a $\mathcal{F}$, è (necessario e) sufficiente che si abbia $\int_A X \, \mathrm{d}P = 0$ per ogni elemento A di $\mathcal{J}$.*

Dimostrazione. Supposta soddisfatta la condizione dell'enunciato, si considerino, sulla tribù $\mathcal{A}$, le due misure ν_1, ν_2 definite, rispetto a P, dalle densità X^+, X^-:

$$\nu_1 = X^+ \cdot P, \qquad \nu_2 = X^- \cdot P.$$

Esse coincidono su $\mathcal{J}$, dunque su $\mathcal{F}$. Si denoti con μ la loro comune restrizione a $\mathcal{F}$. Allora, se U è una variabile aleatoria reale, misurabile rispetto a $\mathcal{F}$, tale che il prodotto UX sia integrabile, si ha

$$\begin{aligned} \int UX \, \mathrm{d}P &= \int UX^+ \mathrm{d}P - \int UX^- \mathrm{d}P \\ &= \int U \, \mathrm{d}\nu_1 - \int U \, \mathrm{d}\nu_2 \\ &= \int U \, \mathrm{d}\mu - \int U \, \mathrm{d}\mu = 0 \end{aligned}$$

(dove la penultima eguaglianza è dovuta all'Osservazione (1.1)). Ciò prova che X è ortogonale a $\mathcal{F}$. □

(1.4) Corollario. *(a) Ogni variabile aleatoria reale trascurabile secondo P è ortogonale a $\mathcal{F}$ nello spazio $\mathcal{L}^1(P)$.*

(b) Nello spazio $\mathcal{L}^1(P)$, gli elementi ortogonali a $\mathcal{F}$ formano un sottospazio vettoriale.

(c) Se un elemento X di $\mathcal{L}^1(P)$ è ortogonale a $\mathcal{F}$, tale è anche ogni elemento di $\mathcal{L}^1(P)$ della forma UX, con U variabile aleatoria reale misurabile rispetto a $\mathcal{F}$.

(d) Affinché una variabile aleatoria reale V, misurabile rispetto a $\mathcal{F}$, sia, al tempo stesso, ortogonale a $\mathcal{F}$, occorre (e basta) che essa sia trascurabile.

Dimostrazione. Le quattro affermazioni si deducono immediatamente dalla proposizione precedente. Per provare l'affermazione (c), basta sfruttare, per ogni elemento A di $\mathcal{F}$, la relazione $\int_A UX\,\mathrm{d}P = \int (I_A U)X\,\mathrm{d}P$. Per provare l'affermazione (d), basta osservare che, se V è una variabile aleatoria reale misurabile rispetto a $\mathcal{F}$, i due eventi $\{V > 0\}$, $\{V < 0\}$ appartengono a $\mathcal{F}$, sicché l'ortogonalità di V a $\mathcal{F}$ implica la relazione

$$\int |V|\,\mathrm{d}P = \int_{\{V>0\}} V\,\mathrm{d}P - \int_{\{V<0\}} V\,\mathrm{d}P = 0. \qquad \square$$

(1.5) Definizione. Dato un elemento X di $\mathcal{L}^1(P)$, si chiama *versione della speranza condizionale* di X rispetto alla sottotribù $\mathcal{F}$ un elemento V di $\mathcal{L}^1(P)$, misurabile rispetto a $\mathcal{F}$, tale che la differenza $X - V$ sia ortogonale a $\mathcal{F}$.

È evidente che, se X è misurabile rispetto a $\mathcal{F}$, allora una versione della speranza condizionale di X rispetto a $\mathcal{F}$ è la variabile aleatoria X stessa. Così pure, se X è ortogonale a $\mathcal{F}$, allora una versione della speranza condizionale di X rispetto a $\mathcal{F}$ è la costante 0.

(1.6) Teorema. *Dato un elemento X di $\mathcal{L}^1(P)$, esiste sempre una versione V della speranza condizionale di X rispetto a $\mathcal{F}$. Inoltre V si può prendere positiva se tale è X.*

Dimostrazione. Mettiamoci senz'altro nel caso in cui sia $X \geq 0$. (A questo caso ci si può ricondurre considerando la decomposizione $X = X^+ - X^-$.) Consideriamo la misura $X \cdot P$. Questa è (in modo banale) assolutamente continua rispetto a P. Perciò la sua restrizione ν alla sottotribù $\mathcal{F}$ è assolutamente continua rispetto all'analoga restrizione di P. È dunque possibile, grazie al teorema di Radon-Nikodým, scrivere ν nella forma $\nu = V \cdot Q$, dove Q denota la restrizione di P a $\mathcal{F}$, mentre V è un opportuno elemento positivo di $\mathcal{L}^1(Q)$. Per ogni elemento A di $\mathcal{F}$, si ha allora (tenendo conto di (1.1))

$$\int_A (X - V)\,\mathrm{d}P = \int_A X\,\mathrm{d}P - \int_A V\,\mathrm{d}Q = \nu(A) - \nu(A) = 0.$$

Dalla Proposizione (1.3) (nella quale si prenda $\mathcal{J}$ coincidente con l'intera tribù $\mathcal{F}$) discende dunque che la differenza $X - V$ è ortogonale a $\mathcal{F}$. Ciò basta per concludere che V è una versione della speranza condizionale di X rispetto a $\mathcal{F}$. $\square$

(1.7) Teorema. *Siano dati un elemento X di $\mathcal{L}^1(P)$, una versione V della speranza condizionale di X rispetto a $\mathcal{F}$ e un'ulteriore variabile aleatoria reale V', misurabile rispetto a $\mathcal{F}$. Le due condizioni seguenti sono allora equivalenti:*

(a) V' è anch'essa una versione della speranza condizionale di X rispetto a $\mathcal{F}$.

(b) V' è equivalente a V secondo P.

Dimostrazione. Per provare l'implicazione (a) $\Rightarrow$ (b), osserviamo che, se è soddisfatta la condizione (a), allora la differenza $V' - V$, da un lato, è misurabile rispetto a $\mathcal{F}$; dall'altro, è ortogonale a $\mathcal{F}$ in virtù di (1.4) (b) perché può essere messa nella forma $(X-V)-(X-V')$, dove le due differenze $X-V$, $X-V'$ sono ortogonali a $\mathcal{F}$. Essa è dunque trascurabile in virtù di (1.4) (d).

L'implicazione (b) $\Rightarrow$ (a) discende poi dalle affermazioni (a) e (b) di (1.4). □

In conclusione, i risultati (1.6) e (1.7) mostrano che ogni elemento X di $\mathcal{L}^1(P)$ può essere decomposto, in modo essenzialmente unico, nella somma di due elementi di $\mathcal{L}^1(P)$, dei quali il primo sia misurabile rispetto a $\mathcal{F}$ e il secondo sia ortogonale a $\mathcal{F}$. Precisamente, per avere una tal decomposizione, basta scrivere X nella forma $X = V + (X-V)$, dove V sia una versione della speranza condizionale di X rispetto a $\mathcal{F}$.

(1.8) Definizione. Dato un elemento X di $\mathcal{L}^1(P)$, l'insieme costituito da tutte le versioni della speranza condizionale di X rispetto a $\mathcal{F}$ si chiama *la speranza condizionale* di X rispetto a $\mathcal{F}$ e si denota col simbolo

$$P[X \mid \mathcal{F}]. \tag{1.9}$$

Risulta da (1.7) che, se si denota con Q la restrizione di P a $\mathcal{F}$, il simbolo (1.9) rappresenta un elemento dello spazio quoziente $L^1(Q)$, ossia una delle classi di equivalenza che, nello spazio $\mathcal{L}^1(Q)$, sono determinate dalla relazione di equivalenza indotta da Q.

Nel caso particolare in cui $\mathcal{F}$ sia la tribù generata da una variabile aleatoria Z (a valori in un arbitrario spazio misurabile), si usa anche, in luogo della notazione (1.9), la notazione $P[X \mid Z]$.

Alla luce della Proposizione (1.3), la nozione di speranza condizionale può essere così caratterizzata:

(1.10) Proposizione. *Siano X un elemento di $\mathcal{L}^1(P)$ e $\mathcal{J}$ una base per la tribù $\mathcal{F}$. Sia inoltre V una variabile aleatoria reale misurabile rispetto a $\mathcal{F}$. Le condizioni che seguono sono allora equivalenti:*

(a) V è una versione di $P[X \mid \mathcal{F}]$,

(b) V appartiene a $\mathcal{L}^1(P)$ e, per ogni elemento A di $\mathcal{J}$, si ha $\int_A V \, dP = \int_A X \, dP$.

(1.11) Un esempio elementare. Dato un elemento X di $\mathcal{L}^1(P)$, si denoti con V una versione di $P[X \mid \mathcal{F}]$. Si supponga inoltre che esista un elemento H di $\mathcal{F}$, non trascurabile, che sia un *atomo* per $\mathcal{F}$, nel senso che la traccia di $\mathcal{F}$ su H sia ridotta alla tribù banale $\{H, \emptyset\}$. È chiaro che la variabile aleatoria V è allora necessariamente costante su H. Detto a il suo valore costante su H, dalla definizione di speranza condizionale si ricava

$$\int_H X \, dP = \int_H a \, dP = a \, P(H),$$

ossia

$$a = P(H)^{-1} \int_H X \, dP.$$

In altre parole, il valore costante di V su H coincide con la media integrale di X su H: media che, com'è noto, è eguale alla speranza di X calcolata secondo la misura di probabilità condizionale P_H.

Si supponga ora, in particolare, che esista una partizione numerabile $\mathcal{H}$ di Ω, costituita da eventi non trascurabili, tale che la tribù $\mathcal{F}$ coincida con la tribù generata da $\mathcal{H}$. Allora ogni elemento di $\mathcal{H}$ è un atomo per la tribù $\mathcal{F}$. In questo caso particolare si può dunque affermare che esiste un'unica versione della speranza condizionale di X rispetto a $\mathcal{F}$ e che quest'unica versione V è la funzione che, su ciascun elemento H della partizione $\mathcal{H}$, coincide con la costante $\int X \, dP_H$. In formule:

$$V = \sum_{H \in \mathcal{H}} I_H \int X \, dP_H.$$

2. Proprietà della speranza condizionale

(2.1) Proposizione. *Si supponga che l'elemento X di $\mathcal{L}^1(P)$ sia indipendente da $\mathcal{F}$ (ciò che avviene, in particolare, quando $\mathcal{F}$ sia degenere). Allora una versione della speranza condizionale $P[X \mid \mathcal{F}]$ è la costante $c = \int X \, dP$.*

Dimostrazione. Basta osservare che la variabile aleatoria $X - c$, essendo centrata e indipendente da $\mathcal{F}$, ha integrale nullo su ogni elemento di $\mathcal{F}$, e quindi è ortogonale a $\mathcal{F}$ (si veda (1.3)). □

(2.2) Proposizione. *L'applicazione $X \mapsto P[X \mid \mathcal{F}]$ dello spazio $\mathcal{L}^1(P)$ nello spazio quoziente $L^1(Q)$ (dove Q denota la restrizione di P a $\mathcal{F}$) è lineare e isotona.*

Dimostrazione. La linearità è immediata. L'isotonia, grazie alla linearità, discende dal fatto che, se X è un elemento positivo dello spazio vettoriale ordinato $\mathcal{L}^1(P)$, allora l'elemento $P[X \mid \mathcal{F}]$ dello spazio vettoriale ordinato $L^1(Q)$ è a sua volta positivo perché la speranza condizionale di X rispetto a $\mathcal{F}$ ammette una versione positiva (si veda (1.6)). □

(2.3) Proposizione. *Siano X un elemento di $\mathcal{L}^1(P)$ e U una variabile aleatoria reale, misurabile rispetto a $\mathcal{F}$, tale che il prodotto UX sia integrabile.*

Allora una versione della speranza condizionale $P[UX \mid \mathcal{F}]$ si ottiene moltiplicando per U una versione di $P[X \mid \mathcal{F}]$.

Dimostrazione. Ci si può limitare a dimostrare il risultato parziale relativo al caso in cui le due variabili aleatorie X, U siano entrambe positive. Infatti basterà, nel caso generale, applicare questo risultato parziale a ciascuna delle quattro speranze condizionali seguenti:

$$P\big[U^+X^+ \,|\, \mathcal{F}\big], \quad P\big[U^+X^- \,|\, \mathcal{F}\big], \quad P\big[U^-X^+ \,|\, \mathcal{F}\big], \quad P\big[U^-X^- \,|\, \mathcal{F}\big].$$

Supponiamo dunque X e U positive, e denotiamo con V una versione, anch'essa positiva, di $P\big[X \,|\, \mathcal{F}\big]$. Basterà dimostrare che il prodotto UV è integrabile: infatti il prodotto $U(X - V)$ sarà allora anch'esso integrabile, e quindi ortogonale a $\mathcal{F}$ in virtù di (1.4) (c). Per provare che UV è integrabile, consideriamo, per ogni intero positivo n, la variabile aleatoria limitata $U_n = U \wedge n$. Essendo $X - V$ ortogonale a $\mathcal{F}$, tale è anche $U_n(X - V)$. Si ha dunque $P\big[U_n(X - V)\big] = 0$, ossia $P\big[U_nV\big] = P\big[U_nX\big]$. Di qui, facendo tendere n all'infinito, si deduce

$$P\big[UV\big] = P\big[UX\big] < \infty. \qquad \square$$

(2.4) Osservazione. Come sottoprodotto del risultato appena dimostrato, si ha che, se X è un elemento di $\mathcal{L}^1(P)$ e V è una versione di $P\big[X \,|\, \mathcal{F}\big]$, allora, per ogni variabile aleatoria reale U misurabile rispetto a $\mathcal{F}$, l'integrabilità di UX implica quella di UV. Questa implicazione non può essere invertita, come mostra l'esempio seguente.

(2.5) Esempio. Sia X una variabile aleatoria a valori nell'insieme $\{-1, 1\}$, che abbia come legge la ripartizione uniforme su questo insieme e che sia indipendente dalla tribù $\mathcal{F}$. Allora X è ortogonale a $\mathcal{F}$. Perciò, come versione V della speranza condizionale $P\big[X \,|\, \mathcal{F}\big]$, si può prendere la costante 0. Se ora U è una variabile aleatoria reale, misurabile rispetto alla tribù $\mathcal{F}$, ma non integrabile, la relazione $P\big[|UX|\big] = P\big[|U|\big] = \infty$ mostra che il prodotto UX non è integrabile, pur essendo nullo il prodotto UV.

(2.6) Proposizione. (Forma condizionale del Teorema di Beppo Levi) *Siano X un elemento di $\mathcal{L}^1(P)$ e $(X_n)_{n\geq 1}$ una successione crescente di elementi di $\mathcal{L}^1(P)$ con $\sup_n X_n = X$. Si denoti con V una versione di $P\big[X \,|\, \mathcal{F}\big]$ e, per ogni n, si denoti con V_n una versione di $P\big[X_n \,|\, \mathcal{F}\big]$. Allora l'evento $\{V_n \uparrow V\}$ è quasi certo.*

Dimostrazione. Poniamo $H = \bigcap_n \{V_n \leq V_{n+1} \leq V\}$. Grazie all'isotonia dell'operazione di speranza condizionale, H è un evento quasi certo. Inoltre, dal teorema di Beppo Levi discende la relazione $P[X_n] \uparrow P[X]$, che può anche essere scritta in ciascuna delle forme seguenti:

$$P[V_n] \uparrow P[V], \quad P[V_nI_H] \uparrow P[VI_H], \quad P\big[VI_H - \sup_n V_nI_H\big] = 0.$$

Si vede così che è quasi certo l'evento $H \cap \{V = \sup_n V_n\}$ (identico a $\{V_n \uparrow V\}$). $\square$

(2.7) Proposizione. (Forma condizionale del Lemma di Fatou) *Sia $(X_n)_{n\geq 1}$ una successione di elementi di $\mathcal{L}^1(P)$, tutti minorati da un medesimo elemento Z di $\mathcal{L}^1(P)$. Supposto che la variabile aleatoria $X = \liminf_n X_n$ appartenga a $\mathcal{L}^1(P)$, si denoti con V una versione di $P[X \mid \mathcal{F}]$ e con V_n una versione di $P[X_n \mid \mathcal{F}]$. Allora l'evento $\{V \leq \liminf_n V_n\}$ è quasi certo.*

Dimostrazione. Poniamo $X'_n = X_n \wedge X_{n+1} \wedge \cdots$ e denotiamo con V'_n una versione di $P[X'_n \mid \mathcal{F}]$. Si ha allora

$$Z \leq X'_n \leq X_n, \qquad X'_n \uparrow X.$$

Perciò i due eventi $\bigcap_n \{V'_n \leq V_n\}$ e $\{V'_n \uparrow V\}$ sono quasi certi. Basta allora osservare che, per ogni elemento ω della loro intersezione, si ha

$$V(\omega) = \lim_n V'_n(\omega) \leq \liminf_n V_n(\omega). \qquad \square$$

Naturalmente, la precedente forma condizionale del Lemma di Fatou ammette una versione "speculare", nella quale interviene il limite superiore, anziché il limite inferiore. Combinando insieme le due versioni, si ottiene il corollario seguente.

(2.8) Corollario. (Forma condizionale del Teorema di Lebesgue) *Sia $(X_n)_{n\geq 1}$ una successione di elementi di $\mathcal{L}^1(P)$, dominata in $\mathcal{L}^1(P)$ e convergente puntualmente verso una variabile aleatoria X. Si denoti con V una versione di $P[X \mid \mathcal{F}]$ e con V_n una versione di $P[X_n \mid \mathcal{F}]$. Allora la successione $(V_n)_{n\geq 1}$ converge quasi certamente verso V.*

È chiaro che, se X è un elemento di $\mathcal{L}^1(P)$, e V è una versione di $P[X \mid \mathcal{F}]$, allora, per ogni coppia a, b di numeri reali, una versione di $P[aX+b \mid \mathcal{F}]$ è la variabile aleatoria $aV + b$. Se, anziché considerare una variabile aleatoria della forma $aX + b$ (ossia la composizione di X con una funzione lineare affine $x \mapsto ax + b$), si considera la composizione di X con una funzione convessa, si ottiene il teorema seguente.

(2.9) Teorema. (Forma condizionale della diseguaglianza di Jensen) *Dato un elemento X di $\mathcal{L}^1(P)$, si denoti con V una versione di $P[X \mid \mathcal{F}]$. Sia inoltre g una funzione reale convessa su $\mathbb{R}$, tale che la funzione composta $g \circ X$ sia integrabile, e si denoti con W una versione di $P[g \circ X \mid \mathcal{F}]$. Allora l'evento $\{g \circ V \leq W\}$ è quasi certo.*

Dimostrazione. La funzione g è l'inviluppo superiore di una famiglia numerabile $(f_i)_{i\in I}$ di funzioni lineari affini. Per ogni indice i, la diseguaglianza $f_i \leq g$ implica

$$P[f_i \circ X \mid \mathcal{F}] \leq P[g \circ X \mid \mathcal{F}].$$

Poiché una versione del primo membro di quest'ultima diseguaglianza è $f_i \circ V$, ne segue che l'evento $\{f_i \circ V \leq W\}$ è quasi certo. Tale è dunque anche l'intersezione degli eventi di questa forma, ossia l'evento $\{g \circ V \leq W\}$. $\square$

(2.10) Corollario. *Dato un elemento X di $\mathcal{L}^p(P)$, con $1 \le p < \infty$, siano V una versione di $P[X \mid \mathcal{F}]$ e W una versione di $P[|X|^p \mid \mathcal{F}]$. Allora l'evento $\{|V|^p \le W\}$ è quasi certo. In particolare, si ha*

$$\int |V|^p \mathrm{d}P \le \int W \,\mathrm{d}P = \int |X|^p \mathrm{d}P$$

(ossia: "l'operatore di speranza condizionale, applicato a un elemento di $\mathcal{L}^p(P)$, ne riduce la norma").

Dimostrazione. Basta applicare il teorema precedente con $g(x) = |x|^p$. □

(2.11) Osservazione. Affinché un elemento X di $\mathcal{L}^1(P)$ verifichi la relazione $P[X \mid \mathcal{F}] = 0$, occorre e basta che esso sia ortogonale a $\mathcal{F}$. Pertanto, grazie alla linearità dell'operazione di speranza condizionale, affinché due elementi X, Y di $\mathcal{L}^1(P)$ verifichino la relazione $P[X \mid \mathcal{F}] = P[Y \mid \mathcal{F}]$, occorre e basta che la loro differenza sia ortogonale a $\mathcal{F}$. Ne discende, in particolare, la proposizione seguente.

(2.12) Proposizione. *Sia $\mathcal{A}'$ una tribù, con $\mathcal{F} \subset \mathcal{A}' \subset \mathcal{A}$. Sia inoltre X un elemento di $\mathcal{L}^1(P)$, e si denoti con X' una versione di $P[X \mid \mathcal{A}']$. Si ha allora*

$$P[X \mid \mathcal{F}] = P[X' \mid \mathcal{F}].$$

Dimostrazione. Basta osservare che $X - X'$ è ortogonale a $\mathcal{A}'$, quindi a $\mathcal{F}$. □

La proposizione appena dimostrata esprime una proprietà molto intuitiva, e spesso utile, dell'operazione di speranza condizionale: per ottenere una versione della speranza condizionale di X rispetto a $\mathcal{F}$, si può prendere dapprima una versione X' della speranza condizionale di X rispetto a una tribù ausiliaria $\mathcal{A}'$, compresa tra $\mathcal{F}$ e $\mathcal{A}$, e quindi passare a una versione della speranza condizionale di X' rispetto a $\mathcal{F}$. Per questo motivo, è naturale dare a questa proprietà il nome di *Proprietà della tribù intermedia*.

3. Speranza condizionale secondo una nuova misura*

Ci proponiamo di dimostrare una proposizione che è utile in situazioni nelle quali intervenga una speranza condizionale calcolata, anziché secondo la misura di probabilità P, secondo una nuova misura di probabilità P', assolutamente continua rispetto a P. A questo scopo, cominciamo con l'introdurre una notazione. Se U è una variabile aleatoria reale, denoteremo con U^* la variabile aleatoria così definita:

$$U^*(\omega) = \begin{cases} 1/U(\omega) & \text{se } U(\omega) \neq 0, \\ 0 & \text{altrimenti.} \end{cases} \tag{3.1}$$

Si osservi che, con questa notazione, si ha

$$UU^* = I_{\{U \neq 0\}}. \tag{3.2}$$

Supponiamo ora che K sia una versione della densità di P' rispetto a P, ossia un elemento di $\mathcal{L}^1(P)$ tale che si abbia $P' = K \cdot P$. Vogliamo provare che, se X è una variabile aleatoria reale integrabile rispetto a P' (cioè tale che XK sia integrabile rispetto a P), allora una versione di $P'[X \mid \mathcal{F}]$ si ottiene dividendo una versione di $P[XK \mid \mathcal{F}]$ per una versione di $P[K \mid \mathcal{F}]$. Più precisamente, vogliamo provare la proposizione seguente.

(3.3) Proposizione. *Nelle ipotesi precedenti, siano V una versione di $P[XK \mid \mathcal{F}]$ e U una versione di $P[K \mid \mathcal{F}]$. Allora una versione di $P'[X \mid \mathcal{F}]$ è VU^* (dove U^* è la variabile aleatoria definita da (3.1)).*

Dimostrazione. Cominciamo con l'osservare che si ha

$$P'\{U = 0\} = \int_{\{U=0\}} K \, dP = \int_{\{U=0\}} U \, dP = 0.$$

Ne segue, grazie a (3.2),

$$UU^* \sim 1 \pmod{P'}. \tag{3.4}$$

Sia ora W una versione di $P'[X \mid \mathcal{F}]$. Grazie al teorema sull'integrazione rispetto a una misura definita da una densità, l'appartenenza di X e di W a $\mathcal{L}^1(P')$ si traduce nell'appartenenza di XK e di WK a $\mathcal{L}^1(P)$. Così pure, l'ortogonalità di $X - W$ a $\mathcal{F}$ nello spazio $\mathcal{L}^1(P')$ si traduce nell'ortogonalità di $XK - WK$ a $\mathcal{F}$ nello spazio $\mathcal{L}^1(P)$, ossia (si veda (2.11)) nell'eguaglianza

$$P[WK \mid \mathcal{F}] = P[XK \mid \mathcal{F}].$$

A sua volta, questa eguaglianza, poiché $P[WK \mid \mathcal{F}]$ ammette WU come versione (si veda (2.3)), si può esprimere mediante la relazione $WU \sim V \pmod{P}$, la quale, per l'assoluta continuità di P' rispetto a P, implica $WU \sim V \pmod{P'}$. Da quest'ultima relazione discende infine, grazie a (3.4),

$$W \sim WUU^* \sim VU^* \pmod{P'}.$$

Si vede così che VU^*, essendo una variabile aleatoria reale misurabile rispetto a $\mathcal{F}$ ed equivalente a W secondo P', è, al pari di W, una versione di $P'[X \mid \mathcal{F}]$. □

(3.5) Corollario. *Nelle stesse ipotesi della proposizione precedente, si supponga per giunta che la misura P' coincida con P sulla tribù $\mathcal{F}$. Si ha allora*

$$P'[X \mid \mathcal{F}] = P[XK \mid \mathcal{F}].$$

Dimostrazione. L'ipotesi che P' coincida con P su $\mathcal{F}$ significa che una versione di $P[K \mid \mathcal{F}]$ è la costante 1. Basta dunque applicare la proposizione precedente nella quale si prenda $U = 1$. □

Uniforme integrabilità

1. Definizione e criterio di la Vallée-Poussin

Supporremo fissato uno spazio probabilizzato $(\Omega, \mathcal{A}, P)$. Tutte le variabili aleatorie considerate saranno tacitamente intese come variabili aleatorie *reali* su $(\Omega, \mathcal{A}, P)$. Cominciamo con una proposizione del tutto elementare.

(1.1) Proposizione. *Sia X una variabile aleatoria, e si denoti con μ la sua legge. Le condizioni che seguono sono tra loro equivalenti:*

(a) X è integrabile, cioè $\int |X| \, \mathrm{d}P < \infty$.

(b) Per ogni numero reale ϵ maggiore di zero, esiste un numero reale positivo t tale che si abbia $\int_{\{|X|>t\}} |X| \, \mathrm{d}P \leq \epsilon$, ossia $\int_{[-t,t]^c} |x| \, \mu(\mathrm{d}x) \leq \epsilon$.

Dimostrazione. (a) $\Rightarrow$ (b): Si supponga che X sia integrabile, cioè che la misura $\nu = |X| \cdot P$ (definita dalla densità $|X|$ rispetto alla misura base P) sia finita. Si scelga una successione crescente $(t_n)_n$ di numeri reali positivi, con $t_n \uparrow \infty$, e si ponga $A_n = \{|X| > t_n\}$. Si ha allora $A_n \downarrow \emptyset$, e quindi, per la continuità di ν sulle successioni decrescenti, $\nu(A_n) \downarrow 0$. Dunque, fissato il numero ϵ maggiore di zero, si ha, per n abbastanza grande, $\nu(A_n) \leq \epsilon$, ossia $\int_{\{|X|>t_n\}} |X| \, \mathrm{d}P \leq \epsilon$. La condizione (b) è dunque verificata.

(b) $\Rightarrow$ (a): Si supponga verificata la condizione (b). Allora (prendendo, in particolare, $\epsilon = 1$) si vede che esiste un numero reale positivo t tale che si abbia $\int_{\{|X|>t\}} |X| \, \mathrm{d}P \leq 1$, e quindi

$$\int |X| \, \mathrm{d}P = \int_{\{|X|\leq t\}} |X| \, \mathrm{d}P + \int_{\{|X|>t\}} |X| \, \mathrm{d}P \leq t + 1.$$

Ciò prova che X è integrabile. □

(1.2) Definizione. Sia $\mathcal{H}$ un insieme di variabili aleatorie. Si dice che esso è *uniformemente integrabile* se la condizione (b) della proposizione precedente è verificata per ogni elemento X di $\mathcal{H}$, e per giunta *in modo uniforme* al variare di X, cioè se, per ogni numero reale ϵ maggiore di zero, esiste un numero reale positivo t tale che si abbia $\sup_{X\in\mathcal{H}} \int_{\{|X|>t\}} |X| \, \mathrm{d}P \leq \epsilon$. In modo più conciso, si può dunque dire che l'insieme $\mathcal{H}$ è uniformemente integrabile se risulta

$$\inf_{t\in\mathbb{R}_+} \sup_{X\in\mathcal{H}} \int_{\{|X|>t\}} |X| \, \mathrm{d}P = 0.$$

Se poi $(X_i)_{i\in I}$ è un'arbitraria *famiglia* di variabili aleatorie, si dice che essa è *uniformemente integrabile* se tale è l'insieme costituito dai suoi termini, cioè se risulta

$$\inf_{t\in\mathbb{R}_+} \sup_{i\in I} \int_{\{|X_i|>t\}} |X_i| \, \mathrm{d}P = 0. \tag{1.3}$$

In tal caso, si dice anche, in modo un po' meno formale, che "le X_i sono uniformemente integrabili".

(1.4) Osservazione. L'uniforme integrabilità di una famiglia $(X_i)_{i\in I}$ di variabili aleatorie è una proprietà che dipende solo dalle *leggi* delle X_i. Se infatti, per ogni indice i, si denota con μ_i la legge di X_i, la relazione (1.3) si può scrivere nella forma

$$\inf_{t\in\mathbb{R}_+} \sup_{i\in I} \int_{[-t,t]^c} |x|\, \mu_i(\mathrm{d}x) = 0. \tag{1.5}$$

Di qui, tenendo conto della Proposizione (1.1), si deduce, in particolare, che *una famiglia di variabili aleatorie integrabili e isonome è uniformemente integrabile.*

Una condizione *sufficiente* per l'uniforme integrabilità è fornita dal criterio seguente.

(1.6) Teorema. (la Vallée-Poussin) *Si supponga che esista un numero reale δ maggiore di zero, tale che la famiglia $(X_i)_{i\in I}$ di variabili aleatorie sia limitata in $\mathcal{L}^{1+\delta}(P)$, nel senso che risulti*

$$c := \sup_{i\in I} P\left[|X_i|^{1+\delta}\right] < \infty. \tag{1.7}$$

Allora la famiglia $(X_i)_{i\in I}$ è uniformemente integrabile.

Dimostrazione. Si denoti con c il primo membro di (1.7). Si ha allora, per ogni indice i ed ogni numero reale t maggiore di zero,

$$\int_{\{|X_i|>t\}} |X_i|\,\mathrm{d}P \le \int |X_i|\,(|X_i|/t)^{\delta}\,\mathrm{d}P = t^{-\delta} P\left[|X_i|^{1+\delta}\right] \le t^{-\delta} c.$$

Poiché l'ultimo membro di questa relazione non dipende da i (e tende a zero al tendere di t all'infinito), ciò basta per concludere che la famiglia $(X_i)_{i\in I}$ è uniformemente integrabile. □

La proposizione seguente mostra che un'altra condizione *sufficiente* per l'uniforme integrabilità è quella di *dominazione*, presente nel teorema di Lebesgue. (Storicamente, la nozione di uniforme integrabilità fu introdotta da Giuseppe Vitali, nel 1907, proprio come un indebolimento di quella condizione.)

(1.8) Proposizione. *Si supponga che la famiglia di variabili aleatorie $(X_i)_{i\in I}$ sia dominata in $\mathcal{L}^1(P)$, cioè sia tale che esista una variabile aleatoria integrabile Z che la domini, nel senso che verifichi la relazione $|X_i| \le Z$ per ogni indice i.*

Allora $(X_i)_{i\in I}$ è uniformemente integrabile.

Dimostrazione. Fissato un numero ϵ maggiore di zero, esiste, grazie a (1.1), un numero reale positivo t tale che si abbia $\int_{\{Z>t\}} Z\,\mathrm{d}P \le \epsilon$. Si ha allora, per ogni indice i,

$$\int_{\{|X_i|>t\}} |X_i|\,\mathrm{d}P \le \int_{\{|X_i|>t\}} Z\,\mathrm{d}P \le \int_{\{Z>t\}} Z\,\mathrm{d}P \le \epsilon,$$

e ciò prova che le X_i sono uniformemente integrabili. □

(1.9) Corollario. *Ogni famiglia finita $(X_i)_{i\in I}$ di variabili aleatorie integrabili è uniformemente integrabile.*

Dimostrazione. La variabile aleatoria $Z = \sup_{i\in I} |X_i|$ è integrabile (come inviluppo superiore di una famiglia finita di variabili aleatorie integrabili) e domina evidentemente la famiglia data. Questa è dunque uniformemente integrabile in virtù della proposizione precedente. □

2. Una condizione equivalente all'uniforme integrabilità

(2.1) Teorema. *Per una famiglia $(X_i)_{i\in I}$ di variabili aleatorie, le condizioni che seguono sono equivalenti:*

(a) La famiglia $(X_i)_{i\in I}$ è uniformemente integrabile.

(b) La famiglia $(X_i)_{i\in I}$ è limitata in $\mathcal{L}^1(P)$, ossia risulta

$$\sup_{i\in I} P[|X_i|] < \infty, \tag{2.2}$$

e inoltre, per ogni numero reale ϵ maggiore di zero, esiste un numero reale δ maggiore di zero, tale che, per ogni evento A con $P(A) \leq \delta$, si abbia

$$\sup_{i\in I} \int_A |X_i|\, dP \leq \epsilon. \tag{2.3}$$

Dimostrazione. (a) $\Rightarrow$ (b): Si supponga che le X_i siano uniformemente integrabili. Allora (prendendo, in particolare, $\epsilon = 1$) si vede che esiste un numero reale positivo t tale che si abbia $\sup_{i\in I} \int_{\{|X_i|>t\}} |X_i|\, dP \leq 1$, e quindi, per ogni indice i,

$$P[|X_i|] = \int_{\{|X_i|\leq t\}} |X_i|\, dP + \int_{\{|X_i|>t\}} |X_i|\, dP \leq t + 1.$$

Si vede così che è verificata la relazione (2.2). Per provare che è soddisfatta anche la seconda parte della condizione (b), fissiamo un numero reale ϵ maggiore di zero e osserviamo che, grazie all'ipotesi (a), esiste un numero reale t maggiore di zero, tale che si abbia

$$\sup_{i\in I} \int_{\{|X_i|>t\}} |X_i|\, dP \leq \epsilon/2.$$

Si ha allora, per ogni indice i ed ogni evento A,

$$\int_A |X_i|\, dP = \int_{A\cap\{|X_i|\leq t\}} |X_i|\, dP + \int_{A\cap\{|X_i|>t\}} |X_i|\, dP \leq tP(A) + \epsilon/2.$$

Ne segue che, per ogni evento A tale che si abbia $tP(A) \leq \epsilon/2$, ossia $P(A) \leq \epsilon/(2t)$, si ha

$$\sup_{i\in I} \int_A |X_i|\, dP \leq tP(A) + \epsilon/2 \leq \epsilon.$$

La condizione (b) è dunque soddisfatta.

(b) $\Rightarrow$ (a): Si supponga soddisfatta la condizione (b) e si denoti con c il primo membro di (2.2). Fissato il numero ϵ maggiore di zero, sia δ un numero reale maggiore di zero, tale che, per ogni evento A con $P(A) \leq \delta$, abbia luogo la relazione (2.3). Per ogni indice i ed ogni numero reale t maggiore di zero, la diseguaglianza di Markov fornisce

$$P\{|X_i| > t\} \leq t^{-1} P\,[|X_i|] \leq t^{-1} c.$$

Ne segue che, se il numero reale t maggiore di zero è tale che si abbia $t^{-1}c \leq \delta$, ossia $t \geq c/\delta$, allora, per ogni indice i, si ha $P\{|X_i| > t\} \leq \delta$, e quindi $\int_{\{|X_i|>t\}} |X_i|\,\mathrm{d}P \leq \epsilon$. La condizione (a) è dunque soddisfatta. □

(2.4) Corollario. *Si supponga che un insieme $\mathcal{H}$ di variabili aleatorie sia uniformemente integrabile. Tale è allora anche l'inviluppo convesso di $\mathcal{H}$, ossia l'insieme $\widetilde{\mathcal{H}}$ costituito da tutte le combinazioni lineari finite della forma*

$$\textstyle\sum_{k=1}^{n} a_k X_k, \quad \text{con} \quad X_k \in \mathcal{H}, \quad a_k \geq 0, \quad \sum_{k=1}^{n} a_k = 1.$$

Dimostrazione. Grazie al teorema precedente, l'ipotesi che $\mathcal{H}$ sia uniformemente integrabile significa che si ha $\sup_{X\in\mathcal{H}} P\,[|X|] < \infty$ e che, per ogni numero reale ϵ maggiore di zero, esiste un numero reale δ maggiore di zero, tale che, per ogni evento A con $P(A) \leq \delta$, si abbia $\sup_{X\in\mathcal{H}} \int_A |X|\,\mathrm{d}P \leq \epsilon$. È immediato riconoscere che queste condizioni sono verificate anche con $\widetilde{\mathcal{H}}$ in luogo di $\mathcal{H}$. □

3. Il teorema di Vitali

Ci proponiamo ora di esporre un fondamentale teorema, dovuto a Vitali, il quale estende e precisa, mediante il concetto di uniforme integrabilità, il teorema di Lebesgue sulla convergenza dominata. A questo scopo, cominceremo col dimostrare il lemma seguente, che è soltanto una prima, assai modesta, estensione di un caso particolare del teorema di Lebesgue, consistente semplicemente nel sostituire l'ipotesi di convergenza quasi certa con quella, più debole, di convergenza in probabilità.

(3.1) Lemma. *Sia $(X_n)_{n\geq 1}$ una successione di variabili aleatorie, la quale converga in probabilità verso la costante 0 e sia dominata in $\mathcal{L}^1(P)$.*

Si ha allora, al tendere di n all'infinito, $P\,[X_n] \to 0$.

Dimostrazione. Com'è ben noto, l'ipotesi che la successione $(X_n)_{n\geq 1}$ converga in probabilità verso 0 equivale al fatto che ogni successione estratta da $(X_n)_{n\geq 1}$ ammetta, a sua volta, una sottosuccessione convergente verso 0 quasi certamente.

Si tratta di dimostrare che la successione (limitata) di numeri reali $(P\,[X_n])_{n\geq 1}$ non ammette alcun valore di aderenza diverso da zero. Sia dunque λ un suo valore di aderenza: ciò vuol dire che esiste una successione $(m_k)_{k\geq 1}$ d'interi, strettamente crescente, tale che, al tendere di k all'infinito, si abbia $P\,[X_{m_k}] \to \lambda$. Grazie all'osservazione iniziale, si può supporre, senza ledere la generalità, che la sottosuccessione $(X_{m_k})_{k\geq 1}$ converga verso 0 quasi certamente. Applicando allora ad essa il teorema di Lebesgue sulla convergenza dominata, si trova $\lambda = 0$, come si voleva dimostrare. □

Allo scopo di dimostrare il teorema di Vitali, sarà utile premettere anche la proposizione seguente (non priva di un suo interesse autonomo). Essa si distingue dai criteri di uniforme integrabilità finora incontrati (che riguardano tutti un'arbitraria famiglia, o un arbitrario insieme, di variabili aleatorie) per il fatto che è circoscritta al caso speciale di una *successione* di variabili aleatorie (alle quali s'impone, per giunta, la condizione preliminare di essere *integrabili*).

(3.2) Proposizione. *Sia $(X_n)_{n\geq 1}$ una successione di variabili aleatorie integrabili. Le condizioni che seguono sono allora equivalenti:*

(a) La successione $(X_n)_{n\geq 1}$ è uniformemente integrabile.

(b) Si ha $\inf_{t\in\mathbb{R}_+} \limsup_n \int_{\{|X_n|>t\}} |X_n|\,\mathrm{d}P = 0$.

Dimostrazione. La condizione (a), che equivale alla relazione

$$\inf_{t\in\mathbb{R}_+} \sup_n \int_{\{|X_n|>t\}} |X_n|\,\mathrm{d}P = 0,$$

implica la condizione (b), grazie all'ovvia maggiorazione

$$\limsup_n \int_{\{|X_n|>t\}} |X_n|\,\mathrm{d}P \leq \sup_n \int_{\{|X_n|>t\}} |X_n|\,\mathrm{d}P.$$

(b) $\Rightarrow$ (a): Si supponga soddisfatta la condizione (b). Allora, fissato il numero reale ϵ maggiore di zero, esiste un numero reale positivo t_0, tale che si abbia

$$\limsup_n \int_{\{|X_n|>t_0\}} |X_n|\,\mathrm{d}P \leq \epsilon.$$

In corrispondenza, esiste dunque un intero $m \geq 1$, tale che si abbia

$$\int_{\{|X_n|>t_0\}} |X_n|\,\mathrm{d}P \leq \epsilon \quad \text{per} \quad n \geq m.$$

D'altra parte, poiché la m–upla $(X_n)_{1\leq n\leq m}$ è uniformemente integrabile (si veda (1.9)), esiste un numero reale positivo t_1 tale che si abbia

$$\int_{\{|X_n|>t_1\}} |X_n|\,\mathrm{d}P \leq \epsilon \quad \text{per} \quad n \leq m.$$

Se allora si denota con t il massimo tra i due numeri t_0 e t_1, si vede che la relazione

$$\int_{\{|X_n|>t\}} |X_n|\,\mathrm{d}P \leq \epsilon$$

ha luogo per ogni n, e ciò prova che la successione $(X_n)_{n\geq 1}$ è uniformemente integrabile. □

Siamo ora in grado di dimostrare facilmente il preannunciato teorema di Vitali.

(3.3) Teorema. (Vitali) *Siano date una successione $(X_n)_{n\geq 1}$ di variabili aleatorie integrabili e un'ulteriore variabile aleatoria X. Le condizioni che seguono sono allora equivalenti:*

(a) X è integrabile e $(X_n)_{n\geq 1}$ converge verso X in $\mathcal{L}^1(P)$ (nel senso che risulta $P[|X_n - X|] \to 0$):

(b) La successione $(X_n)_{n\geq 1}$ è uniformemente integrabile e converge in probabilità verso X:

Dimostrazione. (a) $\Rightarrow$ (b): Si supponga soddisfatta la condizione (a). Allora, innanzitutto, la successione $(X_n)_{n\geq 1}$ converge verso X in probabilità, grazie al fatto che, per ogni numero reale ϵ maggiore di zero, si ha (diseguaglianza di Markov)

$$P\{|X_n - X| > \epsilon\} \leq \epsilon^{-1} P[|X_n - X|] \to 0.$$

In secondo luogo, posto $Z_n = |X_n - X|$, si vede che $(Z_n)_{n\geq 1}$ è una successione di variabili aleatorie integrabili, per la quale è soddisfatta la condizione (b) della proposizione precedente: infatti, addirittura per ogni numero reale positivo t, si ha $\limsup_n \int_{\{Z_n > t\}} Z_n \, dP = 0$. Ne segue che la successione $(Z_n)_{n\geq 1}$ è uniformemente integrabile. Tale è dunque l'insieme costituito dalle Z_n e da X, e quindi anche, grazie alla maggiorazione $\frac{1}{2}|Z_n| \leq \frac{1}{2}|X_n| + \frac{1}{2}|X|$, la successione $(X_n)_{n\geq 1}$ (si veda (2.4)).

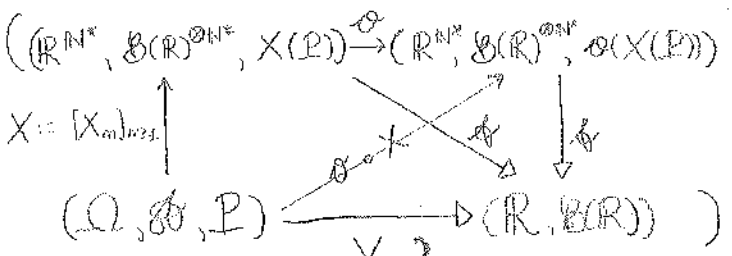
(b) $\Rightarrow$ (a): Si supponga soddisfatta la condizione (b). Allora, detta $(m_k)_{k\geq 1}$ una successione strettamente crescente d'interi, tale che la successione $(X_{m_k})_{k\geq 1}$ converga quasi certamente verso X, si ha innanzitutto, grazie al lemma di Fatou,

$$P[|X|] \leq \liminf_k P[|X_{m_k}|] < \infty$$

(dove la diseguaglianza finale discende dal fatto che la successione $(X_n)_{n\geq 1}$, essendo uniformemente integrabile, è limitata in $\mathcal{L}^1(P)$: si veda la condizione (b) di (2.1)). È così provata l'integrabilità di X. Rimane da provare la convergenza in $\mathcal{L}^1(P)$ di $(X_n)_{n\geq 1}$ verso X. Posto $Z_n = |X_n - X|$, ciò equivale a provare che, al tendere di n all'infinito, si ha $P[Z_n] \to 0$. A questo scopo, osserviamo che, essendo la successione $(X_n)_{n\geq 1}$ uniformemente integrabile, tale è l'insieme costituito dalle X_n e da X, e quindi anche, grazie alla maggiorazione $\frac{1}{2}Z_n \leq \frac{1}{2}|X_n| + \frac{1}{2}|X|$, la successione $(Z_n)_{n\geq 1}$ (si veda (2.4)). Fissato un numero reale ϵ maggiore di zero, esiste dunque un numero reale positivo t, tale che si abbia $\sup_n \int_{\{Z_n > t\}} Z_n \, dP \leq \epsilon/2$. Si ha allora, per ogni n,

$$P[Z_n] = \int_{\{Z_n \leq t\}} Z_n \, dP + \int_{\{Z_n > t\}} Z_n \, dP \leq P[Z_n \wedge t] + \epsilon/2. \tag{3.4}$$

Inoltre, applicando il Lemma (3.1) alla successione $(Z_n \wedge t)_{n\geq 1}$ (dominata dalla costante t), si vede che, al tendere di n all'infinito, si ha $P[Z_n \wedge t] \to 0$. Dalla relazione (3.4) discende dunque che, per n abbastanza grande, si ha $P[Z_n] \leq \epsilon$, e ciò basta per concludere. □



La legge dei grandi numeri per una successione stazionaria

Su uno spazio probabilizzato $(\Omega, \mathcal{A}, P)$ sia $(X_n)_{n\geq 1}$ una successione di variabili aleatorie reali. Si dice che essa è *stazionaria* se la sua legge congiunta è identica a quella della successione "traslata" $(X_{n+1})_{n\geq 1}$, cioè se la legge μ del blocco $[X_n]_{n\geq 1}$ è invariante rispetto all'applicazione θ che, ad ogni successione $(x_n)_{n\geq 1}$ di numeri reali, associa la successione "traslata" $(x_{n+1})_{n\geq 1}$. Quando ciò accade, μ è invariante anche rispetto ad ogni potenza di composizione di θ: in altri termini, il blocco $[X_n]_{n\geq 1}$ è isonomo anche ad ogni blocco della forma $[X_{n+k}]_{n\geq 1}$ (con k arbitrario intero positivo). In particolare, le variabili aleatorie X_n sono tra loro isonome.

Una variabile aleatoria numerica si dice *invariante* (rispetto alle X_n) se è della forma $f \circ [X_n]_{n\geq 1}$, dove f sia una funzione numerica misurabile (sullo spazio d'arrivo del blocco $[X_n]_{n\geq 1}$) che sia invariante rispetto alla composizione con θ, cioè che verifichi la relazione $f \circ \theta = f$. Un evento si dice poi invariante se tale è la sua funzione indicatrice. Le variabili aleatorie limitate e invarianti formano uno spazio di Riesz monotono contenente le costanti. Perciò gli eventi invarianti formano una tribù e, per una variabile aleatoria limitata (dunque anche per una variabile aleatoria numerica), essere misurabile rispetto a questa tribù equivale ad essere invariante.

Osservazione. Data la successione stazionaria $(X_n)_{n\geq 1}$, e presa una funzione reale g misurabile sullo spazio d'arrivo del blocco $[X_n]_{n\geq 1}$, si riconosce facilmente che, se una (e quindi anche l'altra) delle due variabili aleatorie $g \circ [X_n]_{n\geq 1}$, $g \circ [X_{n+1}]_{n\geq 1}$ è integrabile, le due variabili aleatorie hanno la stessa speranza condizionale rispetto alla tribù invariante. In particolare, se una (e quindi ciascuna) delle X_n è integrabile, allora le X_n hanno tutte la stessa speranza condizionale rispetto alla tribù invariante.

Teorema. *Sullo spazio probabilizzato $(\Omega, \mathcal{A}, P)$ sia $(X_n)_{n\geq 1}$ una successione stazionaria di variabili aleatorie reali integrabili. Si ponga*

$$S_n = X_1 + \cdots + X_n,$$

e si denoti con $\mathcal{J}$ la tribù degli eventi invarianti rispetto alle X_n. Allora la successione $(S_n/n)_{n\geq 1}$ converge quasi certamente e in $\mathcal{L}^1$ verso una versione della speranza condizionale di X_1 rispetto a $\mathcal{J}$.

Dimostrazione. Le variabili aleatorie X_k, essendo integrabili e isonome, sono uniformemente integrabili. Pertanto anche le S_n/n sono uniformemente integrabili. Perciò, detta V una versione della speranza condizionale $P[X_1 \mid \mathcal{J}]$, basterà provare la convergenza quasi certa di (S_n/n) verso V. Basterà anzi dimostrare che sussiste quasi dappertutto la diseguaglianza

$$\limsup_n (S_n/n) \leq V, \tag{1}$$

perché da questa si potrà poi dedurre una diseguaglianza analoga, relativa al limite inferiore, considerando la successione $(-X_n)$. D'altra parte, per provare la diseguaglianza (1), basterà dimostrare che, per ogni numero razionale x, l'evento

$$\{V < x < \limsup_n (S_n/n)\}$$

è trascurabile. Pur di considerare la successione $(X_n - x)$ (stazionaria, e ammettente ancora $\mathcal{J}$ come tribù degli eventi invarianti), ci si potrà ridurre al caso in cui x sia nullo. Si tratta allora di provare che la variabile aleatoria V è quasi dappertutto positiva sull'evento (invariante) $A = \{0 < \limsup_n (S_n/n)\}$. In altre parole, si tratta di provare che è quasi dappertutto positiva la variabile aleatoria VI_A. Si osservi che questa variabile aleatoria è una versione della speranza condizionale $P[X_1 I_A \,|\, \mathcal{J}]$. Inoltre, se si pone $T_n = S_1 \vee \cdots \vee S_n$, la relazione $A \subset \bigcup_n \{T_n > 0\}$ mostra che si ha convergenza puntuale di $\big(X_1 I_A I_{\{T_n>0\}}\big)_{n\geq 1}$ verso $X_1 I_A$. Dunque, grazie alla forma condizionale del teorema di Lebesgue sulla convergenza dominata, basterà dimostrare che, per ogni intero n maggiore di 1, si ha

$$P[X_1 I_{\{T_n>0\}} \,|\, \mathcal{J}] \geq 0 \tag{2}$$

(e quindi anche $P[X_1 I_A I_{\{T_n>0\}} \,|\, \mathcal{J}] \geq 0$). A questo scopo, denotiamo con Σ_n, Θ_n le variabili aleatorie definite come S_n, T_n, ma partendo dalla successione traslata $(X_{n+1})_{n\geq 1}$ anziché dalla successione $(X_n)_{n\geq 1}$. Si ha allora

$$\begin{aligned} T_{n+1} - X_1 &= (S_1 - X_1) \vee (S_2 - X_1) \vee \cdots \vee (S_{n+1} - X_1) \\ &= 0 \vee \Sigma_1 \vee \cdots \vee \Sigma_n = \Theta_n^+, \end{aligned}$$

e quindi

$$\begin{aligned} P\big[(T_{n+1} - X_1) I_{\{T_n>0\}} \,|\, \mathcal{J}\big] &\leq P\big[\Theta_n^+ \,|\, \mathcal{J}\big] \\ &= P\big[T_n^+ \,|\, \mathcal{J}\big] = P\big[T_n I_{\{T_n>0\}} \,|\, \mathcal{J}\big] \end{aligned}$$

(dove l'eguaglianza centrale discende dall'osservazione premessa all'enunciato del teorema). Ne segue

$$P\big[X_1 I_{\{T_n>0\}} \,|\, \mathcal{J}\big] \geq P\big[(T_{n+1} - T_n) I_{\{T_n>0\}} \,|\, \mathcal{J}\big] \geq 0.$$

È così provata la relazione (2), e ciò conclude la dimostrazione del teorema. □

È chiaro che ogni evento invariante è terminale. Perciò, se le X_n sono indipendenti, la tribù $\mathcal{J}$ è degenere (grazie alla Legge 0–1 di Kolmogorov). Il teorema sopra dimostrato contiene dunque come caso particolare il teorema seguente (*Legge forte dei grandi numeri di Kolmogorov*):

Teorema. *Su uno spazio probabilizzato $(\Omega, \mathcal{A}, P)$ sia $(X_n)_{n\geq 1}$ una successione di variabili aleatorie reali integrabili, indipendenti e isonome. Allora la successione $\big((X_1 + \cdots + X_n)/n\big)_{n\geq 1}$ converge quasi certamente e in $\mathcal{L}^1$ verso la costante $P[X_1]$.*

Confronto tra il teorema ergodico e la legge dei grandi numeri di Kolmogorov

Ci proponiamo di dimostrare, mediante un controesempio, che il teorema ergodico costituisce un'effettiva estensione della legge forte dei grandi numeri di Kolmogorov.

1. Richiami. Cominceremo col richiamare l'enunciato della legge forte dei grandi numeri di Kolmogorov.

(1.1) Legge forte di Kolmogorov. *Data, su uno spazio probabilizzato $(\Omega, \mathcal{A}, P)$, una successione indipendente $(X_n)_{n\geq 1}$ di variabili aleatorie reali isonome e integrabili, si ponga $S_n = X_1 + \cdots + X_n$.*

Allora la successione $(S_n/n)_{n\geq 1}$ converge quasi certamente e in media verso la comune speranza delle X_n.

Prima di richiamare l'enunciato del teorema ergodico, converrà innanzitutto richiamare i concetti di evento invariante, di evento terminale, di successione stazionaria e di successione ergodica.

Su uno spazio probabilizzato $(\Omega, \mathcal{A}, P)$ sia $(X_n)_{n\geq 1}$ una successione di variabili aleatorie a valori in un medesimo spazio misurabile $(E, \mathcal{E})$. Si denoti con X il blocco $[X_n]_{n\geq 1}$ e con μ la sua legge. Si denoti inoltre con θ l'operatore di *slittamento* nello spazio d'arrivo di X, ossia l'applicazione (misurabile) che ad ogni successione $(x_n)_{n\geq 1}$ di elementi di E associa la successione $(x_{n+1})_{n\geq 1}$. Si denoti infine con θ^k (per ogni intero positivo k) la k-esima iterata di θ, ossia l'applicazione che ad ogni successione $(x_n)_{n\geq 1}$ di elementi di E associa la successione $(x_{n+k})_{n\geq 1}$.

Ciò posto, la successione $(X_n)_{n\geq 1}$ si dice *stazionaria* se risulta $\mu = \theta(\mu)$, cioè se il blocco X è isonomo al blocco $\theta \circ X$ (identico a $[X_{n+1}]_{n\geq 1}$). Se ciò accade, allora le X_n sono tra loro isonome. Inversamente, sussiste la proposizione seguente, di dimostrazione immediata.

(1.2) Proposizione. *Se le X_n sono isonome e indipendenti, allora la successione $(X_n)_{n\geq 1}$ è stazionaria.*

Sia ora Y una variabile aleatoria su $(\Omega, \mathcal{A}, P)$, a valori in un arbitrario spazio misurabile $(F, \mathcal{F})$. Si dice che Y è *invariante* (rispetto alla fissata successione $(X_n)_{n\geq 1}$) se Y è della forma $f \circ X$, con f funzione misurabile sullo spazio d'arrivo di X (e a valori in $(F, \mathcal{F})$) verificante la relazione $f = f \circ \theta$. Si dice invece che Y è *terminale* (rispetto alla fissata successione $(X_n)_{n\geq 1}$) se, per ogni intero positivo k, è possibile mettere Y nella forma

$$Y = f_k \circ \theta^k \circ X = f_k \circ [X_{n+k}]_{n\geq 1},$$

con f_k funzione misurabile sullo spazio d'arrivo di X (e a valori in $(F, \mathcal{F})$).

Un evento si dice poi *invariante* (risp. *terminale*), se tale è la sua funzione indicatrice. È immediato verificare che la classe degli eventi invarianti (risp. terminali) è una tribù. Essa si chiama brevemente la *tribù invariante* (risp. la *tribù terminale*). Si verifica inoltre facilmente che le variabili aleatorie numeriche invarianti (risp. terminali) sono esattamente le variabili aleatorie numeriche misurabili rispetto alla tribù invariante (risp. terminale).

Sussiste la seguente proposizione (di facile dimostrazione).

(1.3) Proposizione. *La tribù invariante è contenuta nella tribù terminale.*

Sussiste inoltre il seguente classico risultato, dovuto a Kolmogorov.

(1.4) Legge 0–1 di Kolmogorov. *Se le X_n sono indipendenti, allora la tribù terminale è degenere (ossia ogni evento terminale ha probabilità eguale a 0 o a 1).*

Quando la tribù invariante sia degenere, si dice che la successione $(X_n)_{n\geq 1}$ è *ergodica*. Grazie a (1.3) e alla legge 0–1 di Kolmogorov, la Proposizione (1.2) può essere così completata:

(1.5) Proposizione. *Se le X_n sono indipendenti e isonome, allora la successione $(X_n)_{n\geq 1}$ è, oltre che stazionaria, anche ergodica.*

Ed ecco l'enunciato del teorema ergodico.

(1.6) Teorema ergodico. *Data, su uno spazio probabilizzato $(\Omega, \mathcal{A}, P)$, una successione stazionaria $(X_n)_{n\geq 1}$ di variabili aleatorie reali integrabili, si ponga, come al solito, $S_n = X_1 + \cdots + X_n$.*

Allora la successione $(S_n/n)_{n\geq 1}$ converge quasi certamente e in media. Se per giunta la successione $(X_n)_{n\geq 1}$ è ergodica, allora $(S_n/n)_{n\geq 1}$ converge (quasi certamente e in media) verso la comune speranza delle X_n.

2. Costruzione del controesempio.

Alla luce di (1.5), l'ultima affermazione del teorema ergodico appare come un'estensione della legge forte di Kolmogorov.

Come annunciato all'inizio, ci proponiamo di dimostrare che si tratta di un'effettiva estensione, ossia che le ipotesi della legge forte dei grandi numeri di Kolmogorov sono effettivamente più restrittive di quelle dell'ultima affermazione del teorema ergodico.

A questo scopo, basterebbe dimostrare che esiste una successione di variabili aleatorie reali integrabili, la quale sia stazionaria ed ergodica, ma non indipendente. In realtà, noi proveremo un risultato leggermente più ricco. Precisamente, noi costruiremo (seguendo un'idea di Pietro Rigo) una successione stazionaria $(X_n)_{n\geq 1}$ di variabili aleatorie a valori nell'insieme $\{-1, 1\}$ (e aventi come comune legge la

ripartizione uniforme su questo insieme), la quale non solo sia ergodica (cioè tale che la tribù invariante sia degenere), ma anche tale che la tribù terminale non sia degenere (e dunque tale che le X_n non siano indipendenti).

Cominciamo col ricordare un metodo ben noto, e molto semplice, per costruire una terna (U_1, U_2, U_3) di variabili aleatorie *a due a due* indipendenti, che non sia una terna di variabili aleatorie indipendenti. Per questo, dopo aver costruito (con lo schema delle prove indipendenti) una coppia (U_1, U_2) di variabili aleatorie indipendenti, a valori nell'insieme $\{-1, 1\}$ e aventi come comune legge la ripartizione uniforme su questo insieme, basta completare la definizione della terna (U_1, U_2, U_3) ponendo $U_3 = U_1 \cdot U_2$. Si vede allora che U_3 ha come legge quella comune alle due variabili aleatorie U_1, U_2 ed è indipendente da ciascuna di esse; tuttavia, essendo eguale al loro prodotto, non è indipendente dal loro blocco. (Inoltre si riconosce immediatamente che *ciascun* elemento della terna (U_1, U_2, U_3) è eguale al prodotto degli altri due.)

Usando ancora lo schema delle prove indipendenti, è ora possibile costruire, su un opportuno spazio probabilizzato $(\Omega, \mathcal{A}, P)$, una successione $(V_n)_{n\geq 1}$ di variabili aleatorie a valori in $\{-1, 1\}$, in modo tale che, per ogni intero positivo h, il blocco $[V_{3h+1}, V_{3h+2}, V_{3h+3}]$ abbia come legge la legge congiunta della terna (U_1, U_2, U_3) sopra costruita, e che inoltre i blocchi del tipo considerato formino una successione di variabili aleatorie indipendenti. Inoltre si può far in modo (sempre usando lo schema delle prove indipendenti) che, sul medesimo spazio $(\Omega, \mathcal{A}, P)$, esista anche una variabile aleatoria J, a valori nell'insieme $\{0, 1, 2\}$ e avente come legge la ripartizione uniforme su questo insieme, la quale sia indipendente dal blocco $[V_n]_{n\geq 1}$. Infine, pur di modificare le V_n su un evento trascurabile, si può far in modo che, per ogni intero positivo h, la variabile aleatoria V_{3h+3} coincida dappertutto col prodotto $V_{3h+1} \cdot V_{3h+2}$.

Si consideri allora la successione $(X_n)_{n\geq 1}$ di variabili aleatorie a valori in $\{-1, 1\}$ così definita: per ciascun intero n strettamente positivo, e ciascun intero j appartenente a $\{0, 1, 2\}$, la variabile aleatoria X_n coincide, sull'evento $\{J = j\}$, con V_{n+j}.

Verifichiamo che la successione $(X_n)_{n\geq 1}$ così costruita possiede le proprietà desiderate. Cominciamo col provare che essa è stazionaria. A questo scopo, poniamo

$$V = [V_n]_{n\geq 1}, \qquad X = [X_n]_{n\geq 1},$$

e denotiamo con θ l'operatore di slittamento nello spazio d'arrivo di questi due blocchi. La legge di X è allora eguale alla media aritmetica delle leggi dei tre blocchi V, $\theta \circ V$, $\theta^2 \circ V$. Dunque, posto $\lambda = V(P)$, $\mu = X(P)$, si ha

$$\mu = \tfrac{1}{3}\big(\lambda + \theta(\lambda) + \theta^2(\lambda)\big), \tag{2.1}$$

e quindi

$$\theta(\mu) = \tfrac{1}{3}\big(\theta(\lambda) + \theta^2(\lambda) + \theta^3(\lambda)\big). \tag{2.2}$$

D'altra parte, la definizione della successione $(V_n)_{n\geq 1}$ mostra che il blocco $\theta^3 \circ V$ è isonomo a V, ossia che la legge $\theta^3(\lambda)$ è identica a λ. Si vede dunque che il secondo

membro di (2.2) coincide col secondo membro di (2.1), sicché risulta $\theta(\mu) = \mu$. È così provato che la successione $(X_n)_{n\geq 1}$ è stazionaria.

Poiché le X_n sono, al pari delle V_n, a valori in $\{-1, 1\}$, provare che esse hanno come comune legge la ripartizione uniforme su $\{-1, 1\}$ e sono a due a due indipendenti equivale a provare che esse sono centrate e a due a due non correlate: e per questo basta impiegare la relazione (2.1) e il fatto che le V_n sono a loro volta centrate e a due a due non correlate.

Proviamo ora che la successione $(X_n)_{n\geq 1}$ è ergodica, ossia che la tribù invariante ad essa associata è degenere. Assegnata dunque una variabile aleatoria numerica della forma $f \circ X$, con f funzione numerica misurabile (sullo spazio d'arrivo di X) verificante la relazione $f = f \circ \theta$, proviamo che $f \circ X$ è degenere. A questo scopo, cominciamo con l'osservare che, per ogni intero positivo k, si ha $f = f \circ \theta^k$, e quindi $f \circ X = f \circ \theta^k \circ X$. In particolare, ciò implica che, per ogni intero h strettamente positivo e ogni intero j appartenente a $\{0, 1, 2\}$, si ha

$$\begin{aligned} I_{\{J=j\}}\, f \circ X &= I_{\{J=j\}}\, f \circ \theta^{3h-j} \circ X \\ &= I_{\{J=j\}}\, f \circ \theta^{3h-j} \circ \theta^j \circ V \\ &= I_{\{J=j\}}\, f \circ \theta^{3h} \circ V. \end{aligned}$$

Ne segue

$$f \circ X = f \circ \theta^{3h} \circ V.$$

Il fatto che questa questa relazione sussista per ogni h mostra che la variabile aleatoria $f \circ X$ è terminale rispetto alla successione (indipendente)

$$V_1, V_2, V_4, V_5, V_7, V_8, \ldots,$$

e quindi è degenere (grazie alle legge 0–1 di Kolmogorov). È così provato che la tribù invariante associata alla successione $(X_n)_{n\geq 1}$ è degenere.

Proviamo infine che la tribù terminale associata alla stessa successione non è degenere. Per questo, basterà esibire una variabile aleatoria numerica che sia terminale rispetto alla successione $(X_n)_{n\geq 1}$, ma non degenere. Cominciamo con l'osservare che, per ogni intero positivo h, il prodotto $V_{3h+1}V_{3h+2}V_{3h+3}$ coincide con la costante 1 (grazie al fatto che l'utimo fattore coincide col prodotto dei primi due), mentre il prodotto $V_{3h+2}V_{3h+3}V_{3h+4}$ è centrato (grazie al fatto che l'ultimo fattore è centrato e indipendente dal prodotto dei primi due). Poniamo

$$Y = \liminf_h X_{3h+1}X_{3h+2}X_{3h+3}.$$

La variabile aleatoria Y così definita è terminale rispetto alla successione $(X_n)_{n\geq 1}$. Basterà provare che essa non è degenere. Osserviamo che, sull'evento $\{J = 0\}$ (non trascurabile), Y coincide con la costante 1. Basterà dunque provare che Y non è

equivalente alla costante 1. Ciò si ottiene osservando che, grazie al lemma di Fatou, si ha

$$\begin{aligned}\int_{\{J=1\}} Y\,\mathrm{d}P &\le \liminf_h \int_{\{J=1\}} X_{3h+1}X_{3h+2}X_{3h+3}\,\mathrm{d}P\\ &= \liminf_h \int_{\{J=1\}} V_{3h+2}V_{3h+3}V_{3h+4}\,\mathrm{d}P\\ &= P\{J=1\}\,\liminf_h \int V_{3h+2}V_{3h+3}V_{3h+4}\,\mathrm{d}P = 0\end{aligned}$$

(dove la penultima eguaglianza è dovuta all'indipendenza di J dal blocco delle V_n).

È così provato che la tribù terminale associata alla successione $(X_n)_{n\ge1}$ non è degenere.

3. Un'osservazione sulla tribù invariante. Nell'enunciare i risultati richiamati nel primo paragrafo, alcuni autori intendono la nozione di variabile aleatoria invariante in un senso un po' più largo di quello da noi impiegato. Questo senso più largo può essere così definito (mantenendo le stesse notazioni del paragrafo 1).

Sia Y una variabile aleatoria su $(\Omega, \mathcal{A}, P)$, a valori in un arbitrario spazio misurabile $(F, \mathcal{F})$. Si dice che Y è *P–invariante* (rispetto alla fissata successione $(X_n)_{n\ge1}$) se Y è equivalente (secondo P) a una variabile aleatoria della forma $f \circ X$, con f funzione misurabile sullo spazio d'arrivo di X (e a valori in $(F, \mathcal{F})$) verificante la relazione

$$f \circ X \sim f \circ \theta \circ X \pmod P,$$

ossia $f \sim f \circ \theta \pmod \mu$.

Si vede però facilmente che, se ci si limita alle variabili aleatorie *numeriche* (e si suppone che l'assegnata successione $(X_n)_{n\ge1}$ sia *stazionaria*), la nozione di variabile aleatoria P–invariante può essere immediatamente ricondotta a quella, più semplice, di variabile aleatoria invariante. Sussiste infatti la proposizione seguente.

(3.1) Proposizione. *Sullo spazio probabilizzato $(\Omega, \mathcal{A}, P)$ sia Y una variabile aleatoria numerica. Le condizioni che seguono sono equivalenti:*

(a) Y è P–invariante rispetto all'assegnata successione stazionaria $(X_n)_{n\ge1}$.

(b) Y è equivalente, secondo P, a una variabile aleatoria numerica invariante rispetto all'assegnata successione stazionaria $(X_n)_{n\ge1}$.

Dimostrazione. Basta provare l'implicazione (a) $\Rightarrow$ (b) (l'implicazione inversa essendo immediata). Supponiamo dunque che Y sia P–invariante. Ciò significa che Y è equivalente, secondo P, a una variabile aleatoria numerica della forma $f \circ X$, con f funzione numerica misurabile (sullo spazio d'arrivo di X) verificante la relazione

$$f \sim f \circ \theta \pmod \mu. \tag{3.2}$$

Cominciamo con l'osservare che, per ogni intero positivo k, grazie all'eguaglianza $\theta^k(\mu) = \mu$, la relazione (3.2) implica

$$f \circ \theta^k \sim f \circ \theta^{k+1} \pmod \mu.$$

Ne segue che, per ogni k, si ha

$$f \sim f \circ \theta^k \pmod{\mu}.$$

Se dunque si pone $g = \liminf_k f \circ \theta^k$, si ha $f \sim g \pmod{\mu}$, ossia $Y \sim g \circ X \pmod{P}$. Poiché la funzione g verifica la relazione $g = g \circ \theta$, si vede che la condizione (b) è soddisfatta. □

Generalità sulle martingale

1. Processi e filtrazioni

(1.1) Definizione. Siano assegnati uno spazio probabilizzato $(\Omega, \mathcal{A}, P)$, uno spazio misurabile $(E, \mathcal{E})$, un insieme ordinato I. Sia poi X un'applicazione di $I \times \Omega$ in E:

$$X : I \times \Omega \to E.$$

Si dice che X è un *processo (stocastico)* su $(\Omega, \mathcal{A}, P)$, ammettente $(E, \mathcal{E})$ come *spazio degli stati* e I come *insieme dei tempi*, se, per ogni elemento t di I, l'applicazione parziale $\omega \mapsto X(t, \omega)$ di Ω in E, denotata anche con $X(t, \cdot\,)$ o con X_t, è un'applicazione misurabile di $(\Omega, \mathcal{A})$ in $(E, \mathcal{E})$.

In tal caso, per ogni elemento ω di Ω, la famiglia $\big(X_t(\omega)\big)_{t \in I}$, ossia l'applicazione $t \mapsto X(t, \omega)$ di I in E, denotata anche con $X(\cdot\,, \omega)$, si chiama la *traiettoria* del processo X associata a ω.

Commettendo un innocuo abuso, si confonde spesso un processo X con la famiglia di variabili aleatorie $(X_t)_{t \in I}$ oppure con il blocco $[X_t]_{t \in I}$ di queste variabili aleatorie, ossia con la variabile aleatoria su $(\Omega, \mathcal{A}, P)$, a valori nello spazio prodotto $(E^I, \mathcal{E}^{\otimes I})$, che a ciascun elemento ω di Ω associa la traiettoria $X(\cdot\,, \omega)$.

Intuitivamente, un processo X si può interpretare come un modello matematico per descrivere l'evoluzione aleatoria, nel tempo, di un determinato sistema fisico, il cui "stato" sia rappresentato, in ciascun istante, da un punto dello spazio degli stati. Per ogni elemento t di I, la variabile aleatoria X_t rappresenta dunque lo stato aleatorio del sistema all'istante t. Per ogni eventualità ω, la traiettoria $X(\cdot\,, \omega)$ rappresenta una delle possibili evoluzioni temporali dello stato del sistema: precisamente, quella che corrisponde all'eventualità ω. Per questa ragione, gli elementi dell'insieme I si chiamano anche *istanti*.

(1.2) Definizione. Nelle ipotesi di (1.1), una *filtrazione* su $(\Omega, \mathcal{A}, P)$, ammettente I come insieme dei tempi, è una famiglia $\mathcal{F} = (\mathcal{F}_t)_{t \in I}$ di sottotribù di $\mathcal{A}$, ammettente I come insieme degl'indici e tale che si abbia $\mathcal{F}_s \subset \mathcal{F}_t$ per ogni coppia s, t di indici con $s \leq t$ (dove $\leq$ designa la relazione d'ordine di I). Gli elementi della tribù $\mathcal{F}_t$ sono detti gli eventi *anteriori all'istante* t (relativamente alla filtrazione $\mathcal{F}$).

Date due filtrazioni $\mathcal{F}$, $\mathcal{G}$ ammettenti lo stesso insieme dei tempi, si dice che $\mathcal{F}$ è *meno fine* di $\mathcal{G}$ (o che $\mathcal{G}$ è *più fine* di $\mathcal{F}$) se si ha $\mathcal{F}_t \subset \mathcal{G}_t$ per ogni t.

(1.3) Definizione. Data una filtrazione $\mathcal{F} = (\mathcal{F}_t)_{t \in I}$, un processo X, ammettente come insieme dei tempi una parte J di I, si dice *adattato* a $\mathcal{F}$ se, per ogni elemento t di J, la variabile aleatoria X_t è misurabile rispetto a $\mathcal{F}_t$.

(1.4) Definizione. Dato un processo X avente I come insieme dei tempi, si chiama *filtrazione naturale* di X la filtrazione $\mathcal{F}$, ammettente I come insieme dei tempi, così definita: per ogni istante t, la tribù $\mathcal{F}_t$ è la tribù generata dalle variabili aleatorie X_s con $s \in I$, $s \leq t$, ossia è la minima tribù che renda misurabile il blocco $[X_s]_{s \in I,\, s \leq t}$.

Evidentemente, affinché un processo X sia adattato a una filtrazione $\mathcal{G}$, occorre e basta che questa sia più fine della filtrazione naturale di X.

(1.5) Esempio. Si supponga che $\mathcal{G}$ sia la filtrazione naturale di un processo $(Z_t)_{t\in I}$ (ammettente come spazio degli stati un arbitrario spazio misurabile) e che X sia un processo *reale*, cioè ammettente come spazio degli stati $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Allora, affinché X sia adattato a $\mathcal{G}$, occorre e basta che, per ogni istante t, la variabile aleatoria reale X_t sia misurabile rispetto alla tribù $\mathcal{G}_t$, ossia rispetto alla minima tribù che renda misurabile il blocco $[Z_s]_{s\in I,\, s\le t}$. Grazie al lemma di misurabilità di Doob, questa condizione equivale ad imporre che, per ogni t, la variabile aleatoria X_t sia della forma

$$X_t = f_t \circ [Z_s]_{s\in I,\, s\le t},$$

con f_t opportuna funzione reale, misurabile sullo spazio d'arrivo di $[Z_s]_{s\in I,\, s\le t}$.

2. Definizione generale di martingala

(2.1) Definizione. Siano dati uno spazio probabilizzato $(\Omega, \mathcal{A}, P)$ e, su di esso, una filtrazione $\mathcal{F} = (\mathcal{F}_t)_{t\in I}$ ammettente come insieme dei tempi l'insieme ordinato I. Un processo reale X, ammettente I come insieme dei tempi, si dice una *martingala* rispetto alla filtrazione $\mathcal{F}$ se possiede le proprietà seguenti:

(a) (*Proprietà di adattazione*) X è adattato a $\mathcal{F}$.

(b) (*Proprietà d'integrabilità*) Ciascuna delle variabili aleatorie X_t è integrabile.

(c) (*Proprietà di ortogonalità degli incrementi*) Per ogni coppia (s,t) d'istanti, con $s \le t$, l'incremento $X_t - X_s$ è ortogonale alla tribù $\mathcal{F}_s$, cioè verifica la relazione

$$\int_A (X_t - X_s)\,\mathrm{d}P = 0 \quad \text{per ogni elemento } A \text{ di } \mathcal{F}_s.$$

Si dice invece che il processo X è (rispetto a $\mathcal{F}$) una *sottomartingala* se esso possiede le proprietà (a), (b) e la proprietà seguente:

(c′) Per ogni coppia (s,t) d'istanti, con $s \le t$, l'incremento $X_t - X_s$ verifica la relazione

$$\int_A (X_t - X_s)\,\mathrm{d}P \ge 0 \quad \text{per ogni elemento } A \text{ di } \mathcal{F}_s.$$

Infine si dice che il processo X è (rispetto a $\mathcal{F}$) una *supermartingala* se il processo $-X$ è una sottomartingala. (Si vede dunque che una martingala è un processo che sia, nello stesso tempo, una sottomartingala e una supermartingala.)

(2.2) Osservazione. Impiegando il concetto di speranza condizionale, si può dire che, nelle ipotesi della definizione precedente, se il processo X verifica le due condizioni (a), (b), allora, affinché esso sia, rispetto a $\mathcal{F}$, una sottomartingala (risp. una supermartingala), occorre e basta che, per ogni coppia (s,t) d'istanti, con $s \le t$, si abbia $P[X_t - X_s \,|\, \mathcal{F}_s] \ge 0$ (risp. $P[X_t - X_s \,|\, \mathcal{F}_s] \le 0$).

Di conseguenza, se il processo X verifica le due condizioni (a), (b), allora, affinché esso sia una martingala, occorre e basta che, per ogni coppia (s,t) d'istanti, con $s \le t$, si abbia $P[X_t - X_s \,|\, \mathcal{F}_s] = 0$.

(2.3) Osservazione. Se un processo X è una martingala rispetto a una filtrazione $\mathcal{F}$, allora esso è una martingala anche rispetto ad ogni filtrazione che sia meno fine di $\mathcal{F}$, ma più fine della filtrazione naturale di X (in modo che X sia ad essa adattato): in particolare, X è una martingala anche rispetto alla propria filtrazione naturale. Un'osservazione analoga vale per i concetti di sottomartingala e di supermartingala.

(2.4) Convenzione. Ogni volta che, nel seguito, si dirà che un processo X è una martingala senza riferirsi esplicitamente a una particolare filtrazione, sarà sempre sottinteso che ci si riferisca alla filtrazione naturale di X. Un'analoga convenzione si adotterà per i concetti di sottomartingala e di supermartingala.

3. Martingale con tempi discreti

Di qui in avanti, tutte le filtrazioni considerate avranno come insieme dei tempi l'insieme $\mathbb{N}$ degli interi positivi. Così pure, tutti i processi considerati saranno (salvo esplicito avviso contrario) processi *reali* aventi $\mathbb{N}$ come insieme dei tempi. Per processi di questo tipo, la definizione di martingala si può leggermente semplificare[*]. Precisamente, dato un processo X, per verificare che esso possieda la proprietà di ortogonalità degli incrementi, basta verificare che questa proprietà valga per gli incrementi della forma $X_n - X_{n-1}$, con $n \ge 1$: infatti l'incremento relativo a un'arbitraria coppia (h,k) d'istanti, con $h \le k$, è somma di un numero finito d'incrementi del tipo particolare sopra descritto. (Una semplificazione analoga vale naturalmente a proposito dei concetti di sottomartingala e di supermartingala.)

(3.1) Osservazione. Se X è una sottomartingala (risp. una martingala), la successione $(P[X_n])_{n\ge 0}$ è crescente (risp. costante).

Ricordando che, se una variabile aleatoria centrata è indipendente da una fissata tribù, le è ortogonale, è facile costruire un paio di esempi elementari di martingala(

(3.2) Esempio. Assegnata, su uno spazio probabilizzato $(\Omega, \mathcal{A}, P)$, una successione $(X_n)_{n\ge 1}$ di variabili aleatorie reali, indipendenti e centrate, si ponga, per ogni intero positivo n,

$$S_n = X_1 + \cdots + X_n \quad \text{(in particolare, } S_0 = 0\text{)},$$

$$\mathcal{F}_n = \mathcal{T}(X_1, \ldots, X_n) \quad \text{(in particolare, } \mathcal{F}_0 = \{\Omega, \emptyset\}\text{)}. \tag{3.3}$$

Allora il processo $(S_n)_{n\ge 0}$ è una martingala rispetto alla filtrazione $(\mathcal{F}_n)_{n\ge 0}$. Infatti esso possiede in modo evidente la proprietà di adattazione e quella d'integrabilità, e inoltre, per ogni intero n strettamente positivo, l'incremento $S_n - S_{n-1}$, essendo identico a X_n, è centrato e indipendente da $\mathcal{F}_{n-1}$ (dunque ortogonale a $\mathcal{F}_{n-1}$).

Si osservi che, essendo $X_n = S_n - S_{n-1}$ per $n \ge 1$, la filtrazione $(\mathcal{F}_n)_{n\ge 0}$ non è altro che la filtrazione naturale del processo $(S_n)_{n\ge 0}$.

(3.4) Esempio. Assegnata, su uno spazio probabilizzato $(\Omega, \mathcal{A}, P)$ una successione $(X_n)_{n\geq 1}$ di variabili aleatorie reali, indipendenti, integrabili e di speranza unitaria, si ponga, per ogni intero positivo n,

$$Y_n = \prod_{j=1}^{n} X_j \qquad \text{(in particolare, } Y_0 = 1\text{)}.$$

Allora il processo $(Y_n)_{n\geq 0}$ è una martingala rispetto alla filtrazione $(\mathcal{F}_n)_{n\geq 0}$ definita da (3.3). Infatti esso possiede in modo ovvio la proprietà di adattazione e quella di d'integrabilità, e inoltre, per ogni intero n strettamente positivo, si ha

$$\begin{aligned} P\big[\, Y_n - Y_{n-1} \,|\, \mathcal{F}_{n-1} \,\big] &= P\big[\, Y_{n-1}X_n - Y_{n-1} \,|\, \mathcal{F}_{n-1} \,\big] \\ &= P\big[Y_{n-1}(X_n - 1) \,|\, \mathcal{F}_{n-1} \,\big] = 0 \end{aligned}$$

(dove l'eguaglianza finale è dovuta al fatto che la variabile aleatoria $X_n - 1$, essendo centrata e indipendente da $\mathcal{F}_{n-1}$, è ortogonale a $\mathcal{F}_{n-1}$).

Si osservi che l'incremento $Y_n - Y_{n-1}$, pur essendo ortogonale alla tribù $\mathcal{F}_{n-1}$, non è (in generale) indipendente da essa (al contrario di ciò che avviene per l'incremento $S_n - S_{n-1}$ nell'Esempio (3.2)).

4. La trasformazione di Burkholder

(4.1) Definizione. Su uno spazio probabilizzato, munito di una filtrazione $\mathcal{F}$, un processo V, *avente* $\mathbb{N}^*$ *come insieme dei tempi*, si dice *prevedibile* (rispetto a $\mathcal{F}$) se, per ogni intero n strettamente positivo, la variabile aleatoria V_n è misurabile rispetto alla tribù $\mathcal{F}_{n-1}$. (Un tal processo è evidentemente adattato alla filtrazione $\mathcal{F}$.)

(4.2) Definizione. Su uno spazio probabilizzato $(\Omega, \mathcal{A}, P)$, munito di una filtrazione $\mathcal{F}$, siano dati un processo X, adattato a $\mathcal{F}$, e un processo V, prevedibile rispetto a $\mathcal{F}$. Sia poi Y il processo caratterizzato dalle relazioni

$$Y_0 = X_0, \qquad Y_n - Y_{n-1} = V_n(X_n - X_{n-1}) \quad \text{per} \quad n \geq 1, \tag{4.3}$$

ossia definito dalla relazione esplicita

$$Y_n = X_0 + \textstyle\sum_{1\leq j\leq n} V_j(X_j - X_{j-1}) \quad \text{per} \quad n \geq 0. \tag{4.4}$$

Il processo Y si dice allora il *trasformato, secondo Burkholder,* del processo X mediante il processo prevedibile V. Esso si denota talvolta col simbolo $X \bullet V$.

(4.5) Teorema. (Burkholder) *Nelle ipotesi della definizione precedente, si supponga per giunta che ciascuna delle variabili aleatorie Y_n sia integrabile. Valgono allora le due affermazioni seguenti:*

(a) Se il processo X è una martingala rispetto a $\mathcal{F}$, tale è il processo Y.

(b) Se il processo X è una sottomartingala rispetto a $\mathcal{F}$, e se il processo V è positivo, allora anche il processo Y è una sottomartingala rispetto a $\mathcal{F}$.

Dimostrazione. Basta osservare che, per ogni intero n strettamente positivo, grazie alla seconda delle relazioni (4.3) e al fatto che V_n è misurabile rispetto a $\mathcal{F}_{n-1}$, una versione della speranza condizionale $P\big[Y_n - Y_{n-1} \mid \mathcal{F}_{n-1}\big]$ si ottiene moltiplicando per V_n una versione di $P\big[X_n - X_{n-1} \mid \mathcal{F}_{n-1}\big]$. □

(4.6) Osservazione. Nell'enunciato di (4.5), l'ipotesi che le Y_n siano integrabili può essere omessa (in quanto automaticamente verificata) quando le V_n siano limitate: ciò risulta immediatamente dalla relazione (4.4).

5. Tempi d'arresto e teorema d'arresto

(5.1) Notazione. Dati, su uno spazio probabilizzato, una variabile aleatoria S a valori in $\mathbb{N} \cup \{\infty\}$ e un processo X, si denota con X_S la variabile aleatoria che è nulla su $\{S = \infty\}$ e che, per ogni intero positivo n, coincide con X_n sull'evento $\{S = n\}$.

(5.2) Definizione. Su uno spazio probabilizzato, munito di una filtrazione $\mathcal{F}$, sia T una variabile aleatoria a valori in $\mathbb{N} \cup \{\infty\}$. Si denoti con V il processo avente $\mathbb{N}^*$ come insieme dei tempi, definito da

$$V_n = I_{\{n \le T\}} \quad \text{per} \quad n \ge 1. \tag{5.3}$$

Si dice che T è un *tempo d'arresto* (relativo alla filtrazione $\mathcal{F}$) se il processo V sopra definito è prevedibile rispetto a $\mathcal{F}$, ossia se, per ogni intero n strettamente positivo, l'evento $\{n \le T\}$ appartiene alla tribù $\mathcal{F}_{n-1}$. Se ciò accade, il processo V si chiama il *processo d'arresto* associato al tempo d'arresto T.

(5.4) Osservazione. Nelle ipotesi della definizione precedente, la condizione affinché T sia un tempo d'arresto si può formulare, in modo equivalente, imponendo che, per ogni intero n strettamente positivo, *il complementare* dell'evento $\{n \le T\}$, ossia l'evento $\{T \le n-1\}$, appartenga a $\mathcal{F}_{n-1}$. In modo ancora equivalente, la stessa condizione si può formulare imponendo che, per ogni intero positivo n, si abbia $\{T \le n\} \in \mathcal{F}_n$.

Nelle ipotesi della Definizione (4.2), è interessante analizzare la forma particolare alla quale si riduce il processo Y (trasformato, secondo Burkholder, del processo X mediante il processo prevedibile V) nel caso speciale in cui V sia il processo d'arresto associato a un tempo d'arresto T. In questo caso, la relazione (4.4) che definisce Y_n si riduce alla forma

$$\begin{aligned} Y_n &= X_0 + \textstyle\sum_{1 \le j \le n} I_{\{j \le T\}} (X_j - X_{j-1}) \overset{(5a)}{=} X_n + \sum_{j=1}^{n \wedge T} (X_0 - X_{j-1}) = \\ &= X_0 + \textstyle\sum_{j \ge 1} I_{\{j \le n \wedge T\}} (X_j - X_{j-1}) \overset{(5)}{=} X_{n \wedge T} \end{aligned} \tag{5.5}$$

(dove l'ultimo membro ha il senso convenuto in (5.1)).

(5.6) Definizione. Nelle ipotesi di (4.2), si supponga in particolare che il processo prevedibile V sia il processo d'arresto associato a un tempo d'arresto T (ossia il processo definito da (5.3)). Allora il trasformato di X mediante V secondo Burkholder, ossia il processo Y definito da (5.5), si chiama il processo ottenuto *arrestando* X al tempo d'arresto T, e si denota con $X^{|T}$. Si ha dunque, per definizione,

$$\left(X^{|T}\right)_n = X_{n\wedge T} \quad \text{per} \quad n \geq 0. \tag{5.7}$$

Si osservi che, per ogni elemento ω di Ω, la traiettoria di $X^{|T}$ associata a ω coincide con $X(\,\cdot\,,\omega)$ su $\mathbb{N}\cap[0,T(\omega)]$ ed è costante su $\mathbb{N}\cap[T(\omega),\infty]$.

Usando la locuzione introdotta nella definizione precedente, si può enunciare la proposizione seguente (conseguenza immediata di (4.5) e (4.6)).

(5.8) Proposizione. *Sia T un tempo d'arresto, relativo a una filtrazione $\mathcal{F}$. Allora, se X è una martingala* (risp. *una sottomartingala*) *rispetto a $\mathcal{F}$, tale è il processo $X^{|T}$ ottenuto arrestando X a T (ossia definito da (5.7)).*

Un'ulteriore utile conseguenza del Teorema (4.5) di Burkholder è il teorema seguente, noto come *Teorema d'arresto*.

(5.9) Teorema. *Sia X una sottomartingala rispetto a una filtrazione $\mathcal{F}$. Siano inoltre S, T due tempi d'arresto limitati, relativi alla filtrazione $\mathcal{F}$, con $S \leq T$. Allora la variabile aleatoria $X_T - X_S$ (incremento di X nel passare dall'istante aleatorio S all'istante aleatorio T) è integrabile, e si ha*

$$P\big[X_T - X_S\big] \geq 0. \tag{5.10}$$

Questa diseguaglianza si riduce a un'eguaglianza nel caso particolare in cui X sia una martingala.

Dimostrazione. Pur di sostituire il processo X con il processo $(X_n - X_0)_{n\geq 0}$ (che è ancora una sottomartingala rispetto a $\mathcal{F}$, e il cui incremento tra gli istanti aleatori S, T è identico a quello di X), si può supporre $X_0 = 0$. Consideriamo allora il processo V (avente $\mathbb{N}^*$ come insieme dei tempi) definito da

$$V_n = I_{\{S<n\leq T\}} = I_{\{n\leq T\}} - I_{\{n\leq S\}} \quad \text{per} \quad n \geq 1.$$

Si tratta di un processo prevedibile, limitato e positivo. Detto Y il trasformato, secondo Burkholder, di X mediante V, si ha $Y = X^{|T} - X^{|S}$. Perciò, se m è un intero che maggiori T, si può scrivere

$$Y_m = X_{m\wedge T} - X_{m\wedge S} = X_T - X_S.$$

Grazie al Teorema (4.5), Y è ancora una sottomartingala. Dunque Y_m è integrabile e inoltre, essendo $Y_0 = 0$, l'Osservazione (3.1) mostra che si ha $P\big[Y_m\big] \geq 0$. È chiaro infine che questa diseguaglianza si riduce a un'eguaglianza nel caso particolare in cui X (e quindi Y) sia una martingala. □

(5.10) Osservazione. Nel teorema precedente, X_T è integrabile (come si vede applicando il teorema con $S = 0$). Di conseguenza, anche X_S è integrabile.

La diseguaglianza di Doob sul numero delle discese

Su uno spazio probabilizzato $(\Omega, \mathcal{A}, P)$ sia X un processo reale, ammettente $\mathbb{N}$ come insieme dei tempi. Fissata una coppia (a, b) di numeri reali, con $a < b$, si ponga

$$S_1(\omega) = \inf\{j \in \mathbb{N} : \ X_j(\omega) \geq b\},$$
$$T_1(\omega) = \inf\{j \in \mathbb{N} : \ X_j(\omega) \leq a, \ j > S_1(\omega)\},$$
$$S_2(\omega) = \inf\{j \in \mathbb{N} : \ X_j(\omega) \geq b, \ j > T_1(\omega)\},$$
$$T_2(\omega) = \inf\{j \in \mathbb{N} : \ X_j(\omega) \leq a, \ j > S_2(\omega)\},$$

e così via. Si ponga inoltre

$$D = \sum_{k\geq 1} I_{\{T_k<\infty\}} \tag{1}$$

e, per ogni intero positivo n,

$$D_n = \sum_{k\geq 1} I_{\{T_k\leq n\}}. \tag{2}$$

La variabile aleatoria D_n si può chiamare "numero degli attraversamenti in discesa dell'intervallo $[a, b]$ da parte del processo X nell'intervallo temporale $[0, n]$". La variabile aleatoria D, che è eguale all'inviluppo superiore della successione crescente $(D_n)_{n\geq 0}$, si può invece chiamare "numero totale degli attraversamenti in discesa dell'intervallo $[a, b]$ da parte del processo X".

Sussiste il fondamentale teorema seguente, dovuto a J. L. Doob.

Teorema. *Nelle ipotesi e con le notazioni precedenti, si supponga che il processo X sia una sottomartingala. Si ha allora, per ogni n,*

$$(b-a)P[D_n] \leq P\left[(X_n - b)^+\right].$$

Dimostrazione. Fissato l'intero positivo n, si tratta di dimostrare la diseguaglianza

$$(b-a)\sum_{k\geq 1} P\{T_k \leq n\} \leq P\left[(X_n - b)^+\right].$$

A questo scopo, poniamo, per ogni intero k strettamente positivo,

$$\Delta_k = X_{T_k\wedge n} - X_{S_k\wedge n}.$$

Le variabili aleatorie S_k, T_k sono tempi d'arresto relativi alla filtrazione naturale del processo X. Pertanto, grazie al teorema d'arresto, la variabile aleatoria Δ_k ha

speranza positiva. D'altra parte, essa è nulla sull'insieme $\{n \le S_k\}$, mentre sull'insieme $\{S_k < n < T_k\}$ è maggiorata da $X_n - b$ e sull'insieme $\{T_k \le n\}$ è maggiorata da $-(b-a)$. Si ha dunque

$$0 \le P[\Delta_k] = \int_{\{S_k<n<T_k\}} \Delta_k \, dP + \int_{\{T_k \le n\}} \Delta_k \, dP,$$
$$\le \int_{\{S_k<n<T_k\}} (X_n - b) \, dP - (b-a) P\{T_k \le n\}.$$

Ne segue

$$(b-a) P\{T_k \le n\} \le \int_{\{S_k<n<T_k\}} (X_n - b) \, dP.$$

Di qui, ponendo $A_k = \{S_k < n < T_k\}$, $A = \bigcup_{k\ge 1} A_k$, e osservando che gli insiemi A_k sono a due a due disgiunti, si ricava

$$(b-a) \textstyle\sum_{k\ge 1} P\{T_k \le n\} \le \int_A (X_n - b) \, dP \le P[(X_n - b)^+].$$

Il teorema è così dimostrato. □

Corollario. *Nelle stesse ipotesi del teorema precedente, si supponga che la sottomartingala X sia limitata in $\mathcal{L}^1$, nel senso che il numero $c = \sup_n P[|X_n|]$ sia finito. Allora la successione $(X_n)_{n\ge 0}$ converge quasi certamente verso una variabile aleatoria reale integrabile.*

Dimostrazione. Cominciamo col provare che le due variabili aleatorie numeriche

$$\liminf_n X_n, \qquad \limsup_n X_n \tag{3}$$

sono equivalenti secondo P, ossia che l'evento $\{\liminf_n X_n < \limsup_n X_n\}$ è trascurabile. Questo evento è l'unione degli eventi della forma

$$\{\liminf_n X_n < a < b < \limsup_n X_n\}, \tag{4}$$

dove (a, b) sia una coppia di numeri razionali, con $a < b$. Basterà dunque provare che ciascuno di questi è trascurabile. Ciò si ottiene osservando che l'evento (4) è contenuto nell'evento $\{D = \infty\}$ (dove D è la variabile aleatoria definita da (1)) e che, grazie alla relazione

$$(b-a)P[D] = (b-a)\sup_n P[D_n] \le \sup_n P[(X_n - b)^+] \le c + |b|,$$

la variabile aleatoria D è integrabile (e quindi quasi certamente finita).

Se poi si denota con X_∞ una variabile aleatoria numerica equivalente, secondo P, a ciascuna delle due variabili aleatorie (3), il lemma di Fatou mostra che si ha

$$P[|X_\infty|] \le \liminf_n P[|X_n|] \le c.$$

Si vede dunque che X_∞ è integrabile, e quindi equivalente, secondo P, a una variabile aleatoria reale integrabile. Il corollario è così dimostrato. □

Martingale chiuse e teorema di Radon-Nikodým

1. Martingale chiuse

Su uno spazio probabilizzato $(\Omega, \mathcal{A}, P)$ siano assegnati un processo reale X e una filtrazione $\mathcal{F}$, ammettenti entrambi l'insieme ordinato I come insieme dei tempi. Si supponga che X sia adattato a $\mathcal{F}$ e che ciascuna delle variabili aleatorie X_t sia integrabile.

Se esiste un elemento Z di $\mathcal{L}^1(P)$, tale che, per ciascun istante t, si abbia

$$\int_A X_t \, \mathrm{d}P = \int_A Z \, \mathrm{d}P \quad \text{per ogni elemento } A \text{ di } \mathcal{F}_t, \tag{1.1}$$

allora X è una martingala rispetto a $\mathcal{F}$. Infatti, per ogni coppia (s,t) d'istanti con $s \le t$ (dove $\le$ denota la relazione d'ordine di I), ed ogni elemento A di $\mathcal{F}_s$, i due integrali $\int_A X_s \, \mathrm{d}P$, $\int_A X_t \, \mathrm{d}P$ sono tra loro eguali, essendo entrambi eguali a $\int_A Z \, \mathrm{d}P$.

L'implicazione inversa è falsa in generale: ossia, se X è una martingala rispetto alla filtrazione $\mathcal{F}$, non sempre esiste un elemento Z di $\mathcal{L}^1(P)$ tale che valga, per ogni t, la relazione (1.1). Quando un siffatto elemento Z esiste, si dice che esso *chiude* la martingala X, e questa si dice una martingala *chiusa*.

(1.2) Esempio. La martingala X è certamente chiusa se l'insieme I dei tempi ammette un massimo elemento. Infatti, in questo caso, detto u il massimo elemento di I, è ovvio che la variabile aleatoria X_u chiude la martingala X.

Da un teorema che dimostreremo più avanti (si veda (2.5)) risulterà che, affinché una martingala X sia chiusa, è *necessario* che le variabili aleatorie X_t siano uniformemente integrabili. Ci si può domandare se questa condizione sia anche *sufficiente*. Il teorema seguente risponde affermativamente alla domanda nel caso particolare in cui l'insieme dei tempi sia $\mathbb{N}$.

(1.3) Teorema. *Su uno spazio probabilizzato $(\Omega, \mathcal{A}, P)$ sia $\mathcal{F}$ una filtrazione, ammettente $\mathbb{N}$ come insieme dei tempi, e sia X una martingala rispetto a $\mathcal{F}$. Si supponga che la successione $(X_n)_{n\ge 0}$ sia uniformemente integrabile. Allora la martingala X è chiusa. Più precisamente, la successione $(X_n)_{n\ge 0}$ converge quasi certamente e in $\mathcal{L}^1(P)$ verso una variabile aleatoria reale Z, la quale chiude la martingala X.*

Dimostrazione. Dall'ipotesi di uniforme integrabilità discende in particolare che la successione (X_n) è limitata in $\mathcal{L}^1(P)$. Un noto corollario della diseguaglianza di Doob sulle discese mostra dunque che la successione $(X_n)_{n\ge 0}$ converge quasi certamente verso una variabile aleatoria reale Z. Grazie al teorema di Vitali, la convergenza ha

luogo anche in $\mathcal{L}^1(P)$. Rimane da provare che Z chiude la martingala X. A questo scopo, fissiamo un intero positivo m e un elemento A di $\mathcal{F}_m$. Si ha allora

$$\int_A X_m \, dP = \int_A X_n \, dP \quad \text{per} \quad n \geq m.$$

Di qui, sfruttando il fatto che la successione $(X_n)_{n\geq 0}$ converge in $\mathcal{L}^1(P)$ verso Z, si ricava l'eguaglianza

$$\int_A X_m \, dP = \int_A Z \, dP.$$

È così provato che Z chiude la martingala X. □

Il teorema che segue estende parzialmente il risultato precedente al caso di una martingala ammettente come insieme dei tempi un insieme ordinato filtrante.

(1.6) Teorema. *Su uno spazio probabilizzato $(\Omega, \mathcal{A}, P)$ sia $\mathcal{F}$ una filtrazione, ammettente come insieme dei tempi un insieme ordinato filtrante I, e sia X una martingala rispetto a $\mathcal{F}$. Si supponga che la famiglia $(X_t)_{t\in I}$ sia uniformemente integrabile. Allora la martingala X è chiusa. Più precisamente, la famiglia $(X_t)_{t\in I}$ converge in $\mathcal{L}^1(P)$ verso una variabile aleatoria reale Z, la quale chiude la martingala X.*

Dimostrazione. Per ogni successione crescente $(t_n)_{n\geq 0}$ di elementi di I, la successione $(X_{t_n})_{n\geq 0}$ è uniformemente integrabile ed è una martingala rispetto alla filtrazione $(\mathcal{F}_{t_n})_{n\geq 0}$. Grazie al Teorema (1.3), essa è dunque di Cauchy nello spazio pseudo-metrico $\mathcal{L}^1(P)$. Per l'arbitrarietà della successione crescente $(t_n)_{n\geq 0}$, ciò permette, com'è noto, di affermare che la famiglia $(X_t)_{t\in I}$ è di Cauchy in $\mathcal{L}^1(P)$. Grazie alla completezza di $\mathcal{L}^1(P)$, essa converge dunque in $\mathcal{L}^1(P)$ verso una variabile aleatoria reale Z. Rimane da provare che Z chiude la martingala X. A questo scopo, fissiamo un elemento s di I e un elemento A di $\mathcal{F}_s$. Si ha allora, indicando con $\leq$ la relazione d'ordine di I,

$$\int_A X_s \, dP = \int_A X_t \, dP \quad \text{per} \quad s \leq t.$$

Di qui, sfruttando il fatto che la famiglia $(X_t)_{t\in I}$ converge in $\mathcal{L}^1(P)$ verso Z, si ricava l'eguaglianza

$$\int_A X_s \, dP = \int_A Z \, dP.$$

È così provato che Z chiude la martingala X. □

2. Il teorema di Radon-Nikodým

Ci proponiamo ora di esporre una semplice dimostrazione del teorema di Radon-Nikodým, fondata sull'impiego del Teorema (1.6).

(2.1) Teorema. (Radon-Nikodým) *Su uno spazio misurabile $(\Omega, \mathcal{A})$ siano μ, ν due misure finite, tali che ν sia assolutamente continua rispetto a μ (cioè tali che ogni insieme trascurabile secondo μ sia trascurabile anche secondo ν). Allora esiste un elemento Z di $\mathcal{L}^1(\mu)$ tale che si abbia $\nu = Z \cdot \mu$, ossia*

$$\nu(A) = \int_A Z \, d\mu \quad \textit{per ogni elemento } A \textit{ di } \mathcal{A}.$$

Dimostrazione. Ci si può limitare al caso in cui ν sia maggiorata da μ. Per convincersi di ciò, si supponga acquisito il teorema in questo caso particolare. Allora, nel caso generale, posto $\lambda = \mu + \nu$, ciascuna delle due misure μ, ν, essendo maggiorata da λ, si potrà esprimere come una misura di base λ:

$$\mu = f \cdot \lambda, \qquad \nu = g \cdot \lambda.$$

Inoltre, posto $A = \{f = 0\}$, l'insieme A sarà trascurabile secondo μ, dunque secondo ν, dunque anche secondo λ, ciò che permetterà di scrivere le relazioni seguenti:

$$\mu = (f + I_A) \cdot \lambda, \quad \Rightarrow \quad \lambda = (f + I_A)^{-1} \cdot \mu, \quad \Rightarrow \quad \nu = \left(g\,(f + I_A)^{-1}\right) \cdot \mu.$$

Possiamo dunque supporre $\nu \leq \mu$. Inoltre, poiché la tesi si riduce a una banalità quando μ sia identicamente nulla, potremo anche supporre che la misura μ sia normalizzata. La indicheremo allora con P.

Sia I l'insieme di tutte le partizioni finite di Ω costituite da elementi di $\mathcal{A}$. Se $\mathcal{P}$ è una siffatta partizione, allora la tribù $\mathcal{T}(\mathcal{P})$ da essa generata è formata da tutte le unioni di elementi di $\mathcal{P}$ (ivi compresa l'unione della famiglia vuota). Considereremo l'insieme I come munito della relazione d'ordine così definita: la partizione $\mathcal{Q}$ segue la partizione $\mathcal{P}$ se $\mathcal{Q}$ è più fine di $\mathcal{P}$, ossia se la tribù $\mathcal{T}(\mathcal{Q})$ contiene la tribù $\mathcal{T}(\mathcal{P})$. Questa relazione fa di I un insieme ordinato filtrante. Consideriamo il processo reale X, avente I come "insieme dei tempi", così definito: per ogni elemento $\mathcal{P}$ di I, la variabile aleatoria $X_{\mathcal{P}}$ è data dalla formula

$$X_{\mathcal{P}} = \sum_{A \in \mathcal{P},\, P(A) \neq 0} \big(\nu(A)/P(A)\big)\, I_A \tag{2.2}$$

Si osservi che $X_{\mathcal{P}}$ è una variabile aleatoria semplice su $(\Omega, \mathcal{A})$, misurabile rispetto alla tribù $\mathcal{T}(\mathcal{P})$. Sul generico elemento A non trascurabile della partizione $\mathcal{P}$, essa assume il valore costante $\nu(A)/P(A)$ (mentre è nulla fuori dell'unione degli elementi non trascurabili di $\mathcal{P}$). Per ogni elemento non trascurabile A di $\mathcal{P}$, si ha dunque

$$\int_A X_{\mathcal{P}}\, \mathrm{d}P = P(A)\,\big(\nu(A)/P(A)\big) = \nu(A).$$

Ma l'eguaglianza

$$\int_A X_{\mathcal{P}}\, \mathrm{d}P = \nu(A) \tag{2.3}$$

vale anche se A è un elemento di $\mathcal{P}$ con $P(A) = 0$. In questo caso, infatti, essa si riduce (grazie all'ipotesi $\nu \leq P$) alla forma $0 = 0$. La relazione (2.3) vale anche se A è un elemento di $\mathcal{T}(\mathcal{P})$: infatti, in questo caso, A è unione di elementi di $\mathcal{P}$ (che sono a due a due disgiunti). Si vede dunque che, se $\mathcal{Q}$ è un elemento di I più fine di $\mathcal{P}$, allora, per ogni elemento A di $\mathcal{T}(\mathcal{P})$, i due numeri $\int_A X_{\mathcal{P}}\, \mathrm{d}P, \int_A X_{\mathcal{Q}}\, \mathrm{d}P$ sono tra loro eguali, essendo entrambi eguali a $\nu(A)$. Ciò mostra che, sullo spazio $(\Omega, \mathcal{A}, P)$, il processo X è una martingala rispetto alla filtrazione

$$(\mathcal{T}(\mathcal{P}))_{\mathcal{P} \in I}.$$

Inoltre le variabili aleatorie $X_{\mathcal{P}}$ sono a valori in $[0,1]$, dunque uniformemente integrabili. Grazie al Teorema (1.6), ne segue che la famiglia $(X_{\mathcal{P}})_{\mathcal{P}\in I}$ converge in $\mathcal{L}^1(P)$ verso una variabile aleatoria Z, la quale chiude la martingala X. È facile vedere che Z possiede la proprietà desiderata. Infatti, assegnato un elemento A di $\mathcal{A}$, è chiaro che esiste una partizione $\mathcal{P}$, appartenente a I, tale che A sia un elemento della tribù $\mathcal{T}(\mathcal{P})$. Se allora si tiene conto dell'eguaglianza (2.3) e del fatto che Z chiude la martingala X, si trova

$$\nu(A) = \int_A X_{\mathcal{P}}\,\mathrm{d}P = \int_A Z\,\mathrm{d}P.$$

Il teorema è così dimostrato. □

È noto che il teorema di Radon-Nikodým è lo strumento fondamentale per dimostrare l'esistenza della speranza condizionale. D'altra parte, una volta acquisita la nozione di speranza condizionale, è possibile impiegarla per caratterizzare nel modo seguente il concetto di martingala chiusa.

(2.4) Proposizione. *Su uno spazio probabilizzato $(\Omega, \mathcal{A}, P)$ siano assegnati un processo reale X e una filtrazione $\mathcal{F}$, ammettenti entrambi l'insieme ordinato I come insieme dei tempi. Sia inoltre Z un elemento di $\mathcal{L}^1(P)$. Affinché X sia, rispetto a $\mathcal{F}$, una martingala chiusa da Z, occorre e basta che, per ogni elemento t di I, la variabile aleatoria X_t sia una versione della speranza condizionale $P\big[\,Z \,|\, \mathcal{F}_t\,\big]$:*

Questa caratterizzazione permette, quando sia assegnata, su uno spazio probabilizzato $(\Omega, \mathcal{A}, P)$, una filtrazione $\mathcal{F} = (\mathcal{F}_t)_{t\in I}$, di costruire, per ogni elemento Z di $\mathcal{L}^1(P)$, un processo X che sia, rispetto a $\mathcal{F}$, una martingala chiusa da Z: per ottenere un tal processo, basta scegliere, per ogni elemento t di I, una versione X_t della speranza condizionale $P\big[\,Z \,|\, \mathcal{F}_t\,\big]$.

Il teorema seguente mostra che, se si parte da una famiglia di variabili aleatorie uniformemente integrabili, e si opera su di esse con operatori di speranza condizionale (eventualmente relativi a sottotribù diverse), si ottiene sempre una famiglia di variabili aleatorie uniformemente integrabili.

(2.5) Teorema. *Su uno spazio probabilizzato $(\Omega, \mathcal{A}, P)$ sia $(Z_i)_{i\in I}$ una famiglia di variabili aleatorie uniformemente integrabili, avente come insieme degli indici un arbitrario insieme I. Sia poi $(\mathcal{F}_i)_{i\in I}$ una famiglia di sottotribù di $\mathcal{A}$, ammettente anch'essa I come insieme degli indici. Per ogni indice i, sia X_i una versione della speranza condizionale $P\big[\,Z_i \,|\, \mathcal{F}_i\,\big]$.*

Allora la famiglia $(X_i)_{i\in I}$ è uniformemente integrabile.

Dimostrazione. Grazie alla decomposizione $Z_i = Z_i^+ - Z_i^-$, si può supporre, senza ledere la generalità, che le Z_i, e quindi le X_i, siano positive. Poiché le Z_i sono uniformemente integrabili, posto

$$c = \sup_i P\big[Z_i\big] = \sup_i P\big[X_i\big],$$

si ha $c < \infty$. Fissato un numero reale ϵ maggiore di zero, si scelga un numero reale δ maggiore di zero, tale che, per ogni evento A con $P(A) < \delta$, si abbia

$$\sup_i \int_A Z_i \, \mathrm{d}P \leq \epsilon.$$

Sia poi t un numero reale maggiore di zero. Per ogni elemento i di I, si ha allora (diseguaglianza di Markov)

$$t\, P\{X_i > t\} \leq P[X_i] \leq c.$$

Dunque, se t è tale che risulti $c/t < \delta$, allora, per ogni i, si ha $P\{X_i > t\} \leq c/t < \delta$, e quindi

$$\int_{\{X_i > t\}} X_i \, \mathrm{d}P = \int_{\{X_i > t\}} Z_i \, \mathrm{d}P \leq \epsilon.$$

È così provato che la famiglia $(X_i)_{i \in I}$ è uniformemente integrabile. □

(2.6) Proposizione. *Su uno spazio probabilizzato $(\Omega, \mathcal{A}, P)$ sia $\mathcal{F}$ una filtrazione ammettente $\overline{\mathbb{N}}$ come insieme dei tempi, e si ponga*

$$\mathcal{F}_\infty = \mathcal{T}\left(\bigcup_{n \geq 0} \mathcal{F}_n\right). \tag{2.7}$$

Siano inoltre Z un elemento di $\mathcal{L}^1(P)$ e $(X_n)_{n \geq 0}$ una martingala (rispetto alla filtrazione $\mathcal{F}$) chiusa da Z, cioè $X_n = P[Z \mid \mathcal{F}_n]$ $\forall n \geq 0$.

Allora $(X_n)_{n \geq 0}$ converge quasi certamente e in $\mathcal{L}^1$ verso una versione della speranza condizionale $P[Z \mid \mathcal{F}_\infty]$.

Dimostrazione. Grazie a (2.5), la martingala $(X_n)_{n \geq 0}$, essendo chiusa, è uniformemente integrabile. Risulta dunque da (1.3) che essa converge, quasi certamente e in media, verso una variabile aleatoria reale U che la chiude. Denotiamo con V la variabile aleatoria reale che coincide con la variabile aleatoria numerica $\liminf_n X_n$ dove questa è finita e con la costante 0 altrove. Poiché V è equivalente a U, è chiaro che V, al pari di U, chiude la martingala $(X_n)_{n \geq 0}$ ed è per essa limite nel senso della convergenza quasi certa e della convergenza in media. Basta dunque provare che V è una versione della speranza condizionale $P[Z \mid \mathcal{F}_\infty]$. A questo scopo, osserviamo innanzitutto che V è misurabile rispetto alla tribù $\mathcal{T}([X_n]_{n \geq 0})$ e quindi, a maggior ragione, rispetto a $\mathcal{F}_\infty$. Osserviamo poi che, per ogni intero positivo n, si ha

$$\int_A V \, \mathrm{d}P = \int_A X_n \, \mathrm{d}P = \int_A Z \, \mathrm{d}P \quad \text{per } A \in \mathcal{F}_n.$$

Ne segue

$$\int_A V \mathrm{d}P = \int_A Z \mathrm{d}P \quad \text{per } A \in \bigcup_{n \geq 0} \mathcal{F}_n.$$

Grazie a (2.7) (e al fatto che l'unione delle $\mathcal{F}_n$ è un'algebra), ciò basta per concludere. □

Convergenza di martingale

Supporremo fissato uno spazio probabilizzato $(\Omega, \mathcal{A}, P)$. Tutte le variabili aleatorie considerate saranno definite su di esso e, se non diversamente specificato, a valori in $\mathbb{R}$. Tutti i processi considerati saranno tacitamente supposti reali, definiti su $(\Omega, \mathcal{A}, P)$ e aventi $\mathbb{N}$ come insieme dei tempi. Assegnati un processo X e una variabile aleatoria T a valori in $\mathbb{N} \cup \{\infty\}$, si denoterà con X_T la variabile aleatoria che è nulla su $\{T = \infty\}$ e che, per ciascun intero positivo n, coincide con X_n su $\{T = n\}$.

È noto che ogni martingala (e, più in generale, ogni sottomartingala) che sia limitata in $\mathcal{L}^1(P)$ converge quasi certamente verso una variabile aleatoria reale integrabile. Ci proponiamo di illustrare alcune conseguenze di questo fondamentale risultato. Cominciamo con l'osservare che, se X è una martingala, allora, grazie alle relazioni

$$P\big[X_n^+\big] - P\big[X_n^-\big] = P\big[X_n\big] = P\big[X_0\big],$$

X è limitata in $\mathcal{L}^1(P)$ non appena sia limitata una delle due successioni numeriche

$$(P[X_n^+])_{n\geq 0}, \qquad (P[X_n^-])_{n\geq 0}.$$

In particolare, X è limitata in $\mathcal{L}^1(P)$ se le variabili aleatorie X_n sono tutte *maggiorate* da un medesimo elemento di $\mathcal{L}^1(P)$ (oppure tutte *minorate* da un medesimo elemento di $\mathcal{L}^1(P)$).

1. Teorema. *Sia $(X_n)_{n\geq 1}$ una successione di variabili aleatorie appartenenti a $\mathcal{L}^2(P)$, centrate, indipendenti, e si ponga $S_n = X_1 + \cdots + X_n$. Se la successione $(S_n)_{n\geq 0}$ converge in $\mathcal{L}^2(P)$ (cioè se si ha $\sum_{n\geq 1} \mathrm{Var}[X_n] < \infty$), allora essa è una martingala limitata in $\mathcal{L}^2(P)$, e quindi converge anche quasi certamente.*

2. Corollario. *Sia $(X_n)_{n\geq 1}$ una successione di variabili aleatorie appartenenti a $\mathcal{L}^2(P)$, centrate, indipendenti, e si ponga $S_n = X_1 + \cdots + X_n$. Sia inoltre $(a_n)_{n\geq 1}$ una successione crescente di numeri reali strettamente positivi, con*

$$a_n \uparrow \infty, \qquad \sum_{n\geq 1} \mathrm{Var}\,[X_n/a_n] < \infty.$$

Allora la serie $\sum_n (X_n/a_n)$ converge quasi certamente, e quindi (per il classico Lemma di Kronecker) la successione $(S_n/a_n)_{n\geq 1}$ converge quasi certamente verso zero.

Se, nel corollario precedente, si prende $a_n = n$, si trova un risultato *analogo* alla legge forte dei grandi numeri di Rajchmann (con l'ipotesi d'indipendenza in luogo della semplice ipotesi di non correlazione, ma, in compenso, con un'ipotesi *più debole* della limitatezza delle varianze).

Una parziale inversione del Teorema 1 è fornita dal teorema seguente (nel quale la semplice appartenenza delle X_n a $\mathcal{L}^2(P)$ è sostituita con un'ipotesi ben più severa: quella della loro equilimitatezza).

3. Teorema. *Sia $(X_n)_{n\geq 1}$ una successione di variabili aleatorie equilimitate, centrate, indipendenti, e si ponga $S_n = X_1 + \cdots + X_n$. Se la successione $(S_n)_{n\geq 0}$ non converge in $\mathcal{L}^2(P)$ (cioè se risulta $\sum_{n\geq 1} \mathrm{Var}[X_n] = \infty$), allora si ha*

$$P\{\sup_n |S_n| = \infty\} = 1.$$

Dimostrazione. Denotiamo con c una costante reale che maggiori i moduli di tutte le X_n. Denotiamo inoltre con $\mathcal{F}$ la filtrazione naturale del processo $(S_n)_{n\geq 0}$. Ricordiamo che questo processo è una martingala. Di conseguenza, il processo $(S_n^2)_{n\geq 0}$ è una sottomartingala. Ricordiamo in che modo se ne possa costruire un compensatore prevedibile (ossia un processo prevedibile V tale che il processo $\big(S_n^2 - V_n\big)_{n\geq 0}$ sia una martingala): basta prendere, per ogni intero n strettamente positivo, una versione D_n della speranza condizionale

$$P\big[\, S_n^2 - S_{n-1}^2 \,|\, \mathcal{F}_{n-1} \,\big] \tag{1}$$

e porre quindi, per ogni intero positivo n,

$$V_n = D_1 + \cdots + D_n \qquad \text{(in particolare, } V_0 = 0\text{)}.$$

D'altra parte, la speranza condizionale (1) è identica (come a suo tempo osservato) alla speranza condizionale $P\big[\, X_n^2 \,|\, \mathcal{F}_{n-1} \,\big]$, la quale, grazie all'indipendenza di X_n da $\mathcal{F}_{n-1}$, ammette come versione la costante $D_n = P\big[X_n^2\big] = \mathrm{Var}\big[X_n\big]$. In conclusione, la sottomartingala $(S_n^2)_{n\geq 0}$ ammette come compensatore prevedibile il processo "deterministico" V definito da

$$V_n = \textstyle\sum_{k=1}^n \mathrm{Var}\,[X_k] = \mathrm{Var}\,[S_n],$$

Si tratta di provare che, per ogni numero reale positivo a, è trascurabile l'evento $\{\sup_n |S_n| \leq a\}$. Questo evento è identico a $\{T = \infty\}$, dove T denota il tempo d'arresto così definito: $T(\omega) = \inf\{n \in \mathbb{N} : |S_n(\omega)| > a\}$. (Si osservi che è $T \geq 1$.) Si tratta perciò di dimostrare la relazione

$$P\{T = \infty\} = 0. \tag{2}$$

Poiché il processo $(S_n^2 - V_n)_{n\geq 0}$ è una martingala nulla all'istante zero, tale è anche il processo ottenuto arrestandolo all'istante T. Dunque, per ogni intero n strettamente positivo, si ha

$$P\big[\, S_{n\wedge T}^2 - V_{n\wedge T} \,\big] = 0,$$

e quindi

$$P\big[V_{n\wedge T}\big] = P\big[S_{n\wedge T}^2\big] \leq (c+a)^2,$$

dove la diseguaglianza finale discende dalla relazione

$$|S_{n\wedge T}| \leq \big|S_{n\wedge T} - S_{(n\wedge T)-1}\big| + \big|S_{(n\wedge T)-1}\big| \leq c + a.$$

Ne segue, a maggior ragione,

$$P\left[V_{n\wedge T}\, I_{\{T=\infty\}}\right] \leq (c+a)^2,$$

ossia $V_n\, P\{T = \infty\} \leq (c+a)^2$. Di qui, essendo $V_n \uparrow \infty$, si ricava la relazione (2) da dimostrare. □

4. Corollario. *Siano assegnate una successione $(a_n)_{n\geq 1}$ di numeri reali e una successione $(X_n)_{n\geq 1}$ di "segni aleatori", ossia di variabili aleatorie indipendenti e centrate, a valori in $\{-1,1\}$. Si denoti con A l'evento costituito dai punti di Ω nei quali la successione $(\sum_{k=1}^n a_k X_k)_{n\geq 1}$ converge. Allora:*

(a) Se $\sum_{n\geq 1} a_n^2 < \infty$, si ha $P(A) = 1$.

(b) Altrimenti, si ha $P(A) = 0$.

Dimostrazione. Si ha $\mathrm{Var}\,[a_k X_k] = a_k^2$. Perciò, nel caso (a), basta applicare il Teorema 1 alla successione $(a_n X_n)_{n\geq 1}$. Nel caso (b), la tesi è ovvia se la successione $(a_n)_{n\geq 1}$ è illimitata; altrimenti si ottiene applicando alla successione $(a_n X_n)_{n\geq 1}$ il teorema precedente. □

Il teorema che segue è un'estensione del Teorema 1.

5. Teorema. *Rispetto a una filtrazione $\mathcal{F}$ fissata su $(\Omega, \mathcal{A}, P)$, sia X una martingala, tale che ciascuna delle variabili aleatorie X_n sia di quadrato integrabile, e si denoti con V un compensatore prevedibile di X^2. Allora la successione $(X_n)_{n\geq 0}$ converge quasi certamente sull'evento $\{\sup_n V_n < \infty\}$.*

Dimostrazione. Senza ledere la generalità, supporremo che sia $X_0 = 0$.

Si tratta di provare che, per ogni numero reale positivo a, la successione $(X_n)_{n\geq 0}$ converge quasi certamente sull'evento $\{\sup_n V_n \leq a\}$. Questo è identico all'evento $\{T = \infty\}$, dove T denota il tempo d'arresto definito da

$$T(\omega) = \inf\{n \in \mathbb{N} : V_{n+1}(\omega) > a\}.$$

Si osservi che T è effettivamente un tempo d'arresto (rispetto alla filtrazione $\mathcal{F}$) perché si può scrivere nella forma

$$T(\omega) = \inf\{n \in \mathbb{N} : V'_n(\omega) > a\},$$

dove V' denota il processo adattato così definito: $V'_n = V_{n+1}$. Poiché il processo $(X_n^2 - V_n)_{n\geq 0}$ è una martingala nulla in 0, si ha, per ogni intero positivo n,

$$P\big[X^2_{n\wedge T}\big] = P\big[V_{n\wedge T}\big] \leq a$$

(dove la diseguaglianza finale segue dal fatto che la variabile aleatoria $V_{n\wedge T}$ è nulla su $\{n \wedge T = 0\}$ ed è maggiorata altrove da a). Ciò prova che la martingala $X^{|T}$ è limitata in $\mathcal{L}^2(P)$, dunque quasi certamente convergente. Poiché essa coincide con X sull'evento $\{T = \infty\}$, ne segue che, su questo evento, anche la martingala X converge quasi certamente. □

6. Teorema. *Sia X una martingala, i cui incrementi $X_n - X_{n-1}$ siano tutti maggiorati da un medesimo elemento Z di $\mathcal{L}^1(P)$. Allora essa converge quasi certamente sull'evento $\{\sup_n X_n < \infty\}$.*

Dimostrazione. Senza ledere la generalità, si può supporre $X_0 = 0$. Basterà provare che, per ogni numero reale positivo a, la successione $(X_n)_{n\geq 0}$ converge quasi certamente sull'evento $\{\sup_n X_n \leq a\}$. Questo evento è identico a $\{T = \infty\}$, dove T denota il tempo d'arresto così definito: $T(\omega) = \inf\{n \in \mathbb{N} : X_n(\omega) > a\}$. (Si osservi che è $T \geq 1$.) Per ogni intero n strettamente positivo, si ha

$$X_{n\wedge T} = \big(X_{n\wedge T} - X_{(n\wedge T)-1}\big) + X_{(n\wedge T)-1} \leq Z + a.$$

Ciò prova che la martingala $X^{\lfloor T}$ è limitata in $\mathcal{L}^1(P)$, dunque quasi certamente convergente. Poiché essa coincide con X sull'evento $\{T = \infty\}$, ne segue che, su questo evento, anche la martingala X converge quasi certamente. □

7. Lemma di Wald. *Sia Z un processo, con Z_0 integrabile, tale che gli incrementi $Z_n - Z_{n-1}$ siano tutti maggiorati in modulo da una medesima costante reale c. Sia poi T una variabile aleatoria, a valori in $\mathbb{N} \cup \{\infty\}$, con $P[T] < \infty$. Allora il processo $Z^{\lfloor T}$ è dominato in $\mathcal{L}^1(P)$ (nel senso che le variabili aleatorie $Z_{n\wedge T}$ sono maggiorate in modulo da un medesimo elemento di $\mathcal{L}^1(P)$).*

Dimostrazione. Per ogni n, si ha $Z_{n\wedge T} = Z_0 + \sum_{k\geq 1} I_{\{k\leq n\wedge T\}}\,(Z_k - Z_{k-1})$, e quindi

$$|Z_{n\wedge T}| \leq |Z_0| + c\sum_{k\geq 1} I_{\{k\leq n\wedge T\}} = |Z_0| + c\,(n \wedge T) \leq |Z_0| + cT. \qquad \square$$

8. Corollario. *Sia X una martingala tale che gli incrementi $X_n - X_{n-1}$ siano tutti maggiorati in modulo da una medesima costante reale. Sia poi a un numero reale, con $P\big[X_0\big] < a$, e si denoti con T il tempo d'arresto così definito:*

$$T(\omega) = \inf\{n \in \mathbb{N} : X_n(\omega) \geq a\}.$$

Allora T non è integrabile.

Dimostrazione. Ragionando per assurdo, si supponga T integrabile. Allora, grazie al Lemma di Wald, la martingala $X^{\lfloor T}$ è dominata in $\mathcal{L}^1(P)$. Perciò essa converge quasi certamente, e in $\mathcal{L}^1(P)$, verso una variabile aleatoria che la chiude. Osserviamo, d'altra parte, che l'evento quasi-certo $\{T < \infty\}$ è contenuto in $\{X_T \geq a\}$ e che, in ciascun punto di $\{T < \infty\}$, la successione $(X_{n\wedge T})_{n\geq 0}$ va definitivamente a coincidere con X_T. Si vede dunque che la martingala $X^{\lfloor T}$ è chiusa da X_T e che si ha

$$P\big[X_0\big] = P\big[X_T\big] \geq a.$$

Poiché ciò contraddice l'ipotesi $P\big[X_0\big] < a$, il corollario è dimostrato. □

Rovina del giocatore

Si suppongono fissati uno spazio probabilizzato $(\Omega, \mathcal{A}, P)$ e, su di esso, una successione $(X_n)_{n\geq 1}$ di variabili aleatorie indipendenti, a valori in $\{-1, 0, 1\}$. Si pone, come al solito, $S_n = X_1 + \cdots + X_n$ (in particolare, $S_0 = 0$), e si considera lo spazio $(\Omega, \mathcal{A}, P)$ come munito della filtrazione naturale associata al processo $(S_n)_{n\geq 0}$.

Inoltre, assegnati due interi a, b strettamente positivi, si considerano i due tempi d'arresto T_b, T_{-a} così definiti:

$$T_b(\omega) = \inf\{n : S_n(\omega) = b\}, \qquad T_{-a}(\omega) = \inf\{n : S_n(\omega) = -a\}.$$

Si pone infine

$$T = T_b \wedge T_{-a}, \qquad A = \{T_b < T_{-a}\}, \qquad B = \{T_{-a} < T_b\}.$$

1. Il caso simmetrico

(1.1) Teorema. *Si supponga* $P\{X_n = 1\} = P\{X_n = -1\} = p_n$ *e* $\sum_n p_n = \infty$. *Valgono allora le conclusioni seguenti:*

(a) L'insieme di non convergenza della successione $(S_n)_{n\geq 0}$ *è quasi certo.*

(b) Ciascuno dei due tempi d'arresto T_b, T_{-a} *è quasi certamente finito, ma non integrabile.*

(c) Si ha $P(A) = a/(a+b)$, $P(B) = b/(a+b)$.

(d) Se risulta $\inf_i p_i = p > 0$, *allora* T *è integrabile.*

(e) Se poi ciascuno dei numeri p_i *coincide con* p, *allora si ha* $P[T] = (ab)/(2p)$.

Dimostrazione. (a) L'insieme di non convergenza della successione $(S_n)_{n\geq 0}$ coincide con l'evento $\limsup_n\{|X_n| = 1\}$, il quale, grazie all'ipotesi $\sum_n p_n = \infty$, è quasi certo per il secondo lemma di Borel-Cantelli.

(b) Per ragioni di simmetria, basta occuparsi di T_b. La martingala $(S_n)_{n\geq 0}$ ha incrementi equilimitati. Pertanto essa converge quasi certamente sull'insieme $\{\sup_n S_n < \infty\}$ e quindi, in particolare, sull'insieme $\{T_b = \infty\}$. Sappiamo, d'altra parte, che il suo insieme di convergenza è trascurabile. Dunque è trascurabile anche l'evento $\{T_b = \infty\}$. Inoltre T_b non è integrabile, come risulta da un corollario del lemma di Wald.

(c) Poiché ciascuno dei due tempi d'arresto T_b, T_{-a} è quasi certamente finito, tale è il loro inviluppo inferiore T. La martingala $S^{|T}$ ottenuta arrestando $(S_n)_{n\geq 0}$ al tempo T, essendo compresa tra le costanti $-a$, b, è chiusa. D'altra parte, in ciascun punto dell'evento quasi certo $\{T < \infty\}$, la successione $(S_{n\wedge T})_{n\geq 0}$ va definitivamente a coincidere con S_T. Dunque S_T chiude la martingala $S^{|T}$. Si ha perciò

$$0 = P[S_0] = P[S_T] = P[b\,I_A - a\,I_B] = b\,P(A) - a\,P(B).$$

La relazione così ottenuta, unita all'ovvia eguaglianza $P(A) + P(B) = 1$, dà luogo alle desiderate espressioni per le probabilità $P(A)$, $P(B)$.

(d) Per ogni indice i, si ha

$$\mathrm{Var}\,[X_i] = P\big[X_i^2\big] = P\{|X_i| = 1\} = 2\,p_i.$$

Pertanto la sottomartingala $(S_n^2)_{n\geq 0}$ ammette, come compensatore prevedibile, il processo deterministico V così definito:

$$V_n = \mathrm{Var}\,[S_n] = \textstyle\sum_{i=1}^n \mathrm{Var}\,[X_i] = 2\,\sum_{i=1}^n p_i.$$

Dalla definizione di p come estremo inferiore dei p_i risulta $V_n \geq 2np$. Per ogni n si ha dunque

$$2p\,P[n\wedge T] \leq P\,[V_{n\wedge T}] = P\,\big[S^2_{n\wedge T}\big] \tag{1.2}$$

(dove l'eguaglianza finale è dovuta al fatto che il processo ottenuto arrestando la martingala $(S_n^2 - V_n)_{n\geq 0}$ all'istante T è ancora una martingala nulla in 0). Dalla relazione (1.2), facendo tendere n all'infinito (e applicando il teorema di Beppo-Levi al primo membro, quello di Lebesgue all'ultimo membro), si ricava

$$2p\,P\,[T] \leq P\big[S_T^2\big] = P\big[b^2 I_A + a^2 I_B\big] = b^2\,\frac{a}{a+b} + a^2\,\frac{b}{a+b} = ab.$$

Questa relazione, essendo p non nullo, mostra che T è integrabile.

(e) Se poi ciascuno dei numeri p_i coincide con p, allora la diseguaglianza che figura nella precedente relazione si riduce a un'eguaglianza, e ciò permette di ricavare, per la speranza di T, il valore esplicito $(ab)/(2p)$. □

2. Il caso non simmetrico*

(2.1) Teorema. *Si supponga che ciascuna delle X_n possa prendere soltanto i valori 1 e -1, con probabilità rispettive p, q, dove p, q sono due numeri reali strettamente positivi, con $p > q$. Si ha allora:*

$$P(A) = \frac{(q/p)^{-a} - 1}{(q/p)^{-a} - (q/p)^b}\,, \quad P(B) = \frac{1 - (q/p)^b}{(q/p)^{-a} - (q/p)^b}\,, \quad P\big\{T_{-a} < \infty\big\} = (q/p)^a$$

Dimostrazione. La legge forte dei grandi numeri assicura che la successione $(S_n/n)_{n\geq 1}$ converge quasi certamente verso la costante $p - q$ (strettamente positiva). Pertanto la successione $(S_n)_{n\geq 1}$ tende quasi certamente a $+\infty$. Ciò prova che il tempo d'arresto T_b è quasi certamente finito. A maggior ragione è quasi certamente finito T. Essendo

$$P\,\big[(q/p)^{X_i}\big] = p\,(q/p) + q\,(q/p)^{-1} = 1,$$

il processo Y definito da

$$Y_n = \prod_{i=1}^{n} (q/p)^{X_i} = (q/p)^{S_n}$$

è una martingala (con $Y_0 = 1$). La martingala arrestata $Y^{|T}$, essendo positiva e maggiorata dalla costante $(q/p)^{-a}$, è chiusa. D'altra parte, in ciascun punto dell'evento quasi certo $\{T < \infty\}$, la successione $(Y_{n\wedge T})_{n\geq 0}$ va definitivamente a coincidere con Y_T. Dunque Y_T chiude la martingala $Y^{|T}$, sicché risulta

$$1 = P\big[Y_0\big] = P\big[Y_T\big] = P\big[(q/p)^b I_A + (q/p)^{-a} I_B\big] = (q/p)^b\, P(A) + (q/p)^{-a}\, P(B).$$

La relazione così trovata, unita all'ovvia eguaglianza $P(A) + P(B) = 1$, fornisce le desiderate espressioni per le probabilità $P(A)$, $P(B)$. Infine, poiché la successione $\big(T_b\big)_{b\in\mathbb{N}^*}$ è crescente e tende a $+\infty$ (grazie alla maggiorazione $b \leq T_b$), si ha

$$P\{T_{-a} < \infty\} = \lim_{b\to\infty} P\{T_{-a} < T_b\} = \lim_{b\to\infty} \frac{1-(q/p)^b}{(q/p)^{-a} - (q/p)^b} = \frac{1}{(q/p)^{-a}} = (q/p)^a. \qquad \square$$

Si osservi che, nelle ipotesi del teorema appena dimostrato, l'evento $\{T_{-a} = \infty\}$ non è trascurabile, sicché la variabile aleatoria T_{-a} non è integrabile. Al contrario, T_b (e a maggior ragione T) è integrabile, come risulta dal teorema seguente.

(2.2) Teorema. *Nelle stesse ipotesi del teorema precedente, si ha*

$$P\big[T_b\big] = b/(p-q), \quad P[T] = \big[b\,P(A) - a\,P(B)\big]/(p-q).$$

Dimostrazione. Consideriamo la martingala Z (nulla in 0 e con incrementi equilimitati) definita da

$$Z_n = S_n - n(p-q).$$

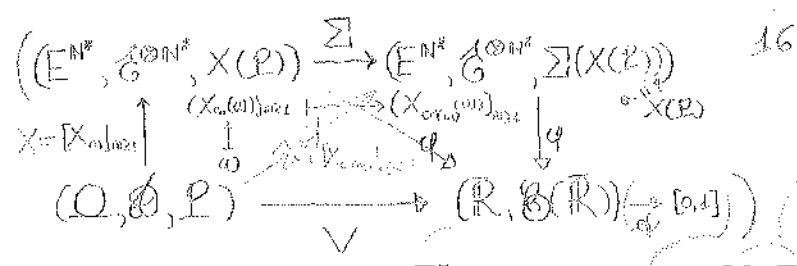
Poiché il processo ottenuto arrestandola all'istante T_b è ancora una martingala nulla in 0, si trova innanzitutto $P\,[Z_{n\wedge T_b}] = 0$, ossia $P\,[S_{n\wedge T_b} - (p-q)(n\wedge T_b)] = 0$, e quindi

$$(p-q)\,P\big[n\wedge T_b\big] = P\big[S_{n\wedge T_b}\big] \leq b.$$

Di qui, facendo tendere n all'infinito, si ricava $(p-q)\,P\big[T_b\big] \leq b$, e ciò prova l'integrabilità di T_b (e, a maggior ragione, di T). Dal lemma di Wald discende allora che le due martingale arrestate $Z^{|T_b}$, $Z^{|T}$ sono entrambe dominate in L^1. Pertanto esse sono chiuse, e precisamente sono chiuse dalle due variabili aleatorie Z_{T_b}, Z_T rispettivamente. Si ha dunque $P\,[Z_{T_b}] = 0$, ossia $P\,[S_{T_b} - (p-q)T_b] = 0$, e un'eguaglianza analoga, con T in luogo di T_b. Ne discendono le relazioni

$$(p-q)\,P\big[T_b\big] = P\big[S_{T_b}\big] = b, \qquad (p-q)\,P[T] = P\big[S_T\big] = b\,P(A) - a\,P(B),$$

dalle quali si ricavano le desiderate espressioni per le speranze di T_b e di T. $\square$



Il teorema di Hewitt e Savage

Definizione. Su uno spazio probabilizzato $(\Omega, \mathcal{A}, P)$ sia $(X_n)_{n\geq 1}$ una successione di variabili aleatorie a valori in un medesimo spazio misurabile $(E, \mathcal{E})$. Si dice che essa è *scambiabile* se, per ogni permutazione σ di $\mathbb{N}^*$, la quale sia *di tipo finito*, cioè tale che l'insieme $\{n \in \mathbb{N}^* : \sigma(n) \neq n\}$ sia finito, la legge del blocco $[X_n]_{n\geq 1}$ è identica a quella del blocco $\left[X_{\sigma(n)}\right]_{n\geq 1}$. Inoltre una variabile aleatoria numerica V si dice una *funzione simmetrica delle* X_n se è della forma $V = \varphi \circ [X_n]_{n\geq 1}$, con φ funzione numerica misurabile sullo spazio d'arrivo del blocco $[X_n]_{n\geq 1}$, tale che, per ogni permutazione σ di $\mathbb{N}^*$, di tipo finito, ed ogni successione $(x_n)_{n\geq 1}$ di elementi di E, si abbia $\varphi(x_1, x_2, \ldots) = \varphi\big(x_{\sigma(1)}, x_{\sigma(2)}, \ldots\big)$.

Teorema. *Nelle ipotesi della definizione precedente, si supponga che la successione $(X_n)_{n\geq 1}$ sia scambiabile e che la variabile aleatoria numerica V sia una funzione simmetrica delle X_n. Allora V è equivalente a una variabile aleatoria terminale rispetto alla successione $(X_n)_{n\geq 1}$.*

Dimostrazione. Senza ledere la generalità, supporremo V limitata. Per ogni intero $n \geq 1$, sia $\varphi_n(X_1, \ldots, X_n)$ (con φ_n misurabile e limitata) una versione della speranza condizionale $P[V \mid X_1, \ldots, X_n]$. Allora la successione

$$\big(\varphi_n(X_1, \ldots, X_n)\big)_{n\geq 1}$$

converge verso V in L^1, ossia verifica la relazione

$$\lim_n P\left[|V - \varphi_n(X_1, \ldots, X_n)|\right] = 0.$$

D'altra parte, grazie alla scambiabilità delle X_i e al carattere simmetrico di V, risulta

$$P\left[|V - \varphi_n(X_1, \ldots, X_n)|\right] = P\left[|V - \varphi_n(X_{n+1}, \ldots, X_{2n})|\right].$$

Dunque anche la successione $(\varphi_n(X_{n+1}, \ldots, X_{2n}))_{n\geq 1}$ converge in L^1 verso V. Una sua opportuna sottosuccessione converge verso V quasi certamente; ne segue che V, essendo equivalente al limite inferiore di questa sottosuccessione, è equivalente a una variabile aleatoria terminale. □

Corollario. *Sia $(X_n)_{n\geq 1}$ una successione di variabili aleatorie reali indipendenti, aventi una stessa legge, diversa dalla legge della costante 0.*

Si ponga $S_n = X_1 + \cdots + X_n$. Allora ciascuna delle due variabili aleatorie numeriche

$$U = \liminf_{n\to\infty} S_n, \qquad V = \limsup_{n\to\infty} S_n$$

è equivalente a una costante non finita, cioè a $-\infty$ o $+\infty$.

Dimostrazione. Occupiamoci ad esempio di U. Grazie al teorema precedente, U è equivalente a una variabile aleatoria numerica terminale, dunque (per la legge zero–uno di Kolmogorov) a una costante numerica c. Se, per assurdo, questa costante fosse finita, si avrebbe

$$c \sim U \overset{(\text{ovvio})}{=} \liminf_{n\geq 1} S_{n+1} \overset{(\text{ovvio})}{=} X_1 + \liminf_{n\geq 1}(X_2 + \ldots + X_{n+1}) \sim X_1 + c,$$

e quindi $X_1 \sim 0$, contrariamente all'ipotesi.

Il corollario è così dimostrato. □

Alcuni risultati riguardanti le passeggiate aleatorie sulla retta

1. Conseguenze elementari del teorema di Hewitt e Savage

Cominciamo col ricordare il seguente risultato, conseguenza del teorema di Hewitt e Savage.

(1.1) Teorema. *Assegnata, su uno spazio probabilizzato $(\Omega, \mathcal{A}, P)$, una successione $(X_n)_{n\geq 1}$ di variabili aleatorie reali indipendenti, di comune legge ν, con $\nu \neq \epsilon_0$, si consideri la passeggiata aleatoria $(S_n)_{n\geq 0}$ definita da $S_n = X_1 + \cdots + X_n$. Si ponga inoltre*

$$U = \liminf_n S_n, \qquad V = \limsup_n S_n. \tag{1.2}$$

Allora le due variabili aleatorie numeriche U, V sono rispettivamente equivalenti, secondo P, a due costanti u, v entrambe infinite.

(1.3) Corollario. *Nelle ipotesi del teorema precedente, si supponga per giunta che la legge ν sia simmetrica (cioè che X_1 sia isonoma a $-X_1$).*

Si ha allora $v = \infty$, $u = -\infty$.

Dimostrazione. Detta φ la funzione che ad ogni successione $(x_n)_{n\geq 1}$ di numeri reali associa il numero $\limsup_n(x_1 + \cdots + x_n)$, si ha

$$V = \varphi \circ [X_n]_{n\geq 1}, \qquad U = -\varphi \circ [-X_n]_{n\geq 1}.$$

D'altra parte, i due blocchi $[X_n]_{n\geq 1}$, $[-X_n]_{n\geq 1}$ sono tra loro isonomi (ciascuno avendo come legge il prodotto di una successione di copie di ν). Dunque la variabile aleatoria U è isonoma a $-V$, sicché risulta $u = -v$. Non potendo essere $v < u$, ne segue $v = \infty$, $u = -\infty$. □

(1.4) Corollario. *Nelle ipotesi di (1.1), si supponga per giunta che la legge ν abbia momento del secondo ordine finito e sia centrata.*

Si ha allora $v = \infty$, $u = -\infty$.

Dimostrazione. Senza ledere la generalità, supporremo che il momento del secondo ordine di ν sia eguale a 1. Basterà provare che non può essere $u = v = \infty$. (Infatti, applicando questo risultato alla successione $(-X_n)_{n\geq 1}$, se ne dedurrà che è parimenti impossibile che sia $u = v = -\infty$.)

Ragionando per assurdo, supponiamo $u = v = \infty$. Ciò vuol dire che la successione $(S_n)_{n\geq 0}$ tende a ∞ quasi certamente, ossia che la successione $\big(\exp(-S_n)\big)_{n\geq 0}$ converge verso zero quasi certamente (e quindi in probabilità). Si ha dunque

$$P\{S_n < 0\} = P\{\exp(-S_n) > 1\} \to 0.$$

Ma ciò non è possibile: infatti il teorema di Lindeberg-Lévy assicura che la successione $(S_n/\sqrt{n})_{n\geq 1}$ converge in legge verso $\mathcal{N}(0,1)$, sicché risulta

$$P\{S_n < 0\} = P\{S_n/\sqrt{n} < 0\} \to \mathcal{N}(0,1)(]-\infty, 0[) = 1/2\,. \qquad \square$$

2. Estensione di uno dei risultati precedenti *

Nel Corollario (1.4), l'ipotesi che il momento del secondo ordine della legge ν sia finito è molto comoda perché permette, come si è visto, di ridurre la dimostrazione a una banale applicazione del teorema di Lindeberg-Lévy. Tuttavia, essa è superflua. È ciò che ci proponiamo di dimostrare nella presente sezione (si veda (2.6)) con l'aiuto di alcuni lemmi preliminari, non privi di un qualche interesse autonomo.

Ci metteremo nelle ipotesi di (1.1). Inoltre, ogni volta che T sia una variabile aleatoria discreta, a valori in $\mathbb{N} \cup \{\infty\}$, indicheremo con S_T la variabile aleatoria reale che è nulla sull'evento $\{T = \infty\}$ e che, per ciascun intero positivo n, coincide con S_n sull'evento $\{T = n\}$.

(2.1) Lemma. *Nelle ipotesi di (1.1), sia T un tempo d'arresto, quasi certamente finito, relativo alla filtrazione naturale del processo $(S_n)_{n\geq 0}$. Allora il processo $(S_{T+n} - S_T)_{n\geq 0}$ è ancora una passeggiata aleatoria su $\mathbb{R}$, uscente dall'origine e avente ν come legge del singolo passo.*

(2.2) Lemma. *Nelle ipotesi di (1.1), si consideri il tempo d'arresto T (relativo alla filtrazione naturale del processo $(S_n)_{n\geq 0}$) così definito:*

$$T(\omega) = \inf\{n \in \mathbb{N}^* : S_n(\omega) \geq 0\}. \tag{2.3}$$

Le condizioni che seguono sono tra loro equivalenti:

(a) Il tempo d'arresto T è quasi certamente finito.

(b) Si ha $v = \infty$.

Dimostrazione. L'implicazione (b) $\Rightarrow$ (a) è ovvia. Per provare l'implicazione inversa, supponiamo verificata la condizione (a). Grazie a (1.1), basterà dimostrare che V è quasi certamente positiva. A questo scopo, definiamo per induzione la successione $(T_j)_{j\geq 1}$ di variabili aleatorie discrete, a valori in $\mathbb{N}^* \cup \{\infty\}$, mediante le due regole seguenti: $T_1 = T$ e

$$T_{j+1}(\omega) = \inf\left\{n \in \mathbb{N}^* : S_{T_1+\cdots+T_j+n}(\omega) - S_{T_1+\cdots+T_j}(\omega) \geq 0\right\}. \tag{2.4}$$

Applicando il lemma precedente, si riconosce facilmente (per induzione) che, per ciascun intero j strettamente positivo, il processo

$$\left(S_{T_1+\cdots+T_j+n} - S_{T_1+\cdots+T_j}\right)_{n\geq 0}$$

è una passeggiata aleatoria uscente dall'origine e avente ν come legge del singolo passo. Allora, per ogni intero j strettamente positivo, confrontando (2.4) con (2.3), si vede che la variabile aleatoria T_{j+1} è isonoma a T, e quindi è anch'essa quasi certamente finita. Di conseguenza, è quasi certo l'evento H così definito:

$$H = \bigcap_{j\geq 1}\{T_j < \infty\}.$$

D'altra parte, le variabili aleatorie $S_{T_1+\cdots+T_j}$ sono tutte positive e, in ciascun punto di H, la successione $(T_1 + \cdots + T_j)_{j\geq 1}$ coincide con una successione strettamente crescente d'interi. La relazione (1.2) mostra dunque che la variabile aleatoria V è quasi certamente positiva, e ciò conclude la dimostrazione. □

(2.5) Lemma. *Nelle ipotesi del Lemma (2.2), si supponga per giunta che la legge ν sia centrata. Allora il tempo d'arresto T non è integrabile.*

Dimostrazione. Ragionando per assurdo, supponiamo che T sia integrabile, e denotiamo con M la martingala ottenuta arrestando $(S_n)_{n\geq 0}$ all'istante T. Allora la relazione

$$|M_n| = |S_{T\wedge n}| \leq \sum_{j\geq 0} I_{\{j<T\wedge n\}}|X_{j+1}|$$

mostra che la martingala M è dominata dalla variabile aleatoria Z così definita:

$$Z = \sum_{j\geq 0} I_{\{j<T\}}|X_{j+1}|.$$

La speranza di questa variabile aleatoria è finita, essendo data da

$$\mathrm{E}\,[Z] = \sum_{j\geq 0} P\{T > j\}\,\mathrm{E}\,[X_{j+1}] = \mathrm{E}\,[X_1]\,\mathrm{E}\,[T]\,.$$

Ne segue che la martingala M è uniformemente integrabile, ossia chiusa. D'altra parte, sull'evento quasi certo $\{T < \infty\}$, la successione $(M_n)_{n\geq 0}$ converge puntualmente verso S_T. (Addirittura, in ciascun punto ω di $\{T < \infty\}$, si ha definitivamente $M_n(\omega) = S_T(\omega)$.) Dunque la martingala M è chiusa dalla variabile aleatoria S_T. Essendo questa positiva, ne segue che ciascuna delle variabili aleatorie M_n è quasi certamente positiva. Tale è, in particolare, la variabile aleatoria M_1 (identica a X_1). Poiché questa è anche centrata, ne segue che essa è trascurabile, contrariamente all'ipotesi che la legge ν sia diversa da ϵ_0. Il lemma è così dimostrato. □

(2.6) Teorema. *Nelle ipotesi di (1.1), si supponga per giunta che la legge ν sia centrata. Si ha allora $v = \infty$, $u = -\infty$.*

Dimostrazione. Basterà provare che si ha $v = \infty$. (Infatti, applicando questo risultato alla passeggiata aleatoria $(-S_n)_{n\geq 0}$, se ne dedurrà $u = -\infty$.) Dunque, grazie al Lemma (2.2), tutto è ridotto a provare che il tempo d'arresto T definito da (2.3) è quasi certamente finito. A questo scopo, cominciamo con l'osservare che, grazie

al Lemma (2.5) (applicato alla passeggiata aleatoria $(-S_n)$), il tempo d'arresto R definito da

$$R(\omega) = \inf\{n \in \mathbb{N}^* : S_n(\omega) \leq 0\} \tag{2.7}$$

non è integrabile. Consideriamo poi, per ogni intero positivo n, l'evento C_n costituito dai punti di Ω in cui S_n è maggiore di ogni S_j con $j \neq n$. L'evento C_n è l'intersezione dei due eventi A_n, B_n così definiti:

$$A_n = \bigcap_{0 \leq j < n} \{S_n - S_j > 0\}, \qquad B_n = \bigcap_{h \geq 1} \{S_{n+h} - S_n < 0\}.$$

Poiché questi sono tra loro indipendenti, si ha

$$P(C_n) = P(A_n) \cdot P(B_n). \tag{2.8}$$

Osserviamo che gli eventi B_n sono equiprobabili. Più precisamente: per ogni n, si ha

$$P(B_n) = P(B_0) = P\big(\bigcap_{h \geq 1} \{S_h < 0\}\big) = P\{T = \infty\}. \tag{2.9}$$

Inoltre l'evento A_n si può scrivere nella forma

$$A_n = \big\{X_1 + \cdots + X_n > 0,\ X_2 + \cdots + X_{n-1} > 0,\ \ldots, X_n > 0\big\}, \tag{2.10}$$

mentre dalla definizione (2.7) risulta

$$\{R > n\} = \big\{X_1 > 0,\ X_1 + X_2 > 0,\ \ldots,\ X_1 + \cdots + X_n > 0\big\}.$$

Confrontando questa relazione con la (2.10), si trova

$$P(A_n) = P\{R > n\}. \tag{2.11}$$

Sfruttando allora le relazioni (2.8), (2.9) e (2.11) (e il fatto che gli eventi C_n sono a due a due disgiunti), si ottiene

$$\begin{aligned} 1 \geq \textstyle\sum_{n \geq 0} P(C_n) &= \textstyle\sum_{n \geq 0} P(A_n)\, P(B_n) \\ &= P\{T = \infty\} \textstyle\sum_{n \geq 0} P\{R > n\} \\ &= P\{T = \infty\}\, \mathrm{E}\,[R]. \end{aligned}$$

Di qui, essendo $\mathrm{E}\,[R] = \infty$, si deduce $P\{T = \infty\} = 0$, e ciò conclude la dimostrazione. □

Una classe di martingale non uniformemente integrabili *

1. Teorema. *Sia $(X_n)_{n\geq 1}$ una successione di variabili aleatorie reali positive indipendenti, dotate di una medesima legge non degenere, con baricentro eguale a 1.*

Allora la martingala $Y = [Y_n]_{n\geq 0}$ definita da $Y_n = \prod_{i=1}^{n} X_i$ converge quasi certamente verso 0 (e quindi non è chiusa).

Alla dimostrazione di questo teorema conviene premettere il seguente lemma, che discende immediatamente dal teorema di Hewitt e Savage (o, più semplicemente, dal paradosso di Borel).

2. Lemma. *Sia $(X_n)_{n\geq 1}$ una successione di variabili aleatorie reali indipendenti, dotate di una medesima legge, diversa dalla legge della costante 0.*

Si ponga $S_n = X_1 + \cdots + X_n$. Allora è quasi certo l'evento

$$\{\sup_n |S_n| = \infty\} = \{\inf_n S_n = -\infty\} \cup \{\sup_n S_n = \infty\}.$$

Impiegando il lemma ora enunciato, il Teorema 1 si può così dimostrare. La tesi è immediata se l'evento $\{X_1 = 0\}$ non è trascurabile. In questo caso, infatti, l'evento $\bigcup_{n\geq 1}\{X_n = 0\}$ è quasi certo e, su di esso, la successione (Y_n) è definitivamente nulla. Si può dunque supporre, senza ledere la generalità, che le X_n siano dappertutto strettamente positive. È lecito allora considerare, per ogni n, la variabile aleatoria reale

$$L_n = \log Y_n = \textstyle\sum_{i=1}^{n} \log X_i\,.$$

Poiché Y è una martingala positiva, la successione (Y_n) converge quasi certamente verso una variabile aleatoria reale positiva. D'altra parte, l'insieme dove (Y_n) converge verso un limite maggiore di zero è contenuto nell'insieme dove (L_n) converge, e quest'ultimo insieme è trascurabile, come risulta dal lemma precedente (applicato alla successione $(\log X_n)$). Dunque (Y_n) converge quasi certamente verso zero. □

Nel teorema seguente considereremo, su uno spazio canonico, una martingala Y del tipo di quella costruita nel teorema precedente, e mostreremo che essa gode di un'importante proprietà: per ogni tempo d'arresto T, che sia *dappertutto finito*, la martingala $Y^{|T}$ è chiusa.

3. Teorema. *Assegnata, sullo spazio misurabile $(E, \mathcal{E})$, una legge μ, si consideri lo schema delle prove indipendenti*

$$(\Omega, \mathcal{A}, P), \quad (X_n)_{n\geq 1}$$

associato a una successione di copie di μ, e si denoti con $\mathcal{F} = (\mathcal{F}_n)_{n\geq 0}$ la filtrazione definita da

$$\mathcal{F}_n = \mathcal{T}(X_i : 1 \leq i \leq n).$$

Sia poi $\nu = f \cdot \mu$ una legge di base μ, e si denoti con Q la misura di probabilità $\nu \otimes \nu \otimes \cdots$ (prodotto di una successione di copie di ν).
Si consideri infine la martingala $Y = [Y_n]_{n\geq 0}$ così definita: $Y_n = \prod_{i=1}^{n} f \circ X_i$. Allora, per ogni tempo d'arresto T, posto

$$Y_T = \sum_{n\in\mathbb{N}} Y_n I_{\{T=n\}},$$

valgono le due proprietà seguenti:

(a) La misura $Y_T \cdot P$ coincide con $I_{\{T<\infty\}} \cdot Q$ sulla tribù $\mathcal{F}_T$ (costituita dagli elementi A di $\mathcal{A}$ tali che, per ogni n, l'insieme $A \cap \{T = n\}$ appartenga a $\mathcal{F}_n$).

(b) Se T è finito, la martingala $Y^{|T}$ è chiusa.

Dimostrazione. (a) Supponiamo dapprima che T coincida con la costante n (nel qual caso la tribù $\mathcal{F}_T$ si riduce semplicemente a $\mathcal{F}_n$). In questo caso, basta provare che la misura $Y_n \cdot P$ coincide con Q su ogni insieme H della forma

$$H = \bigcap_{i=1}^{n} \{X_i \in A_i\}$$

(con $(A_i)_{1\leq i\leq n}$ arbitraria n–upla di elementi di $\mathcal{E}$). Ma ciò è immediato:

$$\int_H Y_n \, \mathrm{d}P = \prod_{i=1}^{n} \int_{A_i} f \, \mathrm{d}\mu = \prod_{i=1}^{n} \nu(A_i) = Q(H).$$

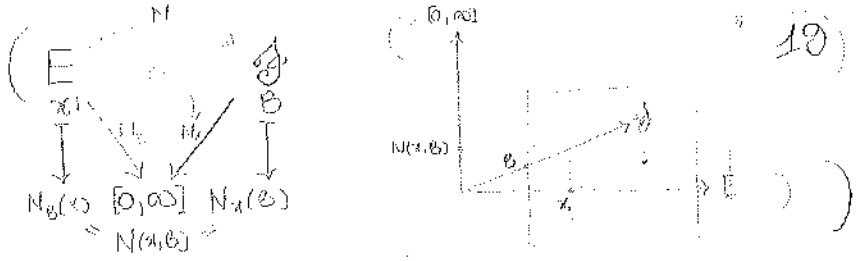
Passando al caso generale, consideriamo ora un qualsiasi elemento A di $\mathcal{F}_T$. Si ha allora, per ogni n,

$$\int_{A\cap\{T=n\}} Y_T \, \mathrm{d}P = \int_{A\cap\{T=n\}} Y_n \, \mathrm{d}P = Q\big(A \cap \{T = n\}\big)$$

(dove l'ultima eguaglianza si ottiene applicando il caso particolare già trattato). Sommando su n, si ottiene la desiderata eguaglianza:

$$\int_A Y_T \, \mathrm{d}P = Q\big(A \cap \{T < \infty\}\big).$$

(b) Supponiamo ora T finito. La successione $\big(Y_{T\wedge n}\big)_{n\geq 0}$ è una successione di densità di probabilità (rispetto a P) e, grazie alla finitezza di T, converge puntualmente verso la variabile aleatoria Y_T, la quale, in virtù di (a), è anch'essa una densità di probabilità. La convergenza ha dunque luogo in L^1 (teorema di Scheffé), e ciò prova che la martingala $Y^{|T}$ è chiusa. □



Nuclei e leggi condizionali

1. Concetto di nucleo

Siano dati due spazi misurabili $(E,\mathcal{E})$, $(F,\mathcal{F})$ e un'applicazione N di $E\times\mathcal{F}$ in $[0,\infty]$, tale che, per ciascun elemento x di E, la funzione $N(x,\cdot\,)$ (cioè la funzione che a ciascun elemento B di $\mathcal{F}$ associa il numero $N(x,B)$) sia una misura sulla tribù $\mathcal{F}$. È chiaro che l'applicazione N è individuata dalla famiglia di misure

$$\big(N(x,\cdot\,)\big)_{x\in E} \tag{1.1}$$

(con la quale può addirittura essere identificata, senza pericolo di ambiguità).

In queste condizioni, se g è una funzione positiva e misurabile su $(F,\mathcal{F})$, denoteremo con Ng la funzione positiva così definita su E:

$$Ng(x)=\langle N(x,\cdot\,),g\rangle=\int N(x,\mathrm{d}y)\,g(y). \tag{1.2}$$

Quando g sia la funzione indicatrice di un elemento B di $\mathcal{F}$, avremo dunque, in particolare,

$$NI_B(x)=N(x,B). \tag{1.3}$$

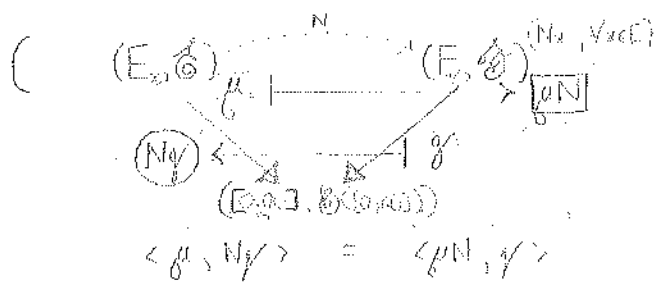
Osserviamo che, per ogni funzione g positiva e misurabile su $(F,\mathcal{F})$ ed ogni successione $(g_n)_{n\geq1}$ di funzioni positive e misurabili su $(F,\mathcal{F})$, valgono le due implicazioni seguenti:

$$g_n\uparrow g\Rightarrow Ng_n\uparrow Ng,\qquad g=\textstyle\sum_{n\geq1}g_n\Rightarrow Ng=\sum_{n\geq1}Ng_n. \tag{1.4}$$

(1.5) Definizione. Nelle ipotesi sopra precisate, si dice che N è un *nucleo* relativo alla coppia di spazi misurabili $(E,\mathcal{E}),(F,\mathcal{F})$ (o anche un nucleo *da* $(E,\mathcal{E})$ *a* $(F,\mathcal{F})$) se, per ogni funzione g positiva e misurabile su $(F,\mathcal{F})$, la funzione positiva Ng definita su E da (1.2) è misurabile su $(E,\mathcal{E})$.

Un nucleo da $(E,\mathcal{E})$ a $(E,\mathcal{E})$ si chiamerà anche un nucleo relativo allo spazio misurabile $(E,\mathcal{E})$.

(1.6) Osservazione. Nelle ipotesi della definizione precedente, è immediato riconoscere che, affinché N sia un nucleo, è (necessario e) sufficiente che, per ogni elemento B di $\mathcal{F}$, la funzione NI_B definita da (1.3) sia misurabile su $(E,\mathcal{E})$. Infatti, supposto che questa condizione sia soddisfatta, si riconosce innanzitutto facilmente che, per ogni funzione g positiva e semplice su $(F,\mathcal{F})$, la funzione Ng è misurabile su $(E,\mathcal{E})$. In seguito si estende questo risultato al caso di un'arbitraria funzione positiva e misurabile g esprimendo g come inviluppo superiore di una successione crescente $(g_n)_{n\geq1}$ di funzioni positive e semplici, e riconducendosi al caso precedente con l'ausilio della prima delle implicazioni (1.4).



(1.7) Notazione. Siano N un nucleo da $(E,\mathcal{E})$ a $(F,\mathcal{F})$ e μ una misura su $(E,\mathcal{E})$. Denoteremo allora con μN la funzione che ad ogni elemento B di $\mathcal{F}$ associa il numero $\mu N(B)$ così definito:

$$\mu N(B) = \int \mu(\mathrm{d}x)\, N(x,B) = \langle \mu,\, NI_B \rangle. \tag{1.8}$$

(1.9) Proposizione. *Siano N un nucleo da $(E,\mathcal{E})$ a $(F,\mathcal{F})$ e μ una misura su $(E,\mathcal{E})$. Valgono allora le due affermazioni seguenti:*

(a) La funzione μN definita in (1.7) è una misura su $(F,\mathcal{F})$.

(b) Per ogni funzione g positiva e misurabile su $(F,\mathcal{F})$, sussiste la seguente "relazione di dualità":

$$\langle \mu N,\, g \rangle = \langle \mu,\, Ng \rangle. \tag{1.10}$$

Dimostrazione. Per ogni elemento B di $\mathcal{F}$ e ogni successione $(B_n)_{n\geq 1}$ di elementi di $\mathcal{F}$, a due a due disgiunti, la cui unione sia eguale a B, l'ovvia relazione $I_B = \sum_{n\geq 1} I_{B_n}$ implica

$$NI_B = \sum_{n\geq 1} NI_{B_n}$$

e quindi $\langle \mu,\, NI_B \rangle = \sum_{n\geq 1} \langle \mu,\, NI_{B_n} \rangle$, ossia $\mu N(B) = \sum_{n\geq 1} \mu N(B_n)$. Ciò prova che μN è una misura su $\mathcal{F}$, inoltre $\mu N \geq 0$ e $\mu N(\emptyset) = 0$.

Proviamo ora l'affermazione (b). Essa è immediata nel caso particolare in cui g sia la funzione indicatrice di un elemento B di $\mathcal{F}$. In questo caso, essa discende infatti direttamente dalla relazione (1.8) che definisce $\mu N(B)$. Immediata è poi l'estensione al caso in cui g sia semplice. Infine, nel caso generale, dopo aver espresso g come inviluppo superiore di una successione crescente $(g_n)_{n\geq 1}$ di funzioni positive e semplici, basta ricondursi al caso precedente servendosi della proprietà (1.4) e del teorema di Beppo Levi. □

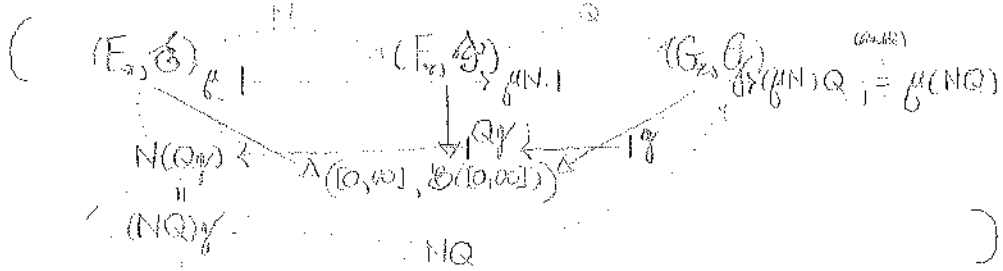
Come risulta dalla proposizione precedente, un nucleo N da $(E,\mathcal{E})$ a $(F,\mathcal{F})$ induce due distinte operazioni:

- un'operazione $\mu \mapsto \mu N$ che trasforma ogni misura μ su $\mathcal{E}$ in una misura μN su $\mathcal{F}$;
- un'operazione $g \mapsto Ng$ che trasforma ogni funzione g positiva e misurabile su $(F,\mathcal{F})$ in una funzione Ng positiva e misurabile su $(E,\mathcal{E})$.

Queste due operazioni sono tra loro legate dalla fondamentale relazione di dualità (1.10). Si osservi che la seconda operazione individua completamente il nucleo, grazie alla relazione $N(x,B) = NI_B(x)$. Anche la prima operazione individua il nucleo: si ha infatti $N(x,\cdot) = \epsilon_x N$, dove ϵ_x designa (per ogni x di E) la misura così definita su $\mathcal{E}$:

$$\epsilon_x(A) = I_A(x) \quad \text{per ogni } A \text{ di } \mathcal{E}$$

(*misura di Dirac* definita da una massa unitaria concentrata nel punto x).



Ci si può domandare come si caratterizzino le "trasformazioni sulle funzioni" indotte dai nuclei. La risposta è semplice: un'applicazione L, che ad ogni funzione g positiva e misurabile su $(F, \mathcal{F})$ associ una funzione $L(g)$ positiva e misurabile su $(E, \mathcal{E})$, è indotta da un nucleo se (e solo se,) possiede le proprietà seguenti:

(a) $L(\alpha g) = \alpha\, L(g)$ per ogni g e ogni scalare positivo α.

(b) $L(\sum_{n\geq 1} g_n) = \sum_{n\geq 1} L(g_n)$ per ogni successione $(g_n)_{n\geq 1}$.

È facile infatti verificare che, se L possiede queste proprietà, e se, per ogni elemento x di E, si considera sulla tribù $\mathcal{F}$ la funzione $B \mapsto (L(I_B))(x)$, allora questa funzione è una misura su $\mathcal{F}$, e la famiglia formata da queste misure è identificabile con un nucleo N verificante la relazione $Ng = L(g)$ per ogni funzione g positiva e misurabile su $(F, \mathcal{F})$.

Siano ora assegnati un nucleo N da $(E, \mathcal{E})$ a $(F, \mathcal{F})$ e un nucleo Q da $(F, \mathcal{F})$ a un ulteriore spazio misurabile $(G, \mathcal{G})$. È facile riconoscere che l'applicazione $g \mapsto N(Qg)$ (ottenuta componendo la trasformazione indotta da Q con quella indotta da N) possiede ancora le proprietà (a), (b), sicché è a sua volta indotta da un nucleo (relativo alla coppia di spazi $(E, \mathcal{E})$, $(G, \mathcal{G})$). Il nucleo così caratterizzato si denota con NQ e si chiama il *nucleo composto* ottenuto componendo i due nuclei N, Q.

Per ogni misura μ su $(E, \mathcal{E})$, la misura $\mu(NQ)$ (trasformata di μ mediante il nucleo composto NQ) coincide con $(\mu N)Q$ (trasformata di μN mediante Q); essa potrà dunque essere denotata, senza pericolo di ambiguità, col simbolo μNQ.

2. Nuclei markoviani

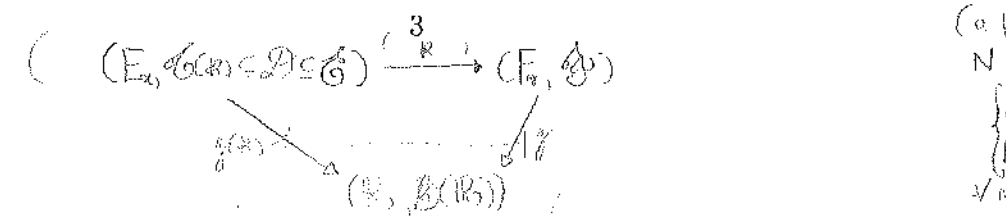
Un nucleo N da $(E, \mathcal{E})$ a $(F, \mathcal{F})$ si dice *markoviano* se ciascuna delle misure $N(x, \cdot)$ è normalizzata, cioè se la trasformata $N1$ della costante 1 su F mediante il nucleo N coincide con la costante 1 su E: $\forall x \in E,\ N_x(F) = 1 \Leftrightarrow N1 = 1_E$.

Si supponga il nucleo N markoviano. Allora, per ogni misura μ su $\mathcal{E}$, si ha

$$\langle \mu N,\, 1\rangle = \langle \mu,\, N1\rangle = \langle \mu,\, 1\rangle,$$

sicché la misura μN ha la stessa massa totale di μ. Inoltre, per ogni funzione g limitata e misurabile su $(F, \mathcal{F})$, si può considerare la funzione Ng definita dalla formula (1.2) esattamente come nel caso di una g positiva e misurabile. Si verifica immediatamente che la funzione Ng così definita (identica alla differenza $Ng^+ - Ng^-$) è limitata e misurabile su $(E, \mathcal{E})$. Inoltre essa verifica la relazione di dualità (1.10) per ogni misura μ su $(E, \mathcal{E})$ che abbia massa totale finita.

(2.1) Definizione. Dati due spazi misurabili $(E, \mathcal{E})$, $(F, \mathcal{F})$, un'applicazione misurabile k di $(E, \mathcal{E})$ in $(F, \mathcal{F})$ e una tribù $\mathcal{D}$ su E, con $\mathcal{T}(k) \subset \mathcal{D} \subset \mathcal{E}$, si vede facilmente che esiste un nucleo markoviano K, da $(E, \mathcal{D})$ a $(F, \mathcal{F})$, che opera così sulle funzioni: per ogni funzione g limitata e misurabile su $(F, \mathcal{F})$, la trasformata di g mediante K è data da $Kg = g \circ k$. Chiameremo K il nucleo, *da* $(E, \mathcal{D})$ *a* $(F, \mathcal{F})$, *associato all'operazione di composizione con* k (o anche, in maniera più concisa, *associato a* k). In particolare, il nucleo, relativo allo spazio $(E, \mathcal{E})$, associato all'applicazione identica di E sarà detto il *nucleo identità* relativo allo spazio $(E, \mathcal{E})$. Esso opera sulle funzioni (risp. sulle misure) lasciando invariata ogni funzione (risp. ogni misura).



(2.2) Osservazione. Nelle ipotesi della Definizione (2.1), si consideri il nucleo K, da $(E, \mathcal{D})$ a $(F, \mathcal{F})$, associato a k. Allora, per ogni misura di probabilità μ su $(E, \mathcal{D})$ ed ogni funzione g limitata e misurabile su $(F, \mathcal{F})$, la relazione di dualità

$$\langle \mu K, g \rangle = \langle \mu, Kg \rangle, \quad \text{ossia} \quad \int g \, \mathrm{d}\mu K = \int g \circ k \, \mathrm{d}\mu,$$

mostra che μK è l'immagine di μ mediante k. Per questa ragione, K si dice anche il nucleo, *da* $(E, \mathcal{D})$ *a* (F, F), *associato all'operazione di immagine mediante* k.

(2.3) Proposizione. *Siano dati due spazi misurabili* $(E, \mathcal{E})$, $(F, \mathcal{F})$ *e un'applicazione* N *di* $E \times \mathcal{F}$ *in* $[0, 1]$, *tale che, per ciascun elemento* x *di* E, *la funzione* $N(x, \cdot)$ *sia una misura normalizzata su* $\mathcal{F}$. *Sia poi assegnato, per la tribù* $\mathcal{F}$, *un sistema di generatori* $\mathcal{J}$, *stabile per l'intersezione binaria. Si supponga che, per ogni elemento* B *di* $\mathcal{J}$, *la funzione* NI_B *(definita da (1.3)) risulti misurabile su* $(E, \mathcal{E})$.

Allora N *è un nucleo markoviano da* $(E, \mathcal{E})$ *a* $(F, \mathcal{F})$.

Dimostrazione. Si denoti con $\mathcal{L}$ la classe costituita dalle funzioni g, limitate e misurabili su $(F, \mathcal{F})$, tali che la funzione Ng definita da (1.2) risulti misurabile su $(E, \mathcal{E})$. Allora $\mathcal{L}$ è uno spazio vettoriale monotono contenente le costanti e le funzioni indicatrici degli insiemi appartenenti a $\mathcal{J}$. Ne segue, grazie al teorema delle classi monotone, che $\mathcal{L}$ contiene tutte le funzioni limitate e misurabili su $(F, \mathcal{F})$, dunque in particolare tutte le funzioni indicatrici di elementi della tribù $\mathcal{F}$. Ciò basta, grazie all'Osservazione (1.6), per concludere che N è un nucleo.

(2.4) Notazione. Sia N un nucleo markoviano da $(E, \mathcal{E})$ a $(F, \mathcal{F})$. Denoteremo allora con $\widetilde{N}$ l'applicazione di $E \times (\mathcal{E} \otimes \mathcal{F})$ in $[0, 1]$ così caratterizzata: per ciascun elemento x di E, la funzione $\widetilde{N}(x, \cdot)$ è la misura normalizzata su $\mathcal{E} \otimes \mathcal{F}$ ottenuta come immagine di $N(x, \cdot)$ mediante l'applicazione $y \mapsto (x, y)$ di $(F, \mathcal{F})$ in $(E \times F, \mathcal{E} \otimes \mathcal{F})$, ossia è la misura che ad ogni elemento C di $\mathcal{E} \otimes \mathcal{F}$ associa il numero $N(x, C(x))$ (dove $C(x)$ denota la sezione $\{y \in F : (x, y) \in C\}$).

(2.5) Proposizione. *Nelle ipotesi di (2.4),* $\widetilde{N}$ *è un nucleo markoviano da* $(E, \mathcal{E})$ *a* $(E \times F, \mathcal{E} \otimes \mathcal{F})$.

Dimostrazione. Sia h una funzione limitata e misurabile su $(E \times F, \mathcal{E} \otimes \mathcal{F})$. Allora, per ogni elemento x di E, si ha

$$\widetilde{N}h(x) = \int N(x, \mathrm{d}y)\, h(x, y). \tag{2.6}$$

In particolare, quando sia $h = f \otimes g$, ossia $h(x, y) = f(x)\, g(y)$ (con f limitata e misurabile su $(E, \mathcal{E})$ e g limitata e misurabile su $(F, \mathcal{F})$), la relazione (2.6) diventa

$$\widetilde{N} f \otimes g\,(x) = \int N(x, \mathrm{d}y)\, f(x)\, g(y) = f(x) \int N(x, \mathrm{d}y)\, g(y) = f(x)\, Ng\,(x).$$

Si può dunque scrivere $\widetilde{N}\, f \otimes g = f\, Ng$. Particolarizzando ulteriormente col prendere $f = I_A$ e $g = I_B$, si trova $\widetilde{N} I_{A \times B} = I_A\, NI_B$. Si vede così che, per ogni "rettangolo"

della forma $A \times B$, con $A \in \mathcal{E}$ e $B \in \mathcal{F}$, la funzione $\widetilde{N} I_{A\times B}$ è misurabile su $(E, \mathcal{E})$. Poiché questi "rettangoli" formano una base per la tribù $\mathcal{E} \otimes \mathcal{F}$, ciò basta, grazie alla Proposizione (2.3), per concludere che $\widetilde{N}$ è un nucleo markoviano. □

(2.7) Osservazione. Come applicazione della proposizione precedente, si può ottenere l'esistenza del prodotto di due misure di probabilità e il relativo teorema di Fubini. Precisamente, assegnati due spazi probabilizzati $(E, \mathcal{E}, \mu)$, $(F, \mathcal{F}, \nu)$, si denoti con N il nucleo da $(E, \mathcal{E})$ a $(F, \mathcal{F})$ così caratterizzato: per ogni elemento x di E, la misura $N(x, \cdot)$ coincide con ν. Si consideri poi il nucleo $\widetilde{N}$ definito in (2.4). Allora si riconosce immediatamente che la misura $\mu\widetilde{N}$ verifica la relazione

$$\mu\widetilde{N}\,(A \times B) = \mu(A)\,\nu(B) \quad \text{per } A \in \mathcal{E} \text{ e } B \in \mathcal{F},$$

ossia è la misura prodotto $\mu \otimes \nu$. Inoltre, per ogni funzione h positiva (o limitata) e misurabile sullo spazio prodotto $(E \times F, \mathcal{E} \otimes \mathcal{F})$, la relazione di dualità

$$\langle \mu\widetilde{N},\, h\rangle = \langle \mu,\, \widetilde{N}h\rangle$$

non è altro che la formula del teorema di Fubini:

$$\langle \mu \otimes \nu,\, h\rangle = \int \mu(\mathrm{d}x) \int \nu(\mathrm{d}y)\, h(x, y).$$

(2.8) Osservazione. Dati un arbitrario nucleo markoviano N, relativo alla coppia di spazi misurabili $(E, \mathcal{E})$, $(F, \mathcal{F})$, e una misura di probabilità μ su $(E, \mathcal{E})$, si può considerare la misura $\mu\widetilde{N}$ (trasformata di μ mediante il nucleo $\widetilde{N}$ definito in (2.4)). Si tratta di una misura su $(E \times F, \mathcal{E} \otimes \mathcal{F})$, caratterizzata dalla relazione

$$\mu\widetilde{N}\,(A \times B) = \int_A \mu(\mathrm{d}x)\, N(x, B) \quad \text{per } A \in \mathcal{E} \text{ e } B \in \mathcal{F}. \tag{2.9}$$

Essa si riduce alla misura prodotto $\mu \otimes \nu$ nel caso particolare considerato nell'osservazione precedente, ossia nel caso in cui il nucleo N sia il nucleo "costante" caratterizzato da $N(x, \cdot) = \nu$ per ogni x. Per questa ragione, nel caso generale, la misura $\mu\widetilde{N}$ sarà talvolta denotata anche col simbolo $\mu \otimes N$.

(2.10) Osservazione. Nelle ipotesi di (2.8), denotiamo con h, k le proiezioni canoniche di $E \times F$ sul primo e sul secondo fattore. Allora la relazione (2.9) fornisce, in particolare,

$$\mu\widetilde{N}\{h \in A\} = \mu\widetilde{N}(A \times F) = \int_A \mu(\mathrm{d}x)\, N(x, F) = \mu(A) \quad \text{per } A \in \mathcal{E},$$

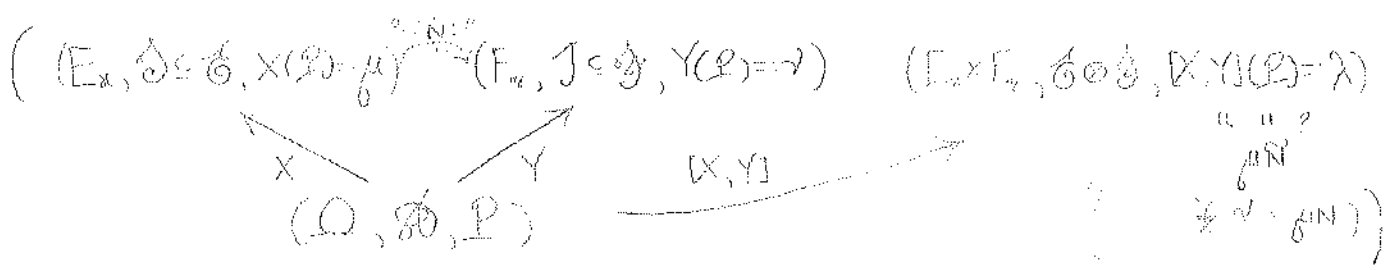
$$\mu\widetilde{N}\{k \in B\} = \mu\widetilde{N}(E \times B) = \int_E \mu(\mathrm{d}x)\, N(x, B) = \mu N(B) \quad \text{per } B \in \mathcal{F}.$$

A parole: l'immagine della misura $\mu\widetilde{N}$ mediante h (risp. k) coincide con μ (risp. μN). Denotando con H il nucleo, da $(E \times F, \mathcal{E} \otimes \mathcal{F})$ a $(E, \mathcal{E})$, associato a h, e con K il nucleo, da $(E \times F, \mathcal{E} \otimes \mathcal{F})$ a $(F, \mathcal{F})$, associato a k, si hanno dunque le eguaglianze tra misure

$$\mu\widetilde{N}H = \mu, \qquad \mu\widetilde{N}K = \mu N,$$

e quindi, denotando con J il nucleo identità relativo a $(E, \mathcal{E})$, le seguenti eguaglianze tra nuclei:

$$\boxed{\widetilde{N}H = J, \qquad \widetilde{N}K = N.}$$



3. Concetto di legge condizionale

(3.1) Ipotesi e notazioni. Su uno spazio probabilizzato $(\Omega, \mathcal{A}, P)$ siano X, Y due variabili aleatorie, a valori in due arbitrari spazi misurabili $(E, \mathcal{E}), (F, \mathcal{F})$. Denoteremo con μ la legge di X e con ν quella di Y.

È ben noto che, se si particolarizzano le ipotesi (3.1) supponendo che Y sia indipendente da X, allora si ha $[X, Y](P) = \mu \otimes \nu$, ossia, per ogni funzione h limitata e misurabile su $(E \times F, \mathcal{E} \otimes \mathcal{F})$, ha luogo la relazione

$$\int h \circ [X, Y] \, \mathrm{d}P = \int \mu(\mathrm{d}x) \int \nu(\mathrm{d}y) \, h(x, y). \tag{3.2}$$

Ci si può domandare se, nel caso generale, si possa assicurare l'esistenza di un nucleo markoviano N, da $(E, \mathcal{E})$ a $(F, \mathcal{F})$, tale che si abbia (con la notazione introdotta in (2.8))

$$[X, Y](P) = \mu \otimes N, \tag{3.3}$$

cioè tale che, per ogni funzione h limitata e misurabile su $(E \times F, \mathcal{E} \otimes \mathcal{F})$, sussista la formula seguente (analoga alla (3.2)):

$$\int h \circ [X, Y] \, \mathrm{d}P = \int \mu(\mathrm{d}x) \int N(x, \mathrm{d}y) \, h(x, y). \tag{3.4}$$

(3.5) Definizione. Nelle ipotesi (3.1), un nucleo markoviano N da $(E, \mathcal{E})$ a $(F, \mathcal{F})$ si dice una *versione della legge condizionale* di Y rispetto a X se esso verifica la relazione (3.3).

(3.6) Osservazione. Nelle ipotesi (3.1), siano $\mathcal{I}$ una base per la tribù $\mathcal{E}$ e $\mathcal{J}$ una base per la tribù $\mathcal{F}$. Allora, com'è noto, la classe

$$\{A \times B : A \in \mathcal{I}, \, B \in \mathcal{J}\} \tag{3.7}$$

è una base per la tribù $\mathcal{E} \otimes \mathcal{F}$. Pertanto, affinché un nucleo markoviano N da $(E, \mathcal{E})$ a $(F, \mathcal{F})$ sia una versione della legge condizionale di Y rispetto a X, è sufficiente che la coincidenza tra le due misure $[X, Y](P)$ e $\mu \otimes N$ abbia luogo sugli insiemi della base (3.7), cioè che si abbia

$$P\{X \in A, \, Y \in B\} = \int_A \mu(\mathrm{d}x) \, N(x, B) \quad \text{per } A \in \mathcal{I} \text{ e } B \in \mathcal{J}. \tag{3.8}$$

(3.9) Proposizione. *Nelle ipotesi (3.1), sia N una versione della legge condizionale di Y rispetto a X. Allora la legge di Y è eguale a μN.*

Dimostrazione. Per ipotesi, la legge del blocco $[X, Y]$ è eguale a $\mu \otimes N$, ossia a $\mu\tilde{N}$. Dunque la legge di Y è eguale all'immagine di $\mu\tilde{N}$ mediante la proiezione canonica di $E \times F$ sul secondo fattore, ossia alla misura μN (si veda (2.10)). □

(3.10) Osservazione. È ben noto che, comunque si assegnino due spazi misurabili $(E, \mathcal{E})$, $(F, \mathcal{F})$, una misura di probabilità μ su $(E, \mathcal{E})$ e una misura di probabilità ν su $(F, \mathcal{F})$, è sempre possibile costruire uno spazio probabilizzato e, su di esso, una coppia di variabili aleatorie indipendenti, a valori in $(E, \mathcal{E})$ e in $(F, \mathcal{F})$ rispettivamente, la prima di legge μ, la seconda di legge ν. Ricordiamo che un modo "canonico" per risolvere il problema consiste nel considerare lo spazio probabilizzato

$$(E \times F,\ \mathcal{E} \otimes \mathcal{F},\ \mu \otimes \nu) \tag{3.11}$$

e, su di esso, la coppia (h, k) delle proiezioni canoniche di $E \times F$ sul primo e sul secondo fattore. (La terna formata dallo spazio (3.11) e dalle due proiezioni canoniche h, k si chiama tradizionalmente lo *schema delle prove indipendenti* associato alla coppia (μ, ν).)

Una naturale estensione del problema precedente consiste nel modificarne i dati, sostituendo la misura ν con un nucleo markoviano N da $(E, \mathcal{E})$ a $(F, \mathcal{F})$ e chiedendo, in corrispondenza, di costruire uno spazio probabilizzato e, su di esso, una coppia di variabili aleatorie, a valori in $(E, \mathcal{E})$ e in $(F, \mathcal{F})$ rispettivamente, in modo tale che μ sia la legge della prima e che N sia una versione della legge condizionale della seconda rispetto alla prima. Per risolvere il problema così esteso, basta evidentemente considerare lo spazio probabilizzato

$$(E \times F,\ \mathcal{E} \otimes \mathcal{F},\ \mu \otimes N) \tag{3.12}$$

e, su di esso, la coppia (h, k) delle proiezioni canoniche. La terna formata dallo spazio probabilizzato (3.12) e dalle due proiezioni canoniche h, k si chiama talvolta lo *schema di Bayes* associato alla coppia (μ, N).

Nelle ipotesi generali (3.1), non si può garantire né che esista una versione della legge condizionale di Y rispetto a X né che una tal versione sia unica. Tuttavia, come vedremo più avanti, l'esistenza è assicurata non appena s'introducano ipotesi "ragionevoli" sullo spazio d'arrivo di Y (per esempio, l'ipotesi che l'insieme F sia munito di una struttura di spazio metrico completo e separabile e che la tribù $\mathcal{F}$ coincida con la relativa tribù boreliana).

4. Legami col concetto di speranza condizionale

(4.1) Teorema. *Nelle ipotesi (3.1), sia N un nucleo markoviano da $(E, \mathcal{E})$ a $(F, \mathcal{F})$. Le condizioni che seguono sono tra loro equivalenti:*

(a) N è una versione della legge condizionale di Y rispetto a X.

(b) Per ogni funzione f limitata e misurabile su $(E, \mathcal{E})$ ed ogni funzione g limitata e misurabile su $(F, \mathcal{F})$, si ha

$$\int (f \circ X)\,(g \circ Y)\,\mathrm{d}P = \int (f \circ X)\,(Ng \circ X)\,\mathrm{d}P. \tag{4.2}$$

(c) Per ogni funzione g limitata e misurabile su $(F, \mathcal{F})$, una versione della speranza condizionale $P[g \circ Y \mid X]$ è la variabile aleatoria $Ng \circ X$.

Dimostrazione. Tenendo conto dell'Osservazione (3.6), si vede che, affinché la condizione (a) sia verificata, cioè abbia luogo la relazione (3.3), occorre e basta che l'eguaglianza (3.4) valga per ogni funzione h della forma particolare $f \otimes g$ (con f limitata e misurabile su $(E, \mathcal{E})$ e g limitata e misurabile su $(F, \mathcal{F})$). D'altra parte, per una funzione h di questo tipo, l'eguaglianza (3.4) si riduce alla forma (4.2): infatti il primo membro di (3.4) può esser messo nella forma

$$\int (f \circ X)\,(g \circ Y)\,\mathrm{d}P,$$

mentre il secondo membro di (3.4) può esser messo in una delle forme seguenti

$$\int \mu(\mathrm{d}x)\, f(x) \int N(x, \mathrm{d}y)\, g(y), = \int \mu(\mathrm{d}x)\, f(x)\, Ng\,(x), = \int (f \circ X)\,(Ng \circ X)\,\mathrm{d}P.$$

È dunque provata l'equivalenza tra (a) e (b). L'equivalenza tra (b) e (c) è poi immediata: infatti, fissata una funzione g limitata e misurabile su $(F, \mathcal{F})$, la variabile aleatoria $Ng \circ X$ è una versione della speranza condizionale $P\big[\, g \circ Y \,|\, X \,\big]$ se, e solo se, la relazione (4.2) ha luogo per ogni funzione f limitata e misurabile su $(E, \mathcal{E})$. □

5. Un risultato di unicità

(5.1) Definizione. Nelle ipotesi (3.1), diremo che due versioni N, N' della legge condizionale di Y rispetto a X sono tra loro *equivalenti secondo* μ se l'insieme

$$\big\{x \in E :\ N(x, \cdot) = N'(x, \cdot)\big\} \tag{5.2}$$

appartiene alla tribù $\mathcal{E}$ e ha misura eguale a 1 secondo μ.

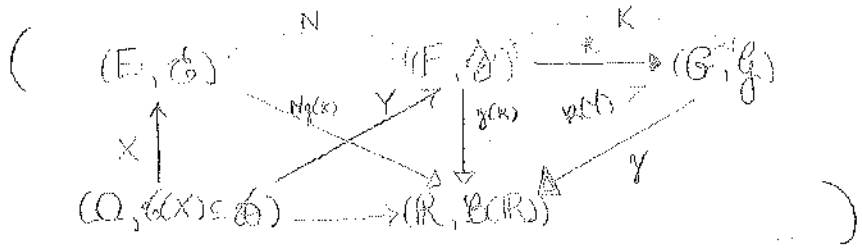
Impiegando la terminologia appena introdotta, possiamo enunciare il teorema seguente, il quale esprime una sorta di "unicità debole" della legge condizionale.

(5.3) Teorema. *Nelle ipotesi (3.1), si supponga per giunta che la tribù $\mathcal{F}$ sia separabile. Allora, se N, N' sono due versioni della legge condizionale di Y rispetto a X, esse sono tra loro equivalenti secondo μ.*

Dimostrazione. La tribù $\mathcal{F}$, essendo separabile, possiede una base numerabile. Sia $\mathcal{J}$ una tal base. Per ogni elemento B di $\mathcal{J}$, le due variabili aleatorie $NI_B \circ X$, $N'I_B \circ X$, come versioni della speranza condizionale $P\big[I_B \circ Y \,|\, X\big]$, sono tra loro equivalenti secondo P. Ciò è come dire che le due funzioni NI_B, $N'I_B$ (misurabili su $(E, \mathcal{E})$) sono tra loro equivalenti secondo μ, ossia che l'insieme

$$\big\{x \in E :\ N(x, B) = N'(x, B)\big\}$$

(appartenente alla tribù $\mathcal{E}$) ha misura 1 secondo μ. D'altra parte, l'intersezione degli insiemi di questa forma (al variare di B in $\mathcal{J}$) coincide, grazie al criterio fondamentale per la coincidenza di due misure, con l'insieme (5.2). Quest'ultimo insieme è dunque anch'esso un elemento di $\mathcal{E}$ di misura 1 secondo μ. Il teorema è così dimostrato. □



6. Legge condizionale di una funzione composta

(6.1) Proposizione. *Nelle ipotesi (3.1), si supponga che il nucleo N sia una versione della legge condizionale di Y rispetto a X. Sia poi k un'applicazione misurabile di $(F,\mathcal{F})$ in uno spazio misurabile $(G,\mathcal{G})$.*

Allora una versione della legge condizionale di $k\circ Y$ rispetto a X è il nucleo composto NK, dove K denota il nucleo, da $(F,\mathcal{F})$ a $(G,\mathcal{G})$, associato a k.

Dimostrazione. Grazie alla caratterizzazione del concetto di legge condizionale (in termini di speranza condizionale) fornita dal Teorema (4.1), si tratta di dimostrare che, per ogni funzione g limitata e misurabile su $(G,\mathcal{G})$, una versione della speranza condizionale

$$P\big[\,g\circ(k\circ Y)\,\big|\,X\,\big] \tag{6.2}$$

è la variabile aleatoria $(NK)g\circ X$. A questo scopo, basta osservare che, grazie all'eguaglianza $g\circ k = Kg$, la speranza condizionale (6.2) può essere scritta nella forma

$$P\big[\,(Kg)\circ Y\,\big|\,X\,\big] \tag{6.3}$$

e quindi, poiché N è una versione della legge condizionale di Y rispetto a X, ammette come versione la variabile aleatoria $N(Kg)\circ X$. □

7. Legge condizionale in presenza di un legame funzionale

Il teorema che segue riguarda un caso particolare del concetto di legge condizionale di Y rispetto a X: quello in cui la variabile aleatoria X sia legata a Y da una relazione del tipo $X = h\circ Y$.

(7.1) Teorema. *Nelle ipotesi (3.1), siano N un nucleo markoviano da $(E,\mathcal{E})$ a $(F,\mathcal{F})$ e h un'applicazione misurabile di $(F,\mathcal{F})$ in $(E,\mathcal{E})$, tale che si abbia*

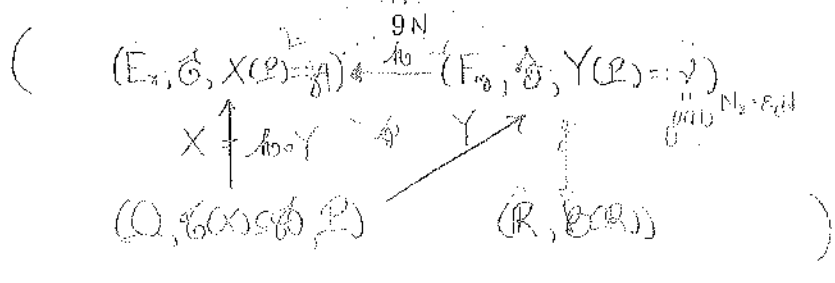
$$X = h\circ Y, \qquad NH = J, \tag{7.2}$$

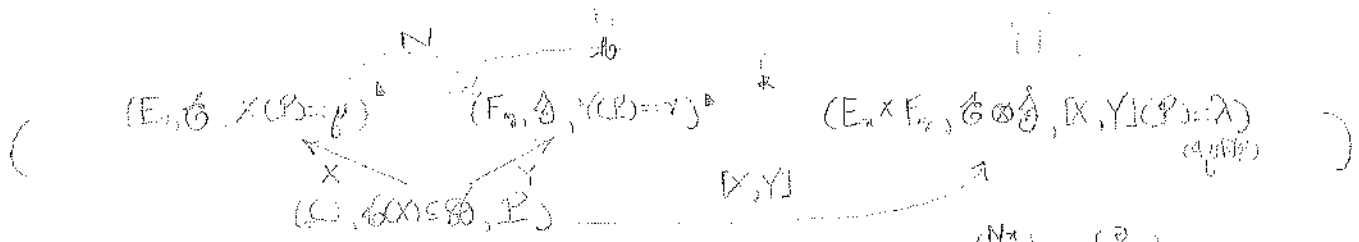
dove H denota il nucleo, da $(F,\mathcal{F})$ a $(E,\mathcal{E})$, associato a h e J il nucleo identità relativo a $(E,\mathcal{E})$.

Allora, affinché N sia una versione della legge condizionale di Y rispetto a X, è (necessario e) sufficiente che la legge ν di Y sia eguale a μN.

Dimostrazione. La necessità è già nota (si veda (3.9)). Per provare la sufficienza, supposto $\nu = \mu N$, fissiamo una funzione f limitata e misurabile su $(E,\mathcal{E})$ e una funzione g limitata e misurabile su $(F,\mathcal{F})$. Si ha allora

$$\int (f\circ X)\,(g\circ Y)\,\mathrm{d}P = \int (f\circ h\circ Y)\,(g\circ Y)\,\mathrm{d}P = \langle \mu N,\,(f\circ h)\,g\rangle. \tag{7.3}$$





D'altra parte, per ogni elemento x di E, poiché la legge di h secondo $\epsilon_x N$ è $\epsilon_x \widetilde{N} H$, ossia ϵ_x, quella di $f \circ h$ è la legge della costante $f(x)$. Dunque, secondo $\epsilon_x N$, la funzione reale $f \circ h$ è equivalente alla costante reale $f(x)$. Si ha perciò

$$\begin{aligned} \langle \mu N, (f \circ h)\, g \rangle &= \int \mu(\mathrm{d}x) \int N(x, \mathrm{d}y)\, (f \circ h)(y)\, g(y) \\ &= \int \mu(\mathrm{d}x)\, f(x) \int N(x, \mathrm{d}y)\, g(y) \\ &= \int \mu(\mathrm{d}x)\, f(x)\, Ng(x) = \int (f \circ X)\, (Ng \circ X)\, \mathrm{d}P. \end{aligned} \tag{7.4}$$

Dalle relazioni (7.3), (7.4) discende l'eguaglianza (4.2). Grazie al Teorema (4.1), ciò basta per concludere che il nucleo N è una versione della legge condizionale di Y rispetto a X. □

(7.5) Corollario. *Nelle ipotesi (3.1), sia N un un nucleo markoviano da $(E, \mathcal{E})$ a $(F, \mathcal{F})$. Le condizioni che seguono sono tra loro equivalenti:*

(a) Il nucleo N è una versione della legge condizionale di Y rispetto a X.

(b) Il nucleo $\widetilde{N}$ è una versione della legge condizionale di $[X, Y]$ rispetto a X.

Dimostrazione. (a) $\Rightarrow$ (b): Denotiamo con h la proiezione canonica di $E \times F$ sul primo fattore, con H il nucleo, da $(E \times F, \mathcal{E} \otimes \mathcal{F})$ a $(E, \mathcal{E})$, associato a h e con J il nucleo identità relativo a $(E, \mathcal{E})$. Si hanno allora le eguaglianze

$$X = h \circ [X, Y], \qquad \widetilde{N} H = J$$

(per la seconda delle quali rimandiamo a (2.10)). Inoltre la condizione (a) si traduce nell'eguaglianza $[X, Y](P) = \mu \widetilde{N}$. Essa implica dunque la condizione (b), come si vede applicando il Teorema (7.1) al nucleo $\widetilde{N}$.

(b) $\Rightarrow$ (a): Denotiamo con k la proiezione canonica di $E \times F$ sul secondo fattore, e con K il nucleo, da $(E \times F, \mathcal{E} \otimes \mathcal{F})$ a $(F, \mathcal{F})$, associato a k. Si hanno allora le eguaglianze

$$Y = k \circ [X, Y], \qquad \widetilde{N} K = N$$

(per la seconda delle quali rimandiamo a (2.10)). Dunque, come si vede applicando la Proposizione (6.1) al nucleo $\widetilde{N}$, la condizione (b) implica che una versione della legge condizionale di Y rispetto a X è il nucleo composto $\widetilde{N} K$, identico a N. □

8. Leggi condizionali per sottotribù

Nelle ipotesi (3.1) della definizione di legge condizionale, le variabili aleatorie X, Y sono *a priori* del tutto arbitrarie. Non è escluso che una di esse, ad esempio Y, sia l'applicazione identica di Ω, considerata come variabile aleatoria su $(\Omega, \mathcal{A}, P)$ a valori nello spazio misurabile ottenuto munendo Ω di un'opportuna sottotribù $\mathcal{K}$ di $\mathcal{A}$. In questo caso, per alleggerire il linguaggio, sarà comodo confondere Y con la stessa sottotribù $\mathcal{K}$, e parlare dunque di legge condizionale *della sottotribù* $\mathcal{K}$

rispetto a X. Analoga convenzione si adotterà nel caso in cui X sia l'applicazione identica di Ω, considerata come variabile aleatoria su $(\Omega, \mathcal{A}, P)$ a valori nello spazio misurabile ottenuto munendo Ω di un'opportuna sottotribù $\mathcal{H}$ di $\mathcal{A}$: si parlerà allora di legge condizionale di Y *rispetto alla sottotribù* $\mathcal{H}$. Quando poi entrambe le variabili aleatorie in gioco siano del tipo particolare sopra descritto, si parlerà di legge condizionale della sottotribù $\mathcal{K}$ rispetto alla sottotribù $\mathcal{H}$.

Un esempio molto semplice d'impiego della terminologia sopra introdotta è fornito dalla proposizione seguente.

(8.1) Proposizione. *Nelle ipotesi (3.1), assegnato un nucleo markoviano N da $(E, \mathcal{E})$ a $(F, \mathcal{F})$, si consideri il nucleo Q, da $(\Omega, \mathcal{T}(X))$ a $(F, \mathcal{F})$, che opera così sulle funzioni: per ogni funzione g limitata e misurabile su $(F, \mathcal{F})$, la sua trasformata mediante Q è data da*

$$Qg = Ng \circ X. \tag{8.2}$$

Le condizioni che seguono sono allora equivalenti:

(a) Il nucleo N è una versione della legge condizionale di Y rispetto a X.

(b) Il nucleo Q è una versione della legge condizionale di Y rispetto a $\mathcal{T}(X)$.

Dimostrazione. Grazie al Teorema (4.1), affinché valga la condizione (a), occorre e basta che, per ogni funzione g limitata e misurabile su $(F, \mathcal{F})$, una versione della speranza condizionale

$$P[\, g \circ Y \,|\, X \,] \tag{8.3}$$

sia la variabile aleatoria $Ng \circ X$.

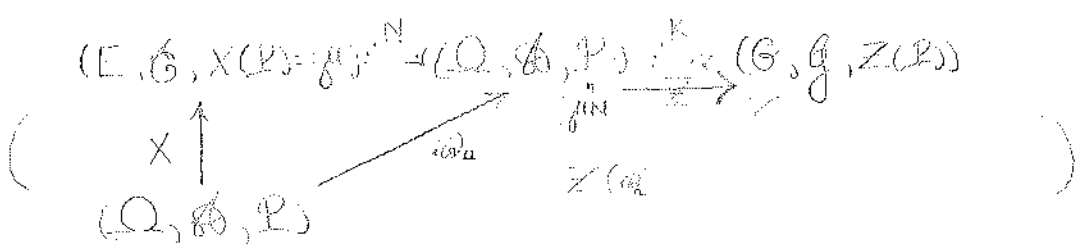
Analogamente, se si denota con X' l'applicazione identica di Ω, pensata come variabile aleatoria su $(\Omega, \mathcal{A}, P)$ a valori in $\big(\Omega, \mathcal{T}(X)\big)$, si può dire che, affinché valga la condizione (b), occorre e basta che, per ogni funzione g limitata e misurabile su $(F, \mathcal{F})$, una versione della speranza condizionale

$$P[\, g \circ Y \,|\, X' \,] \tag{8.4}$$

sia la variabile aleatoria $Qg \circ X'$.

Dunque, per provare l'equivalenza tra (a) e (b), basta osservare che le speranze condizionali (8.3), (8.4) sono identiche, in quanto entrambe eguali a $P\big[\, g \circ Y \,|\, \mathcal{T}(X) \,\big]$, e che anche le variabili aleatorie $Ng \circ X$, $Qg \circ X'$ sono tra loro identiche, in quanto entrambe eguali a Qg. □

(8.5) Osservazione. Il nucleo Q considerato in (8.1) si può vedere come il nucleo composto NK, dove K denota il nucleo, da $(\Omega, \mathcal{T}(X))$ a $(E, \mathcal{E})$, associato a X.



9. Disintegrazione di una misura di probabilità

Vogliamo da ultimo esaminare un caso molto speciale della situazione studiata nella sezione precedente.

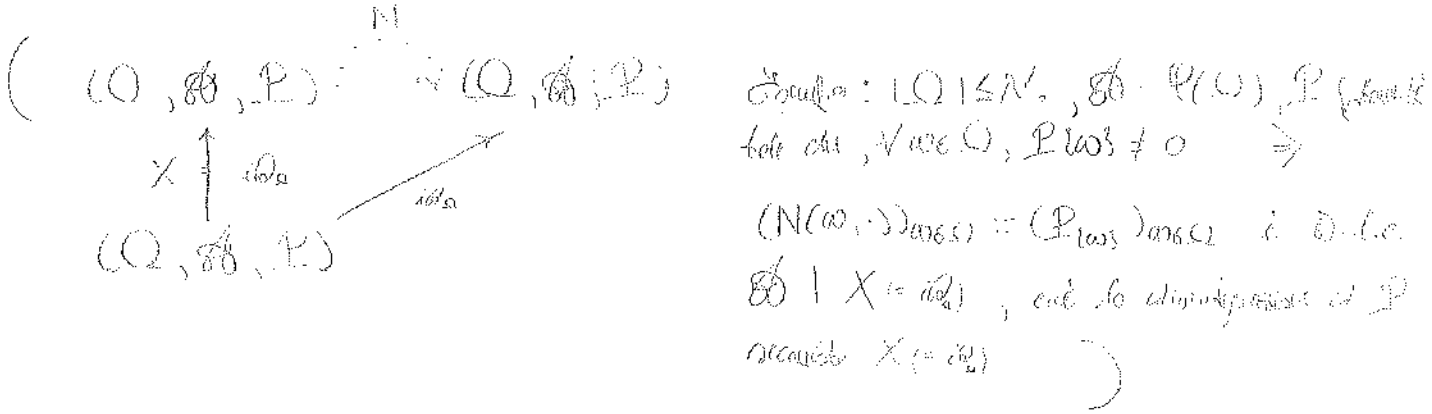
(9.1) Definizione. Su uno spazio probabilizzato $(\Omega, \mathcal{A}, P)$ sia X una variabile aleatoria a valori in uno spazio misurabile $(E, \mathcal{E})$. Si chiama *disintegrazione* di P secondo X un nucleo markoviano N, da $(E, \mathcal{E})$ a $(\Omega, \mathcal{A})$, che sia una versione della legge condizionale di $\mathcal{A}$ rispetto a X (cioè che sia una versione della legge condizionale, rispetto a X, dell'applicazione identica di Ω pensata come variabile aleatoria su $(\Omega, \mathcal{A}, P)$ a valori in $(\Omega, \mathcal{A})$),

Si dice che la misura P è *disintegrabile* secondo X se esiste una disintegrazione di P secondo X.

La proposizione seguente, la quale non è che un caso particolare della Proposizione (6.1) riguardante la legge condizionale di una funzione composta, mostra che, nelle ipotesi di (9.1), se P è disintegrabile secondo X, allora, comunque si assegni una variabile aleatoria su $(\Omega, \mathcal{A}, P)$, esiste sempre una versione della sua legge condizionale rispetto a X.

(9.2) Proposizione. *Nelle ipotesi della Definizione (9.1), si supponga che N sia una disintegrazione di P secondo X. Allora, se Z è una variabile aleatoria su $(\Omega, \mathcal{A}, P)$, a valori in un arbitrario spazio misurabile $(G, \mathcal{G})$, una versione della legge condizionale di Z rispetto a X è il nucleo composto NK, dove K denota il nucleo, da $(\Omega, \mathcal{A})$ a $(G, \mathcal{G})$, associato a Z.*

Dimostrazione. Basta applicare la Proposizione (6.1), nella quale si prenda k eguale a Z e Y eguale all'applicazione identica di Ω (pensata come variabile aleatoria su $(\Omega, \mathcal{A}, P)$ a valori in $(\Omega, \mathcal{A})$), in modo da poter scrivere Z nella forma $k \circ Y$. □



Complementi sui nuclei markoviani *

1. Notazioni e risultati preliminari

Per ogni spazio misurabile $(E, \mathcal{E})$, denoteremo con $\mathcal{R}(E, \mathcal{E})$ lo spazio di Riesz costituito dalle funzioni limitate e misurabili su $(E, \mathcal{E})$. Tutti i nuclei che considereremo *saranno tacitamente supposti markoviani*. Ricordiamo che un nucleo, relativo a una fissata coppia $(E, \mathcal{E})$, $(F, \mathcal{F})$ di spazi misurabili, può sempre essere identificato con uno speciale *operatore su funzioni*: precisamente, con un'applicazione lineare isotona di $\mathcal{R}(F, \mathcal{F})$ in $\mathcal{R}(E, \mathcal{E})$ che trasformi la costante 1 nella costante 1 e possieda la *proprietà di Daniell* (ossia la proprietà di continuità sulle successioni decrescenti dotate di inviluppo inferiore nullo). Ci sarà utile il risultato seguente.

(1.1) Proposizione. *Siano $(E, \mathcal{E})$, $(F, \mathcal{F})$ due spazi misurabili, e sia $\mathcal{S}$ un sottospazio di Riesz di $\mathcal{R}(F, \mathcal{F})$, il quale contenga le costanti e generi la tribù $\mathcal{F}$.*

Sia poi L_0 un'applicazione lineare isotona $g \mapsto L_0 g$ di $\mathcal{S}$ in $\mathcal{R}(E, \mathcal{E})$ la quale trasformi la costante 1 nella costante 1 e possieda la proprietà di Daniell.

Allora è possibile, in un sol modo, costruire un nucleo L, relativo alla coppia $(E, \mathcal{E})$, $(F, \mathcal{F})$, il quale "prolunghi" L_0 (nel senso che operi come L_0 sugli elementi di $\mathcal{S}$).

Dimostrazione. Per ogni elemento x di E, l'applicazione $g \mapsto L_0 g(x)$ è un funzionale di Daniell su $\mathcal{S}$, assumente il valore 1 sulla costante 1. Il teorema di rappresentazione di Daniell assicura che esiste su $\mathcal{F}$ un'unica misura di probabilità $L(x, \cdot\,)$ tale che, per ogni elemento g di $\mathcal{S}$, si abbia

$$L_0 g(x) = \int L(x, \mathrm{d}y)\, g(y).$$

Dunque, indicando con L la famiglia di misure di probabilità $\big(L(x, \cdot\,)\big)_{x \in E}$, si può scrivere $Lg = L_0 g$ per ogni elemento g di $\mathcal{S}$. Ne segue che la classe costituita dalle funzioni g appartenenti a $\mathcal{R}(F, \mathcal{F})$ e tali che Lg stia in $\mathcal{R}(E, \mathcal{E})$ è monotona e contiene $\mathcal{S}$, sicché coincide con l'intero spazio $\mathcal{R}(F, \mathcal{F})$. Ciò prova che L è un nucleo. È poi evidente che questo nucleo è l'unico che "prolunghi" l'applicazione L_0. □

Ricordiamo che, se N è un nucleo da $(E, \mathcal{E})$ a $(F, \mathcal{F})$, si denota con $\widetilde{N}$ il nucleo, da $(E, \mathcal{E})$ a $(E \times F,\ \mathcal{E} \otimes \mathcal{F})$, che opera sulle funzioni nel modo seguente: per ogni elemento g di $\mathcal{R}(E \times F,\ \mathcal{E} \otimes \mathcal{F})$, la trasformata di g mediante $\widetilde{N}$ è la funzione $\widetilde{N}g$ così definita su E:

$$\widetilde{N}g\,(x) = \int N(x, \mathrm{d}y)\, g(x, y) \quad \text{per} \quad x \in E. \tag{1.2}$$

(1.3) Osservazione. Sullo spazio probabilizzabile $(\Omega, \mathcal{A})$ sia X una variabile aleatoria, a valori nello spazio misurabile $(E, \mathcal{E})$, con $X(\Omega) = E$. Si denoti con H il nucleo, da $(\Omega, \mathcal{T}(X))$ a $(E, \mathcal{E})$, associato a X. Allora l'applicazione $f \mapsto f \circ X$

di $\mathcal{R}(E,\mathcal{E})$ in $\mathcal{R}(\Omega,\mathcal{T}(X))$ indotta dal nucleo H è un isomorfismo di spazi vettoriali ordinati: infatti, oltre ad essere lineare e isotona, è iniettiva (a causa dell'ipotesi $X(\Omega)=E$) e ha come immagine l'intero spazio $\mathcal{R}(\Omega,\mathcal{T}(X))$ (grazie al lemma di misurabilità di Doob). L'applicazione inversa, che è un isomorfismo di $\mathcal{R}(\Omega,\mathcal{T}(X))$ su $\mathcal{R}(E,\mathcal{E})$, è indotta da un nucleo markoviano, relativo alla coppia di spazi misurabili $(E,\mathcal{E})$, $(\Omega,\mathcal{T}(X))$, che sarà naturale chiamare il nucleo *inverso* di H e denotare con H^{-1}. È chiaro che il nucleo composto HH^{-1} coincide col nucleo identità relativo allo spazio misurabile $(\Omega,\mathcal{T}(X))$.

2. Nuclei proiettivi

(2.1) Proposizione. *Siano dati uno spazio probabilizzabile $(\Omega,\mathcal{A})$, una coppia $\mathcal{G}_0,\mathcal{G}_1$ di sottotribù di $\mathcal{A}$, con $\mathcal{G}_0\subset\mathcal{G}_1$, e un nucleo K da $(\Omega,\mathcal{G}_0)$ a $(\Omega,\mathcal{G}_1)$. Le condizioni che seguono sono equivalenti:*

(a) L'applicazione $Z\mapsto KZ$ di $\mathcal{R}(\Omega,\mathcal{G}_1)$ in $\mathcal{R}(\Omega,\mathcal{G}_0)$ lascia fisso ogni elemento di $\mathcal{R}(\Omega,\mathcal{G}_0)$.

(b) Per ogni misura di probabilità P su $(\Omega,\mathcal{G}_0)$, la misura PK (trasformata di P mediante il nucleo K) è un prolungamento di P.

Dimostrazione. La condizione (a) equivale ad imporre che, per ogni elemento Z di $\mathcal{R}(\Omega,\mathcal{G}_0)$ ed ogni misura di probabilità P su $(\Omega,\mathcal{G}_0)$, abbia luogo l'eguaglianza

$$\langle P,\,KZ\rangle=\langle P,\,Z\rangle. \tag{2.2}$$

D'altra parte, questa eguaglianza si può scrivere nella forma equivalente

$$\langle PK,\,Z\rangle=\langle P,\,Z\rangle. \tag{2.3}$$

Basta dunque osservare che, per un'assegnata misura di probabilità P su $(\Omega,\mathcal{G}_0)$, il il fatto che l'eguaglianza (2.3) abbia luogo per ogni elemento Z di $\mathcal{R}(\Omega,\mathcal{G}_0)$ equivale al fatto che la restrizione di PK a $\mathcal{G}_0$ coincida con P, cioè che la misura PK sia un prolungamento di P. Si vede dunque che le due condizioni (a),(b) sono equivalenti. □

(2.4) Definizione. Quando, nelle ipotesi della proposizione precedente, siano verificate le condizioni equivalenti (a), (b), si dirà che K è un *nucleo proiettivo* o anche un *nucleo di prolungamento.*

(2.5) Proposizione. *Assegnate, su uno spazio probabilizzabile $(\Omega,\mathcal{A})$, due variabili aleatorie X,Y, a valori negli spazi misurabili $(E,\mathcal{E})$, $(F,\mathcal{F})$, con*

$$[X,Y](\Omega)=E\times F, \tag{2.6}$$

si ponga $\mathcal{G}_0=\mathcal{T}(X)$, $\mathcal{G}_1=\mathcal{T}(X,Y)$. Sia poi N un nucleo da $(E,\mathcal{E})$ a $(F,\mathcal{F})$.

Allora esiste un nucleo proiettivo K, da $(\Omega, \mathcal{G}_0)$ a $(\Omega, \mathcal{G}_1)$, tale che, per ogni misura di probabilità P su $(\Omega, \mathcal{A})$, indicando con P_0, P_1 le restrizioni di P a $\mathcal{G}_0$ e a $\mathcal{G}_1$ rispettivamente, si abbia equivalenza tra le due condizioni seguenti:

(a) Secondo la misura P, il nucleo N è una versione della legge condizionale di Y rispetto a X.

(b) Si ha $P_1 = P_0 K$.

Dimostrazione. Denotiamo con H_0 il nucleo, da $(\Omega, \mathcal{G}_0)$ a $(E, \mathcal{E})$, associato a X e con H_1 il nucleo, da $(\Omega, \mathcal{G}_1)$ a $(E \times F, \mathcal{E} \otimes \mathcal{F})$, associato al blocco $[X, Y]$. Allora, assegnata una misura di probabilità P su $(\Omega, \mathcal{A})$ e indicate con P_0, P_1 le sue restrizioni a $\mathcal{G}_0$ e a $\mathcal{G}_1$ rispettivamente, la legge di X secondo P (o, ciò ch'è lo stesso, secondo P_0) si può scrivere nella forma $P_0 H_0$ e, analogamente, la legge di $[X, Y]$ secondo P si può scrivere nella forma $P_1 H_1$. Perciò, grazie alla definizione di legge condizionale, è possibile tradurre la condizione (a) mediante una delle due relazioni

$$P_1 H_1 = P_0 H_0 \widetilde{N}, \qquad P_1 = P_0 H_0 \widetilde{N} H_1^{-1},$$

o anche mediante la relazione $P_1 = P_0 K$, pur di porre

$$K = H_0 \widetilde{N} H_1^{-1}. \tag{2.7}$$

Per completare la dimostrazione, basta osservare che il nucleo K così definito è proiettivo, grazie al fatto che, per ogni elemento Z di $\mathcal{R}(\Omega, \mathcal{G}_0)$, cioè della forma $Z = f \circ X$, con $f \in \mathcal{R}(E, \mathcal{E})$, si ha

$$Z = (f \otimes 1) \circ [X, Y] = H_1 (f \otimes 1)$$

e quindi

$$KZ = H_0 \widetilde{N} (f \otimes 1) = H_0 f = Z. \qquad \square$$

(2.8) Definizione. Nelle ipotesi di (2.5), il nucleo K costruito nella dimostrazione precedente si chiamerà il nucleo *indotto* da N *per mezzo della coppia* (X, Y).

3. Prodotto di una catena di nuclei proiettivi

(3.1) Ipotesi e notazioni. Si suppone assegnata una successione $\big((E_n, \mathcal{E}_n)\big)_{n \geq 0}$ di spazi misurabili, e si pone

$$\Omega = \prod_{n \geq 0} E_n, \qquad \mathcal{A} = \bigotimes_{n \geq 0} \mathcal{E}_n.$$

Si denota con $(X_n)_{n \geq 0}$ la successione delle proiezioni canoniche di Ω sui singoli fattori. Per ogni intero positivo n, si pone

$$F_n = \prod_{j=0}^{n} E_j, \qquad \mathcal{F}_n = \prod_{j=0}^{n} \mathcal{E}_j, \qquad \mathcal{G}_n = \mathcal{T}(X_0, \ldots, X_n)$$

e si denota più semplicemente con $\mathcal{R}_n$ lo spazio di Riesz $\mathcal{R}(\Omega, \mathcal{G}_n)$. Si pone infine $\mathcal{R}_\infty = \bigcup_n \mathcal{R}_n$.

(3.2) Definizione. Nelle ipotesi (3.1), chiameremo *catena di nuclei proiettivi* adattata alla filtrazione $(\mathcal{G}_n)_{n\geq 0}$ una successione $(K_n)_{n\geq 0}$ di nuclei tale che, per ogni n, il nucleo K_n sia un nucleo proiettivo da $(\Omega, \mathcal{G}_n)$ a $(\Omega, \mathcal{G}_{n+1})$.

Chiameremo poi *prodotto di composizione* (o, semplicemente, *prodotto*) di una tal catena $(K_n)_{n\geq 0}$ un nucleo proiettivo L, da $(\Omega, \mathcal{G}_0)$ a $(\Omega, \mathcal{A})$, il quale "prolunghi" ciascuno dei prodotti di composizione della forma $K_0 \cdots K_n$, nel senso che, per ciascun n, il modo di operare di L sugli elementi di $\mathcal{R}_{n+1}$ sia identico a quello del nucleo composto $K_0 \cdots K_n$.

(3.3) Teorema. *Nelle ipotesi (3.1), sia $(K_n)_{n\geq 0}$ una catena di nuclei proiettivi, adattata alla filtrazione $(\mathcal{G}_n)_{n\geq 0}$. Allora $(K_n)_{n\geq 1}$ ammette un unico prodotto.*

Dimostrazione. L'unicità è immediata: infatti due nuclei, ciascuno dei quali sia prodotto della catena $(K_n)_{n\geq 0}$, operano allo stesso modo su ciascun elemento di $\mathcal{R}_\infty$, dunque sono identici (si veda (1.1)). Proviamo l'esistenza. A questo scopo, per ogni coppia n, h d'interi positivi, denotiamo con $L_{n,n+h}$ il nucleo proiettivo, relativo alla coppia $(\Omega, \mathcal{G}_n)$, $(\Omega, \mathcal{G}_{n+h})$, definito nel modo seguente: $L_{n,n}$ è semplicemente il nucleo identità relativo a $(\Omega, \mathcal{G}_n)$, mentre, per h strettamente positivo, $L_{n,n+h}$ è il nucleo $K_n \cdots K_{n+h-1}$ (prodotto di composizione dei nuclei K_j con $n \leq j < n+h$). È chiaro che, per h strettamente positivo, si ha

$$L_{n,n+h} = K_n L_{n+1,n+h}. \tag{3.4}$$

Fissati un intero positivo n e un elemento Z di $\mathcal{R}_\infty$, poiché i nuclei K_j sono proiettivi, le variabili aleatorie della forma $L_{n,n+h}Z$ (con h tale che si abbia $Z \in \mathcal{R}_{n+h}$) sono tutte eguali a un medesimo elemento di $\mathcal{R}_n$. Denoteremo questo elemento con $L_n Z$. Rimane così definita un'applicazione L_n (lineare e isotona) di $\mathcal{R}_\infty$ in $\mathcal{R}_n$. Tutto è ridotto a dimostrare che L_0 possiede la proprietà di Daniell: infatti, grazie a (1.1), L_0 sarà allora prolungabile in un nucleo L, relativo alla coppia $(\Omega, \mathcal{G}_0)$, $(\Omega, \mathcal{A})$, il quale sarà il desiderato prodotto della catena $(K_n)_{n\geq 0}$.

Cominciamo con l'osservare che, fissato l'elemento Z di $\mathcal{R}_\infty$, si ha, per ogni n,

$$L_n Z = K_n L_{n+1} Z. \tag{3.5}$$

Infatti, scelto un intero h strettamente positivo e tale che si abbia $Z \in \mathcal{R}_{n+h}$, si può scrivere, tenendo conto di (3.4),

$$L_n Z = L_{n,n+h} Z = K_n L_{n+1,n+h} Z = K_n L_{n+1} Z.$$

Osserviamo poi che si ha banalmente

$$\lim_n L_n Z = Z. \tag{3.6}$$

Infatti, non appena n sia abbastanza grande da render vera la relazione $Z \in \mathcal{R}_n$, si ha addirittura $L_n Z = L_{n,n} Z = Z$.

Ciò premesso, volgiamoci ora a dimostrare che l'applicazione L_0 possiede la proprietà di Daniell. Assegnata una successione decrescente $(Z_k)_{k\geq 1}$ di elementi di $\mathcal{R}_\infty$, con $Z_k \downarrow 0$, poniamo

$$Y_n = \inf_{k\geq 1} L_n Z_k = \lim_k L_n Z_k \quad \text{per } n \geq 0. \tag{3.7}$$

Basterà provare che Y_0 è la costante 0. A questo scopo, cominciamo con l'osservare che, grazie a (3.7), si ha

$$0 \leq Y_n \leq L_n Z_k \quad \text{per } n \geq 0 \text{ e } k \geq 1.$$

Facendo tendere n all'infinito in questa relazione, si trova, grazie a (3.6),

$$0 \leq \limsup_n Y_n \leq \lim_n L_n Z_k = Z_k \quad \text{per } k \geq 1.$$

Dall'ipotesi $Z_k \downarrow 0$ discende dunque la relazione $\limsup_n Y_n = 0$, ossia la convergenza puntuale di $(Y_n)_{n\geq 0}$ verso 0.

Osserviamo ora che, per ogni n, essendo $Y_n \in \mathcal{R}_n$, esiste un (unico) elemento f_n di $\mathcal{R}(F_n, \mathcal{F}_n)$ tale che si abbia

$$Y_n = f_n \circ [X_0, \ldots, X_n] = H_n f_n, \tag{3.8}$$

dove H_n denota il nucleo, da $(\Omega, \mathcal{G}_n)$ a $(F_n, \mathcal{F}_n)$, associato al blocco $[X_0, \ldots, X_n]$. Essendo, in particolare, $Y_0 = f_0 \circ X_0$, siamo ridotti a provare che f_0 è la costante 0. A questo scopo, osserviamo che, grazie a (3.5), si ha

$$L_n Z_k = K_n L_{n+1} Z_k \quad \text{per } n \geq 0 \text{ e } k \geq 1.$$

Di qui, facendo tendere k all'infinito (e sfruttando il fatto che K_n possiede la proprietà di Daniell), si deduce $Y_n = K_n Y_{n+1}$. Grazie a (3.8), è possibile mettere questa eguaglianza in una delle due forme

$$H_n f_n = K_n H_{n+1} f_{n+1}, \qquad f_n = H_n^{-1} K_n H_{n+1} f_{n+1},$$

o anche, ponendo $M_n = H_n^{-1} K_n H_{n+1}$, nella forma più concisa

$$f_n = M_n f_{n+1}. \tag{3.9}$$

Quest'ultima eguaglianza implica, per ogni elemento $(x_0, \ldots, x_n)$ di $E_0 \times \cdots \times E_n$, la relazione

$$\begin{aligned} f_n(x_0, \ldots, x_n) &= \int M_n(x_0, \ldots, x_n, \mathrm{d}y)\, f_{n+1}(x_0, \ldots, x_n, y) \\ &\leq \sup_{y \in E_{n+1}} f_{n+1}(x_0, \ldots, x_n, y). \end{aligned} \tag{3.10}$$

Supposto, per assurdo, che la funzione f_0 non sia la costante 0, scegliamo un elemento x_0 di $\{f_0 > 0\}$ e un numero reale ϵ con

$$0 < \epsilon < f_0(x_0).$$

Sfruttando la relazione (3.10), è allora possibile costruire (per induzione) un elemento ω di Ω, con $X_0(\omega) = x_0$, in modo tale che, ponendo $x_n = X_n(\omega)$ per $n \geq 1$, abbia luogo, per ogni intero positivo n, la relazione $f_n(x_0, \ldots, x_n) > \epsilon$, ossia, grazie a (3.8), la relazione $Y_n(\omega) > \epsilon$. Poiché ciò contraddice la già osservata convergenza puntuale di $(Y_n)_{n\geq 0}$ verso 0, il teorema è dimostrato. □

21 Il teorema di Ionescu Tulcea

1. Il teorema di Ionescu Tulcea in un quadro canonico

(1.1) Ipotesi e notazioni. Si suppone assegnata una successione $\big((E_n, \mathcal{E}_n)\big)_{n\geq 0}$ di spazi misurabili e si pone

$$F = \prod_{n\geq 0} E_n, \qquad \mathcal{F} = \bigotimes_{n\geq 0} \mathcal{E}_n.$$

Si denota con $(\xi_n)_{n\geq 0}$ la successione delle proiezioni canoniche di F sui singoli fattori e con $(\mathcal{F}_n)_{n\geq 0}$ la corrispondente *filtrazione canonica*, definita ponendo

$$\mathcal{F}_n = \mathcal{T}(\xi_0, \ldots, \xi_n).$$

Si suppone inoltre assegnato, per ogni intero positivo n, un nucleo markoviano N_n da $\big(E_0 \times \cdots \times E_n,\ \mathcal{E}_0 \otimes \cdots \otimes \mathcal{E}_n\big)$ a $\big(E_{n+1}, \mathcal{E}_{n+1}\big)$.

(1.2) Teorema. (Ionescu-Tulcea) *Nelle ipotesi (1.1), esiste un nucleo Q, relativo alla coppia $(E_0, \mathcal{E}_0)$, $(F, \mathcal{F})$, caratterizzato dalla proprietà seguente:*

Per ogni misura di probabilità μ su $(E_0, \mathcal{E}_0)$, la misura μQ (trasformata di μ mediante il nucleo Q) è l'unica misura di probabilità ν su $(F, \mathcal{F})$ per la quale valgano le due condizioni seguenti:

(a) Secondo ν, la misura μ è la legge di ξ_0.

(b) Secondo ν, per ogni intero positivo n, il nucleo N_n è una versione della legge condizionale di ξ_{n+1} rispetto al blocco $[\xi_0, \ldots, \xi_n]$.

Dimostrazione. Denoteremo con H_0 il nucleo, da $(F, \mathcal{F}_0)$ a $(E_0, \mathcal{E}_0)$, associato a ξ_0. Inoltre, per ogni intero positivo n, denoteremo con K_n il nucleo proiettivo, da $(F, \mathcal{F}_n)$ a $(F, \mathcal{F}_{n+1})$, indotto dal nucleo N_n per mezzo della coppia

$$\big([\xi_0, \ldots, \xi_n],\ \xi_{n+1}\big).$$

Sia ν una misura di probabilità su $(F, \mathcal{F})$ e denotiamo, per ogni n, con ν_n la restrizione di ν alla tribù $\mathcal{F}_n$. Allora, affinché ν verifichi la condizione (a), occorre e basta che valga la relazione $\xi_0(\nu_0) = \mu$, ossia $\nu_0 H_0 = \mu$, che, grazie all'invertibilità del nucleo H_0, può anche esser messa nella forma

$$\nu_0 = \mu H_0^{-1}. \tag{1.3}$$

Affinché ν verifichi la condizione (b), occorre e basta, grazie alle proprietà dei nuclei K_n, che si abbia

$$\nu_{n+1} = \nu_n K_n \quad \text{per } n \geq 0. \tag{1.4}$$

Indicando con L il prodotto della catena di nuclei proiettivi $(K_n)_{n\geq 0}$, osserviamo che, affinché la misura ν verifichi la relazione (1.4), occorre e basta che, per ogni n, essa coincida, sulla tribù $\mathcal{F}_{n+1}$, con la trasformata di ν_0 mediante il nucleo proiettivo $K_0 \cdots K_n$ o, ciò ch'è lo stesso, con la trasformata di ν_0 mediante il nucleo L. In altri termini, occorre e basta che ν verifichi la relazione

$$\nu = \nu_0 L. \tag{1.5}$$

Si vede allora che, affinché ν verifichi entrambe le condizioni (a), (b), occorre e basta che ν verifichi entrambe le relazioni (1.3), (1.5). D'altra parte, queste due relazioni, poiché L è un nucleo di prolungamento, possono essere compendiate nell'unica relazione seguente: $\nu = \mu H_0^{-1} L$. Si riconosce dunque che l'unico nucleo Q dotato della proprietà descritta nell'enunciato del teorema si ottiene ponendo $Q = H_0^{-1} L$. □

(1.6) Definizione. Nelle ipotesi del teorema precedente, il nucleo Q sarà detto *il nucleo di Ionescu Tulcea* associato alla successione di nuclei $(N_n)_{n\geq 0}$ e sarà denotato con $\bigotimes_{n\geq 0} N_n$.

2. Impiego del nucleo di Ionescu Tulcea

Abbandonando il quadro "canonico" nel quale ci siamo fin qui posti, vogliamo ora applicare il concetto di nucleo di Ionescu Tulcea a un'arbitraria successione di variabili aleatorie, definite su un arbitrario spazio probabilizzato.

(2.1) Proposizione. *Nelle ipotesi (1.1), sia data, su un arbitrario spazio probabilizzato $(\Omega, \mathcal{A}, P)$, una successione $(X_n)_{n\geq 0}$ di variabili aleatorie, tale che, per ogni n, lo spazio d'arrivo di X_n sia $(E_n, \mathcal{E}_n)$. Si denoti con X il blocco $[X_n]_{n\geq 0}$ (da considerare come una variabile aleatoria a valori nello spazio prodotto $(F, \mathcal{F})$ definito in (1.1)) e si ponga*

$$\nu = X(P), \quad \mu = X_0(P) = \xi_0(\nu). \tag{2.2}$$

Si denoti inoltre con Q il nucleo di Ionescu Tulcea $\bigotimes_{n\geq 0} N_n$. Le condizioni che seguono sono allora equivalenti:

(a) Per ogni n, il nucleo N_n è, secondo P, una versione della legge condizionale di X_{n+1} rispetto al blocco $[X_0, \ldots, X_n]$.

(b) Per ogni n, il nucleo N_n è, secondo μ, una versione della legge condizionale di ξ_{n+1} rispetto al blocco $[\xi_0, \ldots, \xi_n]$.

(c) Si ha $\nu = \mu Q$.

(d) Secondo P, il nucleo Q è una versione della legge condizionale del blocco X rispetto a X_0.

Dimostrazione. La condizione (a) equivale al fatto che, per ogni n ed ogni coppia f, g di funzioni misurabili e limitate sugli spazi $(E_0 \times \cdots \times E_n, \mathcal{E}_0 \otimes \cdots \mathcal{E}_n)$, $(E_{n+1}, \mathcal{E}_{n+1})$, abbia luogo la relazione

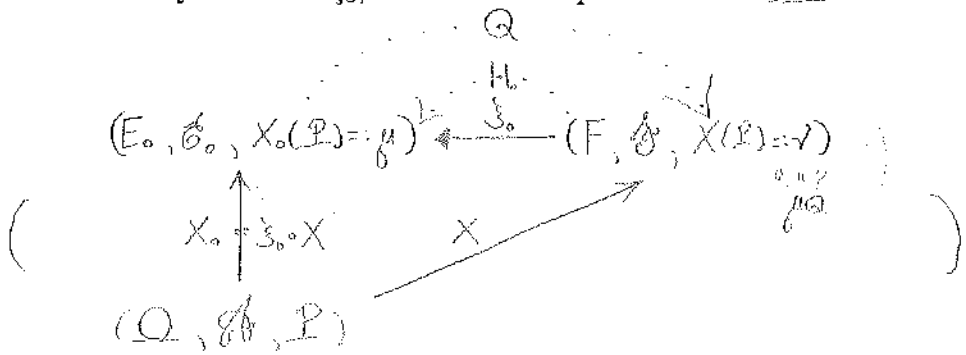
$$\int (f \circ [X_0, \ldots, X_n])(g \circ X_{n+1}) \mathrm{d}P = \int (f \circ [X_0, \ldots, X_n])(N_n g \circ [X_0, \ldots, X_n]) \mathrm{d}P. \tag{2.3}$$

Per provare l'equivalenza tra (a) e (b) basta dunque osservare che, essendo $\nu = X(P)$ e $X_j = \xi_j \circ X$, la relazione (2.3) è equivalente (grazie al teorema sulla speranza di una funzione composta) all'analoga relazione

$$\int (f \circ [\xi_0, \ldots, \xi_n])(g \circ \xi_{n+1}) \mathrm{d}\nu = \int (f \circ [\xi_0, \ldots, \xi_n])(N_n g \circ [\xi_0, \ldots, \xi_n]) \mathrm{d}\nu. \tag{2.4}$$

– L'equivalenza tra (b) e (c) è poi immediata: basta osservare che, grazie a (2.2), la misura ν verifica la relazione $\xi_0(\nu) = \mu$ e quindi, per la definizione stessa del nucleo Q, coincide con μQ se, e solo se, verifica la condizione (b).

– Infine l'equivalenza tra (c) e (d) non è che un caso particolare di un noto criterio riguardante la legge condizionale in presenza di un legame funzionale. Infatti, nel caso presente, ha luogo il legame funzionale $X_0 = \xi_0 \circ X$ e inoltre, denotando ancora con H_0 il nucleo, dallo spazio $(F, \mathcal{F}_0)$ a $(E_0, \mathcal{E}_0)$, associato a ξ_0, si vede subito che il nucleo composto QH_0 coincide col nucleo identità relativo a $(E_0, \mathcal{E}_0)$: basta osservare che, per ogni misura di probabilità λ su $(E_0, \mathcal{E}_0)$, la misura λQH_0, essendo l'immagine di λQ mediante ξ_0, coincide con λ per la definizione stessa del nucleo Q. □



22 Prodotto di misure di probabilità

1. Osservazioni e risultati preliminari

Data una famiglia $\big((E_i, \mathcal{E}_i, \mu_i)\big)_{i\in I}$ di spazi probabilizzati (con I arbitrario insieme non vuoto), ci proponiamo di dimostrare che esiste una e una sola *misura prodotto* della famiglia $(\mu_i)_{i\in I}$. Com'è noto, con questa locuzione s'intende una misura di probabilità definita sulla tribù prodotto $\bigotimes_{i\in I}\mathcal{E}_i$ e che, a ciascun "rettangolo misurabile" della forma $\prod_{i\in I} A_i$, con $A_i \in \mathcal{E}_i$ per ogni i, e con $A_i \neq E_i$ solo per un numero finito di indici i, attribuisca come misura il numero $\prod_{i\in I}\mu_i(A_i)$.

È chiaro che due misure prodotto della stessa famiglia $(\mu_i)_{i\in I}$ sono necessariamente identiche perché coincidono sulla classe dei rettangoli sopra descritti, la quale è una base per la tribù prodotto. Rimane da provare l'esistenza. A questo scopo, cominciamo col ricordare che, se $(E, \mathcal{E}, \mu)$, $(F, \mathcal{F}, \nu)$ sono due spazi probabilizzati, e se N denota il banale nucleo-markoviano, da $(E, \mathcal{E})$ a $(F, \mathcal{F})$, definito da $N(x, \cdot) = \nu$ per ogni x di E, allora la misura prodotto $\mu \otimes \nu$ è la trasformata di μ mediante il nucleo $\widetilde{N}$. In secondo luogo, si prova facilmente, per induzione, che, data una famiglia finita $\big((E_i, \mathcal{E}_i, \mu_i)\big)_{0\le i\le n}$ di spazi probabilizzati, esiste sempre la misura prodotto $\mu_0 \otimes \cdots \otimes \mu_n$. Per $n = 2$, il risultato si riduce a quello già ricordato. Per passare da n a $n+1$, basta osservare che la misura prodotto $\mu_0 \otimes \cdots \otimes \mu_{n+1}$ è l'immagine di $(\mu_0 \otimes \cdots \otimes \mu_n) \otimes \mu_{n+1}$ mediante l'applicazione di $(E_0 \times \cdots \times E_n) \times E_{n+1}$ in $E_0 \times \cdots \times E_{n+1}$ che ad ogni elemento $\big((x_0, \cdots, x_n), x_{n+1}\big)$ del primo spazio associa $(x_0, \ldots, x_{n+1})$.

Sarà poi utile anche l'osservazione seguente.

(1.1) Osservazione. Se ν è una misura prodotto per la famiglia $(\mu_i)_{i\in I}$, e se σ è un'applicazione biunivoca di un insieme J sull'insieme I, allora anche la famiglia $(\mu_{\sigma(j)})_{j\in J}$ ammette una misura prodotto: precisamente, essa ammette come misura prodotto l'immagine di ν mediante l'applicazione dello spazio $\prod_{i\in I} E_i$ sullo spazio $\prod_{j\in J} E_{\sigma(j)}$ che, ad ogni elemento $(x_i)_{i\in I}$ del primo spazio, associa $(x_{\sigma(j)})_{j\in J}$.

Ciò premesso, dimostriamo un primo risultato parziale:

(1.2) Lemma. *Ogni famiglia numerabile di misure di probabilità ammette una misura prodotto.*

Dimostrazione. Naturalmente, grazie all'Osservazione (1.1), basterà limitarsi al caso di una famiglia numerabile *infinita*, anzi addirittura al caso di una *successione*. Consideriamo dunque una successione $\big((E_n, \mathcal{E}_n, \mu_n)\big)_{n\in\mathbb{N}}$ di spazi probabilizzati. Per ogni intero positivo n, denotiamo con N_n il banale nucleo markoviano, relativo alla coppia di spazi misurabili

$$(E_0 \times \cdots \times E_n,\ \mathcal{E}_0 \otimes \cdots \otimes \mathcal{E}_n), \quad (E_{n+1},\ \mathcal{E}_{n+1}),$$

così definito:

$$N_n\big((x_0,\dots,x_n),\,\cdot\,\big)=\mu_{n+1}.$$

Consideriamo, sullo spazio $\prod_{n\geq 0}E_n$, la successione $(\xi_n)_{n\geq 0}$ delle proiezioni canoniche. Grazie al teorema di Ionescu Tulcea, esiste un'unica misura di probabilità ν su $\bigotimes_{n\geq 0}\mathcal{E}_n$ secondo la quale ξ_0 abbia come legge μ_0 e, per ogni intero positivo n, il nucleo N_n sia una versione della legge condizionale di ξ_{n+1} rispetto al blocco $[\xi_0,\dots,\xi_n]$. Per provare che la misura ν così caratterizzata è la desiderata misura prodotto della successione $(\mu_n)_{n\geq 0}$, basta provare che, per ogni intero positivo n e ogni famiglia $(A_j)_{0\leq j\leq n}$ d'insiemi, con $A_j\in\mathcal{E}_j$ per $0\leq j\leq n$, si ha

$$\nu\Big(\textstyle\bigcap_{0\leq j\leq n}\{\xi_j\in A_j\}\Big)=\prod_{j=0}^{n}\mu_j(A_j).$$

Ciò si verifica immediatamente per induzione. □

2. Il risultato generale di esistenza

(2.1) Teorema. *Data una famiglia $\big((E_i,\mathcal{E}_i,\mu_i)\big)_{i\in I}$ di spazi probabilizzati (con I arbitrario insieme non vuoto), esiste sempre una (e una sola) misura prodotto della famiglia $(\mu_i)_{i\in I}$.*

Dimostrazione. Poniamo

$$\Omega=\textstyle\prod_{i\in I}E_i,\qquad \mathcal{A}=\bigotimes_{i\in I}\mathcal{E}_i$$

e denotiamo con $(\xi_i)_{i\in I}$ la famiglia delle proiezioni canoniche di Ω sui singoli fattori. Inoltre, per ogni parte J di I, numerabile e non vuota, posto

$$F_J=\textstyle\prod_{i\in J}E_i,\qquad \mathcal{F}_J=\bigotimes_{i\in J}\mathcal{E}_i,\qquad \mathcal{G}_J=\mathcal{T}\big([\xi_i]_{i\in J}\big),$$

denotiamo con ν_J la misura prodotto (esistente in virtù di (1.2)) della famiglia numerabile $(\mu_i)_{i\in J}$ e con H_J il nucleo di composizione col blocco $[\xi_i]_{i\in J}$. Poiché l'immagine di Ω mediante questo blocco è l'intero spazio F_J, sappiamo che H_J è un nucleo, da $(\Omega,\mathcal{G}_J)$ a $(F_J,\mathcal{F}_J)$, dotato d'inverso. Posto $P_J=\nu_J H_J^{-1}$, possiamo dire che P_J è l'unica misura di probabilità su $(\Omega,\mathcal{G}_J)$ secondo la quale il blocco $[\xi_i]_{i\in J}$ abbia come legge ν_J. Se ora J,L sono due parti di I, numerabili e non vuote, con $J\subset L$, è facile riconoscere che la restrizione di P_L alla tribù $\mathcal{G}_J$ dà anch'essa al blocco $[\xi_i]_{i\in J}$ come legge ν_J (e quindi coincide con P_J). Si tratta semplicemente di verificare che, se $(A_i)_{i\in J}$ è una famiglia d'insiemi tale che si abbia $A_i\in\mathcal{E}_i$ per ogni elemento i di J e che l'insieme $\{i\in J:\,A_i\neq E_i\}$ sia finito, si ha

$$P_L\big(\textstyle\bigcap_{i\in J}\{\xi_i\in A_i\}\big)=\prod_{i\in J}\mu_i(A_i).$$

Ma ciò è evidente: infatti, se si pone $A_i=E_i$ per ogni elemento i di $L\backslash J$, l'eguaglianza da dimostrare si può mettere nella forma

$$P_L\big(\textstyle\bigcap_{i\in L}\{\xi_i\in A_i\}\big)=\prod_{i\in L}\mu_i(A_i).$$

Proviamo ora che la tribù $\mathcal{A}$ coincide con l'unione delle tribù della forma $\mathcal{G}_J$, con J parte di I numerabile e non vuota. Indicata con $\mathcal{H}$ questa unione, osserviamo che essa è contenuta in $\mathcal{A}$ e contiene la classe degli insiemi della forma $\{\xi_i \in A\}$ con $i \in I$ e $A \in \mathcal{E}_i$. Tutto dunque è ridotto a provare che $\mathcal{H}$ è una tribù. Ma per questo basta osservare che, comunque si assegni una successione $(A_n)_{n\geq 1}$ di elementi di $\mathcal{H}$, se, per ciascun indice n, si sceglie una parte J_n di I, numerabile e non vuota, tale che si abbia $A_n \in \mathcal{G}_{J_n}$, e si denota con J l'unione degli insiemi J_n, si ha $A_n \in \mathcal{G}_J$ per ogni n.

Osserviamo che, se A è un fissato elemento della tribù $\mathcal{A}$, le misure P_J corrispondenti alle parti J di I, numerabili e non vuote, tali che A appartenga a $\mathcal{G}_J$ assumono tutte lo stesso valore su A. Infatti, se P_J, P_L sono due siffatte misure, allora, poiché la misura $P_{J\cup L}$ le prolunga entrambe, risulta $P_J(A) = P_{J\cup L}(A) = P_L(A)$. Esiste dunque, sulla tribù $\mathcal{A}$, una funzione P che prolunga simultaneamente tutte le P_J. Si verifica subito che P è una misura. Infatti, data una successione $(A_n)_{n\geq 1}$ di elementi di $\mathcal{A}$, a due a due disgiunti, si può trovare una parte J di I, numerabile e non vuota, tale che tutti gli A_n appartengano alla tribù $\mathcal{G}_J$. Allora, poiché P coincide con P_J su $\mathcal{G}_J$, si ha

$$P\big(\bigcup_{n\geq 1} A_n\big) = \sum_{n\geq 1} P(A_n).$$

Per completare la dimostrazione del teorema, basta osservare che la misura P così costruita è una misura prodotto per l'assegnata famiglia $(\mu_i)_{i\in I}$: infatti, se $(A_i)_{i\in J}$ è una famiglia finita e non vuota d'insiemi, tale che si abbia $A_i \in \mathcal{E}_i$ per ogni elemento i di J, allora, poiché la misura P coincide con P_J su $\mathcal{G}_J$, si ha

$$P\left(\bigcap_{i\in J}\{\xi_i \in A_i\}\right) = \prod_{i\in J} \mu_i(A_i). \qquad \square$$

23

Generalizzazione degli spazi polacchi.

1. Spazi polacchi e loro sottospazi

(1.1) Definizione. Si chiama brevemente *spazio polacco* uno spazio topologico metrizzabile E, di tipo numerabile, tale che esista una distanza, compatibile con la sua topologia, per la quale E sia uno spazio metrico completo.

Sono immediate le due proposizioni seguenti.

(1.2) Proposizione. *Il prodotto di una famiglia numerabile di spazi polacchi è uno spazio polacco.*

(1.3) Proposizione. *Ogni sottospazio chiuso di uno spazio polacco è uno spazio polacco.*

Allo scopo di ottenere ulteriori risultati riguardanti i sottospazi polacchi di un assegnato spazio polacco, cominciamo col dimostrare la proposizione seguente, la quale riguarda i sottospazi aperti di uno spazio metrizzabile.

(1.4) Proposizione. *Siano E uno spazio topologico metrizzabile e U un sottospazio aperto di E. Allora U è omemorfo a un sottospazio chiuso di $E \times \mathbb{R}$.*

Dimostrazione. Sia d una distanza compatibile con la topologia di E. La tesi è immediata se U è eguale all'intero spazio E. Supporremo dunque $U \neq E$. Consideriamo allora la funzione reale continua f così definita su E:

$$f(x) = d(x, U^c) \quad \text{per } x \in E.$$

Si ha allora $U = \{f > 0\}$. Consideriamo poi l'applicazione g di U in $E \times \mathbb{R}$ così definita:

$$g(x) = \big(x, 1/f(x)\big) \quad \text{per } x \in U.$$

Si ha allora

$$g(U) = \big\{(x,t) \in E \times \mathbb{R} : tf(x) = 1\big\}.$$

Dunque $g(U)$ è un insieme chiuso di $E \times \mathbb{R}$. Inoltre l'applicazione g è un omeomorfismo di U su $g(U)$: essa è infatti un'applicazione continua e iniettiva di U su $g(U)$, la cui inversa è la restrizione a $g(U)$ della proiezione canonica di $E \times \mathbb{R}$ sul primo fattore. □

Dalla proposizione appena dimostrata discende il corollario seguente.

(1.5) Corollario. *Ogni sottospazio aperto di uno spazio polacco è uno spazio polacco.*

Dimostrazione. Siano E uno spazio polacco e U un sottospazio aperto di E. Dalla proposizione precedente discende che U è omeomorfo a un sottospazio chiuso dello spazio polacco $E \times \mathbb{R}$, e dunque a uno spazio polacco (si veda (1.3)). □

Ci proponiamo ora di dimostrare un'importante caratterizzazione dei sottospazi polacchi di un assegnato spazio polacco. A questo scopo, cominciamo col dimostrare il risultato seguente, relativo a un arbitrario spazio topologico separato.

(1.6) Proposizione. *Siano E uno spazio topologico separato e $(A_n)_{n\geq 1}$ una successione di sottospazi di E. Allora il sottospazio $\bigcap_{n\geq 1} A_n$ è omeomorfo a un sottospazio chiuso dello spazio prodotto $\prod_{n\geq 1} A_n$.*

Dimostrazione. Denotiamo con D la diagonale dello spazio $E^{\mathbb{N}^*}$. Basta allora osservare che l' omeomorfismo $x \mapsto (x, x, \ldots)$ di E su D trasforma l'insieme $\bigcap_{n\geq 1} A_n$ nell'insieme $D \cap \prod_{n\geq 1} A_n$. □

Il risultato appena dimostrato ammette il corollario seguente.

(1.7) Corollario. *Siano F uno spazio polacco e E un sottospazio di F. Si supponga che E sia l'intersezione di una successione $(V_n)_{n\geq 1}$ d'insiemi aperti di F. Allora E è polacco.*

Dimostrazione. Lo spazio prodotto $\prod_{n\geq 1} V_n$ è polacco, tale essendo ciascuno dei V_n (si veda (1.2) e (1.5)). Inoltre, grazie a (1.6), E è omeomorfo a un sottospazio chiuso di $\prod_{n\geq 1} V_n$. Da (1.3) discende dunque che E è polacco. □

Ci proponiamo ora d'invertire l'affermazione di (1.7). A questo scopo, è utile premettere l'osservazione seguente.

(1.8) Osservazione. Siano dati uno spazio topologico F, un filtro $\mathcal{F}$ su di esso e un sottospazio E di F che incontri ogni elemento di $\mathcal{F}$. È allora possibile considerare su E la *traccia* del filtro $\mathcal{F}$, ossia il filtro $\mathcal{E}$ costituito dagli insiemi della forma $A \cap E$, con $A \in \mathcal{F}$. Nello spazio F, è chiaro che $\mathcal{E}$ è una base di filtro, più fine del filtro $\mathcal{F}$. Inoltre, affinché un assegnato elemento di E sia, nello spazio E, limite del filtro $\mathcal{E}$, occorre e basta che esso sia, nello spazio F, limite della base di filtro $\mathcal{E}$.

Ciò premesso, siamo ora in grado di dimostrare l'inverso del risultato (1.7), ossia il risultato seguente.

(1.9) Teorema. *Siano F uno spazio polacco e E un sottospazio di F. Si supponga che E sia polacco. Allora E è l'intersezione di una successione d'insiemi aperti di F.*

In realtà, dimostreremo qualcosa di più: ossia il risultato seguente, del quale (1.9) è un immediato corollario.

(1.10) Teorema. *Siano F uno spazio topologico separato e E un sottospazio di F. Si supponga che esista su E una distanza d, compatibile con la topologia di E, che faccia di E uno spazio metrico completo. Allora esiste una successione $(V_n)_{n\geq 1}$ d'insiemi aperti di F tale che si abbia (denotando con $\overline{E}$ la chiusura di E in F)*

$$E = \overline{E} \cap \bigcap_{n\geq 1} V_n. \tag{1.11}$$

Dimostrazione. Per ogni parte non vuota A di E, denotiamo con $\mathrm{diam}(A)$ il diametro di A secondo la distanza d. Denotiamo poi con $\mathcal{V}$ la classe degli insiemi aperti di F e, per ogni intero n strettamente positivo, poniamo

$$\mathcal{V}_n = \big\{V \in \mathcal{V} : V \cap E \neq \emptyset,\ \mathrm{diam}(V \cap E) < 1/n\big\}.$$

Denotiamo infine con V_n l'unione degli insiemi appartenenti a $\mathcal{V}_n$. Proviamo che vale allora la relazione (1.11). Per ogni elemento x di E ed ogni intero n strettamente positivo, l'insieme non vuoto $\big\{y \in E : d(y,x) < 1/(2n)\big\}$, essendo un insieme aperto del sottospazio E e avendo diametro minore di $1/n$, è della forma $V \cap E$, con $V \in \mathcal{V}_n$. Perciò x appartiene a V_n. È così provato che il primo membro di (1.11) è contenuto nel secondo. Per provare l'inclusione opposta, consideriamo ora un elemento x del secondo membro e, detto $\mathcal{F}(x)$ il filtro degli intorni di x nello spazio F, denotiamo con $\mathcal{E}(x)$ la traccia di $\mathcal{F}(x)$ sul sottospazio E, cioè poniamo

$$\mathcal{E}(x) = \big\{V \cap E : V \in \mathcal{F}(x)\big\}.$$

Allora $\mathcal{E}(x)$ è un filtro su E che, secondo la distanza d, è un filtro di Cauchy. (Infatti l'appartenenza di x al secondo membro di (1.11) implica che, per ogni intero n strettamente positivo, x appartiene a V_n, sicché $V_n \cap E$ è, nello spazio E, un intorno di x di diametro minore di $1/n$.) Per la completezza dello spazio metrico (E,d), ne segue che, in questo spazio, il filtro $\mathcal{E}(x)$ converge verso un elemento y di E. Grazie all'Osservazione (1.8), ciò equivale a dire che, nello spazio F, la base di filtro $\mathcal{E}(x)$ converge verso y. D'altra parte, essa converge anche verso x (in quanto è più fine del filtro $\mathcal{F}(x)$ degli intorni di x). Dunque, essendo lo spazio F separato, il punto x coincide con y e quindi appartiene a E. Il teorema è così dimostrato. □

Spazi misurabili regolari

1. Definizione e primo esempio di spazio regolare

(1.1) Definizione. Uno spazio misurabile $(F, \mathcal{F})$ sarà detto *regolare* (rispetto al concetto di legge condizionale) se, comunque si prendano uno spazio probabilizzato $(\Omega, \mathcal{A}, P)$ e, su di esso, una variabile aleatoria Y a valori in $(F, \mathcal{F})$ e una variabile aleatoria X a valori in un arbitrario spazio misurabile, esiste sempre una versione della legge condizionale di Y rispetto a X.

Un primo importante esempio di spazio misurabile regolare è fornito dal teorema seguente.

(1.2) Teorema. *Lo spazio misurabile $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ è regolare.*

Dimostrazione. Assegnate, su uno spazio probabilizzato $(\Omega, \mathcal{A}, P)$, una variabile aleatoria X, a valori in un arbitrario spazio misurabile $(E, \mathcal{E})$, e una variabile aleatoria reale Y, si tratta di esibire una versione della legge condizionale di Y rispetto a X. Poiché la classe costituita dagli intervalli della forma $]-\infty, r]$, con r numero razionale, è una base per la tribù $\mathcal{B}(\mathbb{R})$, basterà costruire un nucleo markoviano N, da $(E, \mathcal{E})$ a $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, tale che, per ogni numero razionale r, la speranza condizionale

$$P\big[I_{]-\infty,r]} \circ Y \,\big|\, X\big] \tag{1.3}$$

ammetta come versione la variabile aleatoria $\big(NI_{]-\infty,r]}\big) \circ X$. A questo scopo, scegliamo, per ogni numero razionale r, una versione della speranza condizionale (1.3): per il lemma di Doob, essa sarà della forma $f_r \circ X$, con f_r funzione reale misurabile su $(E, \mathcal{E})$. Grazie alle proprietà della speranza condizionale, ciascuno dei tre insiemi seguenti avrà, secondo la legge μ di X, misura eguale a 1:

$$A = \bigcap_{r,s\in\mathbb{Q},\, r<s} \{f_r \le f_s\}, \qquad B = \bigcap_{r\in\mathbb{Q}} \{f_r = \inf_{n\in\mathbb{N}} f_{r+1/n}\},$$

$$C = \big\{\sup_{n\in\mathbb{N}} f_n = 1,\ \inf_{n\in\mathbb{N}} f_{-n} = 0\big\}.$$

Ponendo $D = A \cap B \cap C$, si avrà ancora $\mu(D) = 1$. Per ogni elemento x di D, denotiamo con $N(x, \cdot)$ la legge su $\mathcal{B}(\mathbb{R})$ caratterizzata dall'eguaglianza

$$N\big(x,]-\infty, r]\big) = f_r(x) \quad \text{per ogni numero razionale } r. \tag{1.4}$$

Poniamo inoltre $N(x, \cdot) = \varepsilon_0$ per ogni elemento x di $E \setminus D$. Rimane così definito un nucleo N che risolve il problema. Infatti, per ogni numero razionale r, la funzione $NI_{]-\infty,r]}$, cioè $x \mapsto N\big(x,]-\infty, r]\big)$, è misurabile su $(E, \mathcal{E})$ perché è costante su $E \setminus D$ e coincide su D con la funzione f_r (alla quale è dunque equivalente secondo μ); pertanto la variabile aleatoria $\big(NI_{]-\infty,r]}\big) \circ X$, essendo equivalente, secondo P, alla variabile aleatoria $f_r \circ X$, è, al pari di questa, una versione della speranza condizionale (1.3). □

2. Proprietà della classe degli spazi misurabili regolari

(2.1) Definizione. Dati due spazi misurabili $(F, \mathcal{F})$, $(F_0, \mathcal{F}_0)$, si dirà che $(F_0, \mathcal{F}_0)$ è *subordinato* a $(F, \mathcal{F})$ se esistono un'applicazione misurabile f di $(F_0, \mathcal{F}_0)$ in $(F, \mathcal{F})$ e un'applicazione misurabile g di $(F, \mathcal{F})$ in $(F_0, \mathcal{F}_0)$, tali che $g \circ f$ coincida con l'applicazione identica di F_0.

(2.2) Esempi. (a) Si supponga che $(F_0, \mathcal{F}_0)$ sia un sottospazio *misurabile* di $(F, \mathcal{F})$, ossia che l'insieme F_0 appartenga alla tribù $\mathcal{F}$ e che la tribù $\mathcal{F}_0$ sia costituita dagli elementi di $\mathcal{F}$ contenuti in $\mathcal{F}_0$. Allora $(F_0, \mathcal{F}_0)$ è subordinato a $(F, \mathcal{F})$. Per riconoscerlo, basta prendere come applicazione f l'iniezione canonica di F_0 in F e come applicazione g una qualsiasi applicazione di F su F_0 che lasci fisso ciascun punto di F_0 e che sia costante su $F \setminus F_0$.

(b) Se due spazi misurabili sono tra loro *isomorfi* (cioè se esiste un'applicazione biunivoca del primo nel secondo che sia misurabile insieme con la sua inversa), allora ciascuno dei due spazi è subordinato all'altro.

L'interesse della nozione introdotta nella Definizione (2.1) è dovuto alla proposizione seguente.

(2.3) Proposizione. *Se uno spazio misurabile $(F, \mathcal{F})$ è regolare, tale è ogni spazio misurabile $(F_0, \mathcal{F}_0)$ ad esso subordinato.*

Dimostrazione. Siano f, g come nella Definizione (2.1). Siano date, su un medesimo spazio probabilizzato, una variabile aleatoria Y, a valori in $(F_0, \mathcal{F}_0)$, e una variabile aleatoria X, a valori in un arbitrario spazio misurabile. Allora $f \circ Y$, essendo una variabile aleatoria a valori nello spazio regolare $(F, \mathcal{F})$, ammette una versione della legge condizionale rispetto a X. Dunque anche la variabile Y, che coincide con $g \circ (f \circ Y)$, ammette una versione della legge condizionale rispetto a X. □

La proposizione che segue si ottiene facilmente impiegando il concetto di *nucleo di Ionescu Tulcea*.

(2.4) Proposizione. *Uno spazio misurabile che sia il prodotto di una famiglia numerabile di spazi misurabili regolari è a sua volta regolare.*

Dimostrazione. Uno spazio misurabile che sia il prodotto di una famiglia *finita* di spazi misurabili regolari si può sempre pensare come un sottospazio misurabile di uno spazio misurabile che sia il prodotto di una famiglia *infinita numerabile* di spazi misurabili regolari (e che sia quindi isomorfo al prodotto di una *successione* di spazi misurabili regolari). Si vede dunque (tenendo conto di (2.2)) che è lecito ridursi al caso di uno spazio misurabile che sia il prodotto di una successione $((E_n, \mathcal{E}_n))_{n \geq 1}$ di spazi misurabili regolari. Si tratta allora di provare che, comunque si assegnino uno spazio probabilizzato $(\Omega, \mathcal{A}, P)$ e, su di esso, una successione $(X_n)_{n \geq 1}$ di variabili

aleatorie, tale che, per ogni n, lo spazio d'arrivo di X_n sia $(E_n, \mathcal{E}_n)$, e un'ulteriore variabile aleatoria X_0, a valori in un arbitrario spazio misurabile $(E_0, \mathcal{E}_0)$, esiste una versione della legge condizionale del blocco $[X_n]_{n\geq 1}$ rispetto a X_0 o, ciò ch'è lo stesso, esiste una versione della legge condizionale del blocco $[X_n]_{n\geq 0}$ rispetto a X_0. A questo scopo, scegliamo, per ogni intero positivo n, una versione N_n della legge condizionale di X_{n+1} rispetto al blocco $[X_0, \ldots, X_n]$, e denotiamo con Q il nucleo di Ionescu Tulcea associato alla successione $(N_n)_{n\geq 0}$. Allora, com'è ben noto, Q è una versione della legge condizionale del blocco $[X_n]_{n\geq 0}$ rispetto a X_0. La proposizione è così dimostrata. □

(2.5) Corollario. *Ogni spazio polacco (munito della propria tibù boreliana) è uno spazio misurabile regolare.*

Dimostrazione. Infatti ogni spazio polacco è omeomorfo a un sottospazio boreliano (più precisamente, a un G_δ) dello spazio $[0,1]^{\mathbb{N}}$. □

Sistemi proiettivi di misure di probabilità

1. Definizione di sistema proiettivo

Data una famiglia $\big((E_i, \mathcal{E}_i)\big)_{i\in I}$ di spazi misurabili, avente come insieme degli indici un arbitrario insieme I infinito, poniamo

$$\Omega = \prod_{i\in I} E_i, \qquad \mathcal{A} = \bigotimes_{i\in I} \mathcal{E}_i$$

e denotiamo con $(\xi_i)_{i\in I}$ la famiglia delle proiezioni canoniche di Ω sui singoli fattori. Indichiamo inoltre con $\mathcal{H}$ l'insieme di tutte le parti di I finite e non vuote e, per ogni siffatta parte H, poniamo

$$\mathcal{G}_H = \mathcal{T}\left([\xi_i]_{i\in H}\right).$$

Supponiamo infine assegnata, per ogni elemento H di $\mathcal{H}$, una misura di probabilità P_H sullo spazio $(\Omega, \mathcal{G}_H)$.

(1.1) Definizione. Nelle ipotesi e con le notazioni sopra introdotte, si dice che la famiglia di misure di probabilità $(P_H)_{H\in\mathcal{H}}$ è un *sistema proiettivo*, relativo all'assegnata famiglia di spazi misurabili, se, per ogni coppia H, K di elementi di $\mathcal{H}$, con $H \subset K$, la misura P_K prolunga la misura P_H. Si dice poi che una misura di probabilità su $\mathcal{A}$ è *limite proiettivo* del sistema proiettivo $(P_H)_{H\in\mathcal{H}}$ se essa prolunga tutte le P_H. È chiaro che, se un sistema proiettivo ammette limite proiettivo, questo è unico.

(1.2) Osservazione. Sia J un insieme equipotente a I, e sia φ un'applicazione biunivoca di J su I. Si denoti inoltre con Φ l'applicazione (biunivoca e bimisurabile) dello spazio $(\Omega, \mathcal{A})$ nello spazio

$$\left(\prod_{j\in J} E_{\varphi(j)},\ \bigotimes_{j\in J} \mathcal{E}_{\varphi(j)}\right)$$

che ad ogni elemento $(x_i)_{i\in I}$ del primo spazio associa $(x_{\varphi(j)})_{j\in J}$. Allora, per ogni sistema proiettivo $(P_H)_H$, relativo alla famiglia di spazi $\big((E_i, \mathcal{E}_i)\big)_{i\in I}$, si può costruire un sistema proiettivo $(Q_K)_K$, relativo alla famiglia di spazi $\big((E_{\varphi(j)}, \mathcal{E}_{\varphi(j)})\big)_{j\in J}$, ponendo, per ogni parte K di J, finita e non vuota, $Q_K = \Phi\left(P_{\varphi(K)}\right)$. Si riconosce immediatamente che, se esiste un limite proiettivo per il sistema $(P_H)_H$, allora la sua immagine mediante Φ è limite proiettivo per il sistema $(Q_K)_K$.

Nel caso particolare in cui l'assegnata famiglia di spazi misurabili sia una successione, si può introdurre, accanto alla nozione di sistema proiettivo, una nozione formalmente diversa (e un po' più semplice), ma sostanzialmente equivalente: quella di *successione proiettiva.*

(1.3) Definizione. Data una successione $\big((E_n, \mathcal{E}_n)\big)_{n\geq 0}$ di spazi misurabili, poniamo

$$\Omega = \prod_{n\geq 1} E_n, \qquad \mathcal{A} = \bigotimes_{n\geq 0} \mathcal{E}_n$$

e denotiamo con $(\xi_n)_{n\geq 0}$ la successione delle proiezioni canoniche di Ω sui singoli fattori. Inoltre, per ogni intero positivo n, poniamo $\mathcal{G}_n = \mathcal{T}(\xi_0, \ldots, \xi_n)$. Supponiamo infine assegnata, per ogni intero positivo n, una misura di probabilità P_n sullo spazio $(\Omega, \mathcal{G}_n)$. Allora la successione $(P_n)_{n\geq 0}$ si dice *proiettiva* se, per ogni n, la misura P_{n+1} è un prolungamento di P_n. Si dice poi che una misura di probabilità su $\mathcal{A}$ è un *limite proiettivo* della successione proiettiva $(P_n)_{n\geq 1}$ se essa prolunga tutte le P_n.

(1.4) Osservazione. (a) Nelle ipotesi della definizione precedente, si supponga assegnata una successione proiettiva $(P_n)_{n\geq 0}$. Se, per ogni parte H di $\mathbb{N}$, finita e non vuota, si sceglie un intero n tale che si abbia $H \subset \{0, \ldots, n\}$, e si denota con P'_n la restrizione di P_n alla tribù generata dal blocco $[\xi_j]_{j\in H}$, si definisce, senza ambiguità, una famiglia $(P'_H)_H$ di misure di probabilità che costituisce un sistema proiettivo nel senso della Definizione (1.1). Lo chiameremo il sistema proiettivo *associato* all'assegnata successione proiettiva $(P_n)_{n\geq 0}$. È chiaro che, affinché una misura di probabilità P sullo spazio $(\Omega, \mathcal{A})$ sia limite proiettivo per esso, occorre e basta che P sia limite proiettivo per l'assegnata successione proiettiva.

(b) Inversamente, si supponga assegnato un arbitrario sistema proiettivo $(P_H)_H$ relativo alla successione $\big((E_n, \mathcal{E}_n)\big)_{n\geq 0}$ di spazi misurabili. Per ogni intero positivo n, si ponga $P'_n = P_{\{0,\ldots,n\}}$. Allora è chiaro che la successione $(P'_n)_{n\geq 0}$ è proiettiva e che il sistema proiettivo ad essa associato coincide con l'assegnato sistema proiettivo $(P_H)_H$.

2. Esistenza del limite proiettivo

(2.1) Teorema. *Data una famiglia infinita di spazi misurabili regolari (rispetto al concetto di legge condizionale), ogni sistema proiettivo di leggi ad essa relativo ammette un (unico) limite proiettivo.*

Dimostrazione. Ci limiteremo a dimostrare il teorema nel caso di una famiglia *numerabile*. (L'estensione al caso generale si otterrà con un ragionamento analogo a quello impiegato a proposito delle misure prodotto.) Consideriamo dunque una famiglia infinita numerabile di spazi misurabili regolari. Grazie all'Osservazione (1.2), è lecito, senza ledere la generalità, supporre che la famiglia in questione sia una *successione* $\big((E_n, \mathcal{E}_n)\big)_{n\geq 0}$. Grazie all'Osservazione (1.4), sarà in realtà sufficiente provare che ogni *successione proiettiva* relativa alla successione $\big((E_n, \mathcal{E}_n)\big)_{n\geq 0}$ è dotata di limite proiettivo. Sia dunque $(P_n)_{n\geq 0}$ una tal successione. Poniamo, come al solito,

$$\Omega = \prod_{n\geq 0} E_n, \qquad \mathcal{A} = \bigotimes_{n\geq 0} \mathcal{E}_n$$

e denotiamo con $(\xi_n)_{n\geq 0}$ la successione delle proiezioni canoniche di Ω sui singoli fattori. Poniamo inoltre, per ogni intero positivo n,

$$\mathcal{G}_n = \mathcal{T}(\xi_0, \ldots, \xi_n), \qquad \nu_n = [\xi_0, \ldots, \xi_n](P_n).$$

Fissato un intero positivo n, mettiamoci sullo spazio $(\Omega,\ \mathcal{G}_{n+1},\ P_{n+1})$. Poiché lo spazio misurabile $(E_{n+1},\ \mathcal{E}_{n+1})$ è regolare, è possibile scegliere un nucleo N_n che sia, secondo P_{n+1}, una versione della legge condizionale di ξ_{n+1} rispetto a $[\xi_0,\dots,\xi_n]$. Grazie al teorema di Ionescu Tulcea, esiste, sulla tribù $\mathcal{A}$, un'unica misura di probabilità P secondo la quale la variabile aleatoria ξ_0 ammetta ν_0 come legge e, per ciascun intero positivo n, il nucleo N_n sia una versione della legge condizionale di ξ_{n+1} rispetto al blocco $[\xi_0,\dots,\xi_n]$. Vogliamo provare che la misura P così caratterizzata è limite proiettivo dell'assegnata successione proiettiva. A questo scopo, osserviamo che il nucleo, da $(\Omega,\mathcal{G}_n)$ a $(E_0\times\cdots\times E_n,\ \mathcal{E}_0\otimes\cdots\otimes\mathcal{E}_n)$, associato al blocco $[\xi_0,\dots,\xi_n]$ ammette inverso, sicché due misure di probabilità su $(\Omega,\mathcal{G}_n)$ che diano la stessa legge a questo blocco sono identiche. Dunque, per provare che la misura P prolunga tutte le misure P_n, basterà verificare che, per ogni intero positivo n, essa dà al blocco $[\xi_0,\dots,\xi_n]$ la stessa legge ν_n che gli dà P_n:

$$[\xi_0,\dots,\xi_n](P)=\nu_n. \tag{2.2}$$

La tesi è vera per $n=0$ (per la definizione stessa di P). Supposto che essa sia vera per un certo intero positivo n, dimostriamo che si ha anche $[\xi_0,\dots,\xi_{n+1}](P)=\nu_{n+1}$. In modo equivalente, si tratta di provare l'eguaglianza

$$\big[[\xi_0,\dots,\xi_n],\ \xi_{n+1}\big](P)=\big[[\xi_0,\dots,\xi_n],\ \xi_{n+1}\big](P_{n+1}).$$

Questa discende dai due fatti seguenti.

(a) Secondo P, il blocco $[\xi_0,\dots,\xi_n]$ ha legge ν_n (per l'ipotesi induttiva (2.2)) e la variabile aleatoria ξ_{n+1} ammette N_n come versione della propria legge condizionale rispetto a $[\xi_0,\dots,\xi_n]$ (per la definizione stessa di P).

(b) Secondo P_{n+1}, il blocco $[\xi_0,\dots,\xi_n]$ ha legge ν_n (perché, grazie alla definizione di successione proiettiva, P_{n+1} prolunga P_n) e inoltre la variabile aleatoria ξ_{n+1} ammette N_n come versione della propria legge condizionale rispetto a $[\xi_0,\dots,\xi_n]$ (per la definizione stessa di N_n).

L'affermazione è così dimostrata. □

(2.3) Osservazione. Nella precedente dimostrazione, l'ipotesi di regolarità rispetto al concetto di legge condizionale non è stata sfruttata per lo spazio $(E_0,\mathcal{E}_0)$.

Introduzione alle catene di Markov

1. Definizione di catena di Markov

In tutto il seguito si suppongono fissati uno spazio misurabile $(E, \mathcal{E})$ e un nucleo markoviano N ad esso relativo.

Su uno spazio probabilizzato $(\Omega, \mathcal{A}, P)$ sia data una successione $(X_n)_{n\geq 0}$ di variabili aleatorie a valori in $(E, \mathcal{E})$. Si dice che il processo $X = [X_n]_{n\geq 0}$ è una *catena di Markov* compatibile col nucleo N se, per ogni n, è verificata la condizione seguente:

(a) Per ogni funzione f limitata e misurabile su $(E, \mathcal{E})$, la variabile aleatoria $(Nf)\circ X_n$ è una versione della speranza condizionale $P[f\circ X_{n+1} \mid X_0, \dots, X_n]$.

La condizione (a) si può tradurre, nel linguaggio delle leggi condizionali, in ciascuno dei due modi seguenti:

(a′) Una versione della legge condizionale di X_{n+1} rispetto a $[X_0, \dots, X_n]$ è il nucleo L_n, da $(E^{n+1}, \mathcal{E}^{\otimes n+1})$-a $(E, \mathcal{E})$, così definito:

$$L_n\big((x_0, \dots, x_n), \cdot\big) = N(x_n, \cdot).$$

(a″) Una versione della legge condizionale di X_{n+1} rispetto alla tribù generata da $X_0, \dots, X_n$ è il nucleo M_n, da $(\Omega, \mathcal{T}(X_0, \dots, X_n))$ a $(E, \mathcal{E})$, così definito:

$$M_n(\omega, \cdot) = L_n\big((X_0(\omega), \dots, X_n(\omega)), \cdot\big) = N\big(X_n(\omega), \cdot\big).$$

Si supponga ora che, sullo spazio $(\Omega, \mathcal{A}, P)$, sia assegnata anche una filtrazione

$$\mathcal{F} = (\mathcal{F}_n)_{n\geq 0}.$$

Allora si dice che il processo X è una catena di Markov compatibile col nucleo N e *con la filtrazione* $\mathcal{F}$ se X è adattato a $\mathcal{F}$ e verifica, per ogni n, la condizione seguente:

(b) Per ogni funzione f limitata e misurabile su $(E, \mathcal{E})$, la variabile aleatoria $(Nf)\circ X_n$ è una versione della speranza condizionale $P[f\circ X_{n+1} \mid \mathcal{F}_n]$.

Quest'ultima condizione si può così tradurre nel linguaggio delle leggi condizionali:

(b′) Lo stesso nucleo M_n sopra introdotto, se considerato come nucleo da $(\Omega, \mathcal{F}_n)$ a $(E, \mathcal{E})$, risulta una versione della legge condizionale di X_{n+1} rispetto alla tribù $\mathcal{F}_n$ (contenente la tribù generata da $X_0, \dots, X_n$).

È immediata la proposizione seguente.

(1.1) Proposizione. *Su uno spazio probabilizzato $(\Omega, \mathcal{A}, P)$ sia X una catena di Markov compatibile col nucleo N, e si denoti con μ la legge di X_0. Allora, per ogni intero positivo n, la legge di X_n coincide con μN^n.*

Dimostrazione. L'affermazione è ovvia per $n = 0$. D'altra parte, se si suppone che essa sia vera per l'intero n, si ha

$$\langle P,\, f \circ X_{n+1}\rangle = \langle P,\, (Nf)\circ X_n\rangle = \langle \mu N^n,\, Nf\rangle = \langle \mu N^{n+1},\, f\rangle.$$

L'affermazione rimane dunque provata per induzione. □

2. Realizzazione canonica

Consideriamo lo spazio canonico $E^{\mathbb{N}}$, munito della tribù prodotto $\mathcal{E}^{\otimes\mathbb{N}}$. Inoltre, per ogni intero positivo n, denotiamo con ξ_n la proiezione canonica di indice n di $E^{\mathbb{N}}$ su E.

Applicando il teorema di Ionescu Tulcea, si vede che, per ogni legge μ su $\mathcal{E}$, esiste un'unica misura di probabilità P_μ su $\mathcal{E}^{\otimes\mathbb{N}}$ secondo la quale il processo canonico $[\xi_n]_{n\geq 0}$ sia una catena di Markov compatibile col nucleo N e ammettente μ come "legge iniziale" (cioè come legge di ξ_0). Inoltre l'applicazione (iniettiva) $\mu \mapsto P_\mu$ è "indotta da un nucleo markoviano". In altri termini: se, per ogni elemento x di E, si pone $Q(x,\,\cdot\,) = P_{\varepsilon_x}$, la famiglia $Q = (Q(x,\,\cdot\,))_{x\in E}$ è un nucleo markoviano, relativo alla coppia di spazi misurabili $(E,\mathcal{E})$, $\big(E^{\mathbb{N}},\mathcal{E}^{\otimes\mathbb{N}}\big)$, e, per ogni legge μ su $\mathcal{E}$, si ha $P_\mu = \mu Q$.

Il nucleo Q si chiama la *realizzazione canonica* del nucleo N.

È facile riconoscere che, per un assegnato processo, il fatto di essere una catena di Markov compatibile col nucleo N e avente un'assegnata legge iniziale è una proprietà che dipende solo dalla legge del processo (pensato come variabile aleatoria a valori nello spazio canonico). Più precisamente:

(2.1) Proposizione. *Su uno spazio probabilizzato $(\Omega,\mathcal{A},P)$ sia assegnato un processo $X = [X_n]_{n\geq 0}$, avente $(E,\mathcal{E})$ come spazio degli stati. Denotiamo con ν la sua legge e con μ la sua legge iniziale, cioè poniamo*

$$\nu = X(P), \qquad \mu = X_0(P) = \xi_0(\nu).$$

Denotiamo inoltre con Q la realizzazione canonica del nucleo N.

Le condizioni che seguono sono allora equivalenti:

(a) Il processo X è una catena di Markov compatibile col nucleo N.

(b) Si ha $\nu = \mu Q$.

(c) Q è una versione della legge condizionale di X rispetto a X_0.

Dimostrazione. Affinché valga la condizione (a), occorre e basta che, per ogni n e ogni coppia f, g di funzioni misurabili e limitate sugli spazi E^{n+1}, E rispettivamente, si abbia

$$P\big[f(X_0,\dots,X_n)\, g(X_{n+1})\big] = P\big[f(X_0,\dots,X_n)\, Ng\,(X_n)\big],$$

ossia

$$\nu\big[f(\xi_0,\dots,\xi_n)\,g(\xi_{n+1})\big] = \nu\big[f(\xi_0,\dots,\xi_n)\,Ng\,(\xi_n)\big].$$

In altri termini: affinché valga la condizione (a), occorre e basta che la misura di probabilità ν renda il processo canonico una catena di Markov compatibile col nucleo N, ossia coincida con μQ (che è l'unica misura dotata di questa proprietà). È così provata l'equivalenza tra (a) e (b).

Per provare l'equivalenza tra (b) e (c), osserviamo che è $X_0 = \xi_0 \circ X$ e che, per ogni fissato elemento x di E, l'immagine di $Q(x,\cdot)$ mediante ξ_0 è eguale a ϵ_x. Ciò basta per concludere, grazie a un noto risultato sulla legge condizionale in presenza di un legame funzionale. □

(2.2) Corollario. *Sia $X = [X_n]_{n\geq 0}$ una catena di Markov compatibile col nucleo N. Allora, per ogni intero positivo n, una versione della legge condizionale di X_n rispetto a X_0 è il nucleo N^n.*

Dimostrazione. Con le notazioni di (2.1), si ha $X_n = \xi_n \circ X$ e Q è una versione della legge condizionale di X rispetto a X_0. Perciò una versione della legge condizionale di X_n rispetto a X_0 è QK_n, dove K_n denota il nucleo associato a ξ_n. Ma, per ogni legge μ su $\mathcal{E}$, si ha $\mu QK_n = \xi_n(\mu Q) = \mu N^n$ (dove l'eguaglianza finale discende da (1.1)) Si vede dunque che QK_n coincide con N^n. □

3. Proprietà forte di Markov

Assegnata, su uno spazio probabilizzato $(\Omega, \mathcal{A}, P)$, una filtrazione $(\mathcal{F}_n)_{n\geq 0}$, se T è un tempo-d'arresto ad essa relativo, si denota con $\mathcal{F}_T$ la tribù costituita dagli eventi A, appartenenti a $\bigvee_n \mathcal{F}_n$, tali che si abbia

$$A \cap \{T \leq n\} \in \mathcal{F}_n \quad \text{per ogni intero positivo } n.$$

Questa tribù è detta la tribù degli eventi *anteriori* a T. Essa si riduce a $\mathcal{F}_m$ quando T coincida con la costante m. (Ciò rende coerente la notazione adottata.)

In tutto il seguito, supporremo fissato un elemento ∂ di E e adotteremo la seguente

Convenzione: Se $X = [X_n]_{n\geq 0}$ è un processo avente $(E,\mathcal{E})$ come spazio degli stati, e se T è una variabile aleatoria discreta a valori in $\mathbb{N}\cup\{\infty\}$, denoteremo con X_T la variabile aleatoria che coincide con la costante ∂ su $\{T=\infty\}$ e che, per ogni intero positivo n, coincide con X_n su $\{T=n\}$.

(3.1) Teorema. *Su uno spazio probabilizzato $(\Omega, \mathcal{A}, P)$ sia X una catena di Markov compatibile col nucleo N e con la filtrazione $(\mathcal{F}_n)_{n\geq 0}$. Sia poi T un tempo-d'arresto, relativo a questa filtrazione, con $P\{T<\infty\} \neq 0$.*

Allora, secondo la misura di probabilità $P(\,\cdot\,|\{T<\infty\})$, il processo $[X_{T+n}]_{n\geq 0}$ è una catena di Markov compatibile col nucleo N e con la filtrazione $(\mathcal{F}_{T+n})_{n\geq 0}$.

Dimostrazione. Che il processo $[X_{T+n}]_{n\geq 0}$ sia adattato alla filtrazione $(\mathcal{F}_{T+n})_{n\geq 0}$ è facile da verificare. Per provare il resto, posto

$$H = \{T < \infty\}, \qquad P_H = P(\,\cdot\,|\,H)$$

e fissati un intero positivo n e una funzione f limitata e misurabile su $(E, \mathcal{E})$, occorre provare che si ha

$$P_H\big[f\circ X_{T+n+1}\,|\,\mathcal{F}_{T+n}\big] \stackrel{\circ}{=} (Nf)\circ X_{T+n}.$$

Pur di sostituire T con $T+n$, si può supporre, senza ledere la generalità, che sia $n = 0$. Si tratta allora di dimostrare che, per ogni elemento A di $\mathcal{F}_T$, risulta

$$\int_A f\circ X_{T+1}\,\mathrm{d}P_H = \int_A (Nf)\circ X_T\,\mathrm{d}P_H.$$

Moltiplicando per $P(H)$ entrambi i membri di questa eguaglianza, si ottiene l'eguaglianza equivalente

$$\int_{A\cap H} f\circ X_{T+1}\,\mathrm{d}P = \int_{A\cap H} (Nf)\circ X_T\,\mathrm{d}P.$$

Per provarla, basta verificare che, per ogni intero positivo m, si ha

$$\int_{A\cap\{T=m\}} f\circ X_{m+1}\,\mathrm{d}P = \int_{A\cap\{T=m\}} (Nf)\circ X_m\,\mathrm{d}P.$$

Ora, quest'ultima relazione discende dal fatto che l'evento $A\cap\{T=m\}$ appartiene a $\mathcal{F}_m$ e che X è una catena di Markov compatibile col nucleo N e con la filtrazione $(\mathcal{F}_n)_{n\geq 0}$. Il teorema è dunque dimostrato. □

(3.2) Corollario. *Su uno spazio probabilizzato $(\Omega, \mathcal{A}, P)$ sia X una catena di Markov compatibile col nucleo N e con la filtrazione $(\mathcal{F}_n)_{n\geq 0}$. Si denoti inoltre con Q la realizzazione canonica di N. Allora, per ogni intero positivo m, una versione della legge condizionale del processo traslato $[X_{m+n}]_{n\geq 0}$ rispetto a $\mathcal{F}_m$ è il nucleo K_m così definito:*

$$K_m(\omega,\,\cdot\,) = Q\big(X_m(\omega),\,\cdot\,\big).$$

In altri termini: per ogni funzione f limitata e misurabile sullo spazio $(E^{\mathbb{N}}, \mathcal{E}^{\otimes\mathbb{N}})$, una versione della speranza condizionale

$$P\big[f\circ[X_{m+n}]_{n\geq 0}\,|\,\mathcal{F}_m\big]$$

è la variabile aleatoria $Qf\circ X_m$.

Dimostrazione. Poiché X_m è la variabile aleatoria iniziale del processo $[X_{m+n}]_{n\geq 0}$, il quale, grazie al teorema precedente, è una catena di Markov compatibile col nucleo N e con la filtrazione $(\mathcal{F}_{m+n})_{n\geq 0}$, si può supporre, senza ledere la generalità, che sia $m = 0$ (nel qual caso, il processo traslato coincide col processo X stesso). Si tratta allora di dimostrare che, fissato un elemento H di $\mathcal{F}_0$, risulta

$$\int_H f\circ X\,\mathrm{d}P = \int_H (Qf)\circ X_0\,\mathrm{d}P.$$

Naturalmente, si può supporre H non trascurabile. Applicando allora il teorema precedente al tempo d'arresto T che coincide con 0 su H e con $+\infty$ altrove, si trova che, secondo P_H, il processo Y definito da $Y_n = X_{T+n}$ è una catena di Markov

compatibile col nucleo N, ossia (vedi (2.1)) ammette Q come legge condizionale rispetto a Y_0. In particolare, si ha $P_H\big[f \circ Y\big] = P_H\big[(Qf) \circ Y_0\big]$. Questa relazione equivale a quella da dimostrare (grazie al fatto che Y coincide su H con X). □

(3.3) Corollario. *Nelle stesse ipotesi del corollario precedente, le condizioni che seguono sono equivalenti:*

(a) Il processo traslato $[X_{m+n}]_{n\geq 0}$ è indipendente dalla tribù $\mathcal{F}_m$.

(b) La variabile aleatoria X_m è degenere.

Dimostrazione. (a) $\Rightarrow$ (b): se l'intero processo traslato è indipendente da $\mathcal{F}_m$, allora a maggior ragione la variabile aleatoria X_m è indipendente da $\mathcal{F}_m$, e quindi degenere.

(b) $\Rightarrow$ (a): se X_m è degenere, allora, per ogni funzione f limitata e misurabile sullo spazio canonico $E^{\mathbb{N}}$, la variabile aleatoria reale $(Qf) \circ X_m$ è degenere, ossia equivalente a una costante. Perciò la speranza condizionale $P\big[f \circ [X_{m+n}]_{n\geq 0} \,\big|\, \mathcal{F}_m\big]$ ammette come versione una costante, ossia la costante $\int f \, d\nu$, dove ν denota la legge del processo traslato. Per l'arbitrarietà di f, ciò vuol dire che il nucleo costante identificato a ν è una versione della legge condizionale del processo traslato rispetto a $\mathcal{F}_m$, ossia che questo processo è indipendente da $\mathcal{F}_m$. □

Come ulteriore applicazione del Teorema (3.1), dimostriamo infine la proposizione seguente, che ci sarà utile nella prossima sezione.

(3.4) Proposizione. *Si denoti con Q la realizzazione canonica del nucleo N, con θ il processo $[\xi_{n+1}]_{n\geq 0}$ sullo spazio canonico $\big(E^{\mathbb{N}}, \mathcal{E}^{\otimes\mathbb{N}}\big)$ e con Θ il nucleo di composizione con θ, ossia il nucleo, relativo allo spazio canonico, che opera sulle funzioni nel modo seguente: $\Theta f = f \circ \theta$. Si ha allora*

$$\left\{ \quad Q\Theta = NQ. \quad \right\} \tag{3.5}$$

Dimostrazione. Sia μ un'arbitraria legge su $(E, \mathcal{E})$. Se, mettendoci sullo spazio

$$\big(E^{\mathbb{N}}, \mathcal{E}^{\otimes\mathbb{N}}, \mu Q\big), \tag{3.6}$$

applichiamo il Teorema (3.1) al processo canonico $[\xi_n]_{n\geq 0}$ e al tempo d'arresto costantemente eguale a 1, troviamo che, sullo spazio (3.6), il processo θ è una catena di Markov compatibile col nucleo N. Inoltre questa catena ha come legge iniziale la legge di ξ_1, ossia μN. Ciò equivale a dire che la legge di θ secondo μQ è μNQ, cioè che, per ogni funzione f limitata e misurabile sullo spazio canonico $\big(E^{\mathbb{N}}, \mathcal{E}^{\otimes\mathbb{N}}\big)$, si ha

$$\langle \mu Q, f \circ \theta \rangle = \langle \mu NQ, f \rangle,$$

ossia

$$\langle \mu, Q\Theta f \rangle = \langle \mu, NQf \rangle.$$

Quest'ultima eguaglianza, grazie all'arbitrarietà di μ e di f, prova l'identità (3.5). □

(3.7) Osservazione. Nelle ipotesi della proposizione precedente, θ è semplicemente l'operatore di *slittamento*, ossia l'operatore che, ad ogni elemento $(x_n)_{n\geq 0}$ di $E^{\mathbb{N}}$, associa l'elemento $(x_{n+1})_{n\geq 0}$. Di questo operatore si può considerare, per ogni intero positivo m, la m–esima potenza di composizione θ^m. (Essa non è altro che il processo traslato $[\xi_{m+n}]_{n\geq 0}$.) Il nucleo di composizione con θ^m coincide evidentemente con la m–esima potenza di composizione Θ^m del nucleo Θ. Ragionando per induzione, è possibile dedurre da (3.5) che, per ogni intero positivo m, si ha $Q\Theta^m = N^m Q$.

4. Funzioni eccessive e funzioni invarianti

(4.1) Definizione. Chiameremo *eccessiva* (rispetto al nucleo N) una funzione numerica f, positiva e misurabile su $(E,\mathcal{E})$, tale che si abbia $f \geq Nf$. Una funzione eccessiva f si dirà poi *invariante* se verifica la precedente relazione col segno di eguaglianza.

(4.2) Osservazione. Se la funzione f è eccessiva, allora la successione $(N^n f)_{n\geq 0}$ è decrescente (come si riconosce immediatamente per induzione). È vera anche l'implicazione inversa: infatti la decrescenza della successione $(N^n f)_{n\geq 0}$ implica in particolare, per $n = 0$, la diseguaglianza $f \geq Nf$. In modo analogo si vede che f è invariante se, e solo se, la successione $(N^n f)_{n\geq 0}$ è costante.

(4.3) Osservazione. Si riconosce immediatamente che, se una funzione f è eccessiva, tale è anche ogni funzione della forma $f \wedge c$, con c costante reale positiva.

(4.4) Proposizione. *Sia f una funzione eccessiva e limitata. Inoltre, su uno spazio probabilizzato $(\Omega, \mathcal{A}, P)$, siano date una filtrazione $(\mathcal{F}_n)_{n\geq 0}$ e una catena di Markov X compatibile col nucleo N e con la filtrazione $(\mathcal{F}_n)_{n\geq 0}$. Allora, rispetto alla stessa filtrazione, il processo $[f \circ X_n]_{n\geq 0}$ è una supermartingala (e una martingala qualora f sia invariante).*

Dimostrazione. Si ha infatti, per ogni intero positivo n ed ogni elemento A di $\mathcal{F}_n$,

$$\int_A (f \circ X_{n+1} - f \circ X_n)\, dP = \int_A (Nf - f) \circ X_n \, dP \leq 0.$$

Il prossimo esempio è di fondamentale importanza per il seguito.

(4.5) Esempio. Fissato, nello spazio degli stati $(E,\mathcal{E})$, un insieme misurabile A, consideriamo, sullo spazio canonico, le due funzioni seguenti (a valori in $\{0,1\}$) che chiameremo, rispettivamente, la *funzione delle visite* ad A e la *funzione delle infinite visite* ad A:

$$v_A = \sup_{n\geq 0} I_A \circ \xi_n, \qquad w_A = \limsup_n (I_A \circ \xi_n) \tag{4.6}$$

Allora Qv e Qw sono due funzioni misurabili sullo spazio degli stati, dal significato molto intuitivo. Si ha infatti, per ogni elemento x di E,

$$Qv(x) = \epsilon_x Q\Big(\bigcup_{n\geq 0}\{\xi_n \in A\}\Big), \quad Qw(x) = \epsilon_x Q\Big(\limsup_n \{\xi_n \in A\}\Big).$$

A parole: $Qv(x)$ (risp. $Qw(x)$) è la probabilità che una catena compatibile col nucleo N e uscente da x visiti almeno una volta (risp. infinite volte) l'insieme A.

È chiaro che w è maggiorata da v. Si ha inoltre $w = w \circ \theta$ e quindi $w = w \circ \theta^n$ per ogni intero positivo n. Utilizzando il nucleo Θ introdotto in (3.4), si può scrivere la relazione $w = w \circ \theta$ nella forma $w = \Theta w$. Applicando il nucleo Q a entrambi i membri di quest'ultima relazione, si trova $Qw = Q\Theta w = NQw$. Si vede così che la funzione Qw è invariante. Si ha inoltre $v \circ \theta^n \downarrow w$, ossia

$$\Theta^n v \downarrow w. \tag{4.7}$$

Applicando il nucleo Q a entrambi i membri di questa relazione, si trova $Q\Theta^n v \downarrow Qw$, ossia, grazie a (3.7),

$$N^n Q v \downarrow Qw.$$

In particolare, la decrescenza della successione $(N^n Qv)_{n\geq 0}$ mostra che la funzione Qv è eccessiva. Si osservi che, per ogni elemento x di E, si ha

$$NQv(x) = Q\Theta v(x) = \langle \epsilon_x Q, v \circ \theta \rangle = \epsilon_x Q\left(\bigcup_{n\geq 0}\{\xi_{n+1} \in A\}\right). \tag{4.8}$$

A parole: $NQv(x)$ è la probabilità che una catena di Markov compatibile col nucleo N e uscente da x visiti l'insieme A in qualche istante diverso da 0.

(4.9) Teorema. *Nelle ipotesi e con le notazioni dell'Esempio (4.5), ciascuna delle due successioni*

$$(Qv \circ \xi_n)_{n\geq 0}, \qquad (Qw \circ \xi_n)_{n\geq 0} \tag{4.10}$$

converge quasi certamente verso w secondo ogni misura di probabilità della forma μQ, con μ legge sullo spazio degli stati.

Dimostrazione. Fissiamo una legge μ sullo spazio degli stati. Denotiamo inoltre con $(\mathcal{F}_n)_{n\geq 0}$ la filtrazione naturale del processo canonico. Per ogni intero positivo n, poiché w coincide con $w \circ \theta^n$, risulta da (3.2) che la variabile aleatoria $Qw \circ \xi_n$ è una versione della speranza condizionale $\mu Q[w \mid \mathcal{F}_n]$. In altre parole, rispetto alla filtrazione $(\mathcal{F}_n)_{n\geq 0}$, la seconda delle due successioni (4.10) è, secondo μQ, una martingala chiusa da w, e quindi convergente quasi certamente verso w.

Poiché la funzione Qv, come si è visto nell'Esempio (4.5), è eccessiva (e limitata), la prima delle due successioni (4.10) è una supermartingala limitata, dunque convergente quasi certamente. Per provare che essa converge quasi certamente verso w, basterà mostrare che la differenza tra la prima e la seconda delle due successioni (4.10) converge in media verso zero. A questo scopo, osserviamo che, grazie a (3.2), $Qv \circ \xi_n$ è una versione della speranza condizionale $\mu Q[v \circ \theta^n \mid \mathcal{F}_n]$. Passando alle speranze, e tenendo conto della relazione $v \circ \theta^n \downarrow w$, si trova dunque

$$\mu Q\big[Qv \circ \xi_n - Qw \circ \xi_n\big] = \mu Q\big[v \circ \theta^n - w\big] \downarrow 0.$$

Il teorema è così dimostrato. □

5. Caso discreto: stati transitori e stati ricorrenti

In tutto il seguito, supporremo che lo spazio $(E, \mathcal{E})$ sia discreto. Per semplicità tipografica, ogni volta che M sia un nucleo ad esso relativo, e che x, y siano due stati (cioè due elementi di E), scriveremo "$M(x,y)$" per indicare $M(x,\{y\})$. Inoltre parleremo semplicemente di "catena" per riferirci a una catena di Markov compatibile col fissato nucleo N, e indicheremo sempre con Q la realizzazione canonica di N. Diremo poi che una catena "parte da x" (o "esce da x") se la sua legge iniziale è la misura di Dirac ϵ_x. Come semplice applicazione della proprietà di Markov, si ha il seguente risultato:

(5.1) Proposizione. *Data, su uno spazio probabilizzato $(\Omega, \mathcal{A}, P)$, una catena X uscente da x, e fissati due stati y, z e due interi positivi m, n, si ha*

$$P\{X_m = y,\ X_{m+n} = z\} = N^m(x,y)\, N^n(y,z).$$

Dimostrazione. Posto $H = \{X_m = y\}$, si ha $P(H) = N^m(x,y)$. Senza ledere la generalità, si può supporre che questa probabilità non sia nulla. Si denoti con T il tempo d'arresto (relativo alla filtrazione naturale di X) che coincide con m su H e con ∞ altrove. Risulta allora da (3.1) che, secondo P_H, il processo $Y = [X_{T+h}]_{h\geq 0}$ è una catena uscente da y. Si ha dunque $P_H\{Y_n = z\} = N^n(y,z)$, ossia la formula da dimostrare. □

(5.2) Definizione. Per un fissato stato x, il numero

$$p_{x,N} = \epsilon_x Q\left(\bigcup_{n\geq 1}\{\xi_n = x\}\right)$$

(che rappresenta la probabilità che una qualsiasi catena uscente da x torni a visitare x in un istante successivo all'istante 0) sarà detto la *probabilità di ritorno* in x (relativa al nucleo N).

(5.3) Teorema. *Data, su uno spazio probabilizzato $(\Omega, \mathcal{A}, P)$, una catena X, e fissato uno stato x, si consideri la successione crescente $(T_k)_{k\geq 1}$ degli "istanti di visita a x diversi dall'istante 0", così definita per induzione:*

$$\begin{cases} T_1(\omega) = \inf\{n \in \mathbb{N}^* : X_n(\omega) = x\}, \\ T_{k+1}(\omega) = \inf\{n \in \mathbb{N}^* : n > T_k(\omega),\ X_n(\omega) = x\}. \end{cases}$$

Si ponga inoltre

$$V = \sum_{n\geq 1} I_{\{X_n = x\}} = \sum_{k\geq 1} I_{\{T_k < \infty\}}.$$

Si denoti infine con p la probabilità di ritorno in x (relativa al nucleo N).

Si ha allora, per ogni intero k strettamente positivo:

$$P\{T_{k+1} < \infty\} = p\, P\{T_k < \infty\}.$$

Sussiste pertanto la seguente dicotomia:

(a) Se è $p < 1$, allora la variabile aleatoria V è integrabile (perché la serie di termine generale $P\{T_k < \infty\}$ è convergente).

(b) Se è $p = 1$, allora l'evento $\{T_1 < \infty\}$ (ossia $\{V \neq 0\}$) è equivalente a ciascuno degli eventi $\{T_k < \infty\}$, e quindi alla loro intersezione, ossia all'evento $\{V = \infty\}$: di conseguenza, quest'ultimo evento è quasi certo se la catena esce da x.

Dimostrazione. La variabile aleatoria T_k è un tempo d'arresto per la filtrazione naturale di X. Posto $H = \{T_k < \infty\}$, si può supporre, senza ledere la generalità, che risulti $P(H) \neq 0$. Si ha allora

$$P_H\{T_{k+1} < \infty\} = P_H\Big(\bigcup_{n\geq 1}\{X_{T_k+n} = x\}\Big) = p,$$

dove l'ultima eguaglianza è dovuta al fatto che, secondo P_H, il processo $[X_{T_k+n}]_{n\geq 0}$ è una catena uscente da x. La conclusione discende dunque dal fatto che il primo membro della precedente relazione non è altro che il rapporto

$$P\{T_{k+1} < \infty\} \,/\, P\{T_k < \infty\}. \qquad \square$$

La dicotomia messa in evidenza nel precedente enunciato suggerisce la seguente definizione: lo stato x si dice *transitorio* (per il nucleo N) se la probabilità di ritorno in x (relativa al nucleo N) è minore di 1. Altrimenti, lo stato x si dice *ricorrente*.

(5.4) Corollario. *Dati una catena X e uno stato x, affinché la variabile aleatoria $Z = \sum_{n\geq 0} I_{\{X_n = x\}}$ (che rappresenta il numero totale delle visite di X a x) sia integrabile, occorre e basta che essa sia quasi certamente finita.*

Dimostrazione. Si ha $V \leq Z \leq V + 1$, dove V denota la variabile aleatoria considerata nel teorema precedente. Questa, se x è transitorio, è senz'altro integrabile; se invece x è ricorrente, allora è quasi certamente a valori in $\{0, \infty\}$ (e quindi addirittura *trascurabile*, non appena sia quasi certamente finita). Ciò basta per concludere. $\square$

(5.5) Osservazione. Affinché lo stato x sia transitorio, occorre e basta che la serie di termine generale $N^n(x, x)$ sia convergente. Per convincersene, basta osservare che, se nel Teorema (5.3) si suppone che la catena X parta da x, si ha

$$P[V] = \sum_{n\geq 1} P\{X_n = x\} = \sum_{n\geq 1} N^n(x, x).$$

(5.6) Osservazione. Se A è un insieme finito di stati transitori, allora, per ogni catena X, l'evento

$$\liminf_n \{X_n \in A^c\}$$

è quasi certo. Pertanto, se lo spazio E è finito, esiste uno stato ricorrente.

(5.7) Definizione. Dati due stati x, y, si dice che *x permette di accedere a y* (o che *y è accessibile da x*), e si scrive $x \hookrightarrow y$, se esiste un intero $n \geq 0$ tale che si abbia $N^n(x, y) \neq 0$, cioè se per una (e quindi per *ciascuna*) catena X uscente da x, l'evento $\bigcup_{n\geq 0}\{X_n = y\}$ ha probabilità non nulla. In caso contrario, si dice che y è *inaccessibile* da x.

È facile verificare che la relazione $x \hookrightarrow y$ è riflessiva e transitiva.

(5.8) Teorema. *Sia A una fissata parte dello spazio degli stati e siano v, w le funzioni considerate nell'Esempio (4.5). Se la funzione Qw è nulla (risp. eguale a 1) in un punto x di E, tale essa è in ogni punto accessibile da x.*

Dimostrazione. Sia x un punto di E con $Qw(x) = 0$ (risp. $Qw(x) = 1$). Su uno spazio probabilizzato $(\Omega, \mathcal{A}, P)$ sia X una catena uscente da x. Allora l'evento $\limsup_n\{X_n \in A\}$ è trascurabile (risp. quasi certo) secondo P. Assegnato ora un punto y di E che sia accessibile da x, cioè tale che l'istante T della prima visita di X a y non sia equivalente alla costante ∞, si ponga

$$H = \{T < \infty\}. \tag{5.9}$$

Allora, secondo la misura di probabilità P_H, il processo $[X_{T+n}]_{n\geq 0}$ è una catena uscente da y. Si ha quindi

$$Qw(y) = P_H\Big(\limsup_n\{X_{T+n} \in A\}\Big) = P_H\Big(\limsup_n\{X_n \in A\}\Big),$$

dove l'eguaglianza finale discende dal fatto che, grazie a (5.9), si ha

$$H \cap \limsup_n\{X_{T+n} \in A\} = H \cap \limsup_n\{X_n \in A\}.$$

D'altra parte, l'evento $\limsup_n\{X_n \in A\}$, essendo trascurabile (risp. quasi certo) secondo P, tale è anche secondo P_H. Si ha dunque $Qw(y) = 0$ (risp. $Qw(y) = 1$). □

(5.10) Corollario. *Sia x uno stato ricorrente, e sia y uno stato accessibile da x. Allora:*

(a) Ogni catena uscente da x visita quasi certamente infinite volte lo stato y (e quindi y non può essere transitorio).

(b) Ogni catena uscente da y visita quasi certamente infinite volte lo stato x.

Dimostrazione. (a) Denotiamo con v la funzione delle visite a $\{y\}$ e con w la funzione delle infinite visite a $\{y\}$ (si veda (4.5)). Il Teorema (4.9) mostra allora che si ha

$$\epsilon_x Q\Big\{Qv \circ \xi_n \to w\Big\} = 1.$$

Inoltre la ricorrenza di x e l'accessibilità di y da x si traducono mediante le due relazioni seguenti

$$\epsilon_x Q\Big(\limsup_n\{\xi_n = x\}\Big) = 1, \qquad Qv(x) > 0.$$

Poniamo $Qv(x) = c$. Si ha allora (tenendo conto del fatto che w è a valori in $\{0, 1\}$)

$$\{Qv \circ \xi_n \to w\} \cap \limsup_n\{\xi_n = x\} \subset \{w \geq c\} = \{w = 1\} = \limsup_n\{\xi_n = y\}.$$

Poiché il primo membro di questa relazione è quasi certo secondo $\epsilon_x Q$, tale è l'ultimo membro. Si ha dunque $\epsilon_x Q\left(\limsup_n\{\xi_n = y\}\right) = 1$, e ciò prova l'asserzione (a).

(b) Denotiamo ora invece con v la funzione delle visite a $\{x\}$ e con w la funzione delle infinite visite a $\{x\}$. Dal Teorema (5.8) risulta allora che, essendo, per ipotesi, $\epsilon_x Q\left(\limsup_n\{\xi_n = x\}\right) = 1$, ossia $Qw(x) = 1$, si ha anche $Qw(y) = 1$, ossia $\epsilon_y Q\left(\limsup_n\{\xi_n = x\}\right) = 1$. È così provata anche l'asserzione (b). □

6. Nuclei irriducibili

(6.1) Proposizione. *Sia f una funzione eccessiva, e siano x, y due stati ricorrenti, accessibili l'uno dall'altro. Si ha allora $f(x) = f(y)$.*

Dimostrazione. Pur di sostituire f con $f \wedge c$ (e far poi tendere c a $+\infty$), si può supporre che la funzione f sia limitata. Su uno spazio probabilizzato $(\Omega, \mathcal{A}, P)$ sia X una catena uscente da x, e si denoti con T l'istante della sua prima visita a y. Allora T è un tempo d'arresto (relativo alla filtrazione naturale di X) quasi certamente finito. Applicando il teorema d'arresto alla supermartingala $[f \circ X_n]_{n\geq 0}$, si trova, per ogni intero n,

$$P[f \circ X_0] \geq P[f \circ X_{T\wedge n}].$$

Di qui, facendo tendere n all'infinito (e invocando il teorema di Lebesgue sulla convergenza dominata), si trae

$$P[f \circ X_0] \geq P[f \circ X_T],$$

ossia $f(x) \geq f(y)$. Invertendo i ruoli tra x e y, si ottiene la diseguaglianza opposta, e quindi la voluta eguaglianza. □

Il nucleo N si dice *irriducibile* se non esiste alcuna coppia di stati dei quali uno sia inaccessibile dall'altro. In questo caso, gli stati sono necessariamente o tutti ricorrenti (e allora il nucleo si dice *ricorrente*) o tutti transitori (e allora il nucleo si dice *transitorio*).

(6.2) Proposizione. *Le condizioni che seguono sono equivalenti:*

(a) Il nucleo N è irriducibile ricorrente.

(b) Per ogni coppia x, y di stati, risulta $\sum_{n\geq 0} N^n(x,y) = \infty$.

Dimostrazione. La somma che figura nella condizione (b) rappresenta, per una catena uscente da x, il valor medio del numero delle visite a y. Perciò è chiaro che (a) implica (b). Inversamente, se è soddisfatta la condizione (b), allora, innanzitutto, per ogni coppia x, y di stati, esiste qualche intero positivo n tale che si abbia $N^n(x,y) \neq 0$; in secondo luogo, nessuno stato è transitorio perché, per ogni x, risulta $\sum_{n\geq 0} N^n(x,x) = \infty$. □

(6.3) Teorema. *Le condizioni che seguono sono equivalenti:*

(a) Il nucleo N è irriducibile ricorrente.

(b) Ogni funzione eccessiva limitata è costante.

(c) Per ogni coppia x, y di stati, la probabilità che una catena uscente da x visiti y è eguale a 1.

Dimostrazione. L'implicazione (a) $\Rightarrow$ (b) è già nota (si veda (6.1)). Per provare l'implicazione (b) $\Rightarrow$ (c), si supponga soddisfatta la condizione (b), si fissi uno stato y e si denoti con v la funzione delle visite al singoletto $\{y\}$ (si veda (4.5)). Allora Qv è la funzione che, ad ogni stato x, associa la probabilità che una catena uscente da x visiti y. Questa funzione, come sappiamo da (4.5), è eccessiva. Dunque essa è costante, grazie all'ipotesi (b). Poiché vale 1 nel punto y, coincide con la costante 1.

Proviamo infine l'implicazione (c) $\Rightarrow$ (a). Supponiamo dunque verificata la condizione (c). Allora è chiaro che ogni stato permette di accedere ad ogni altro stato. Rimane da provare che ogni stato è ricorrente. A questo scopo, fissato lo stato x, denotiamo con v la funzione delle visite al singoletto $\{x\}$ (si veda (4.5)). L'ipotesi (c) implica che Qv è la costante 1. Ne segue che anche NQv è la costante 1. D'altra parte, $NQv(x)$ è (si veda (4.8)) la probabilità che una catena uscente da x visiti x in qualche istante diverso da 0, ossia (si veda (5.2)) la probabilità di ritorno in x relativa al nucleo N. Dunque questa probabilità è eguale a 1, e ciò vuol dire appunto che lo stato x è ricorrente. □

7. Esistenza di misure invarianti

Una misura m su E si dirà *eccessiva* (risp. *invariante*) se verifica la relazione $mN \leq m$ (risp. $mN = m$). Ragionando per induzione, si vede che, se la misura m è eccessiva (risp. invariante), allora, per ogni intero positivo k, si ha $mN^k \leq m$ (risp. $mN^k = m$).

Ci proponiamo di dimostrare che, se esiste uno stato ricorrente, allora esiste una misura invariante non banale. A questo scopo, fissato uno stato ricorrente a, consideriamo, su uno spazio probabilizzato $(\Omega, \mathcal{A}, P)$, una catena X uscente da a, e denotiamo con $(T_k)_{k\geq 1}$ la successione dei suoi istanti di visita ad a diversi dall'istante 0 (vedi (5.3)). Allora, per ogni elemento x di E, la variabile aleatoria

$$Z_x := \sum_{n\geq 0} I_{\{X_n = x,\, T_1 > n\}} \tag{7.1}$$

rappresenta il numero di visite allo stato x che la catena X compie in istanti strettamente anteriori a T_1 (che è l'istante del primo ritorno di X in a). Denotiamo ora con m la misura su E così caratterizzata: per ogni elemento x di E, il valore di m sul singoletto $\{x\}$ è il valor medio della variabile aleatoria (7.1):

$$m\{x\} = P\left[\sum_{n\geq 0} I_{\{X_n = x,\, T_1 > n\}}\right] = \sum_{n\geq 0} P\{X_n = x,\, T_1 > n\}. \tag{7.2}$$

Per $x = a$, si ha evidentemente $m\{a\} = 1$. La misura m si può scrivere nella forma

$$m = \sum_{n\geq 0} \lambda_n, \tag{7.3}$$

dove λ_n è la misura definita da

$$\lambda_n = X_n \left(I_{\{T_1 > n\}} \cdot P\right), \tag{7.4}$$

cioè caratterizzata dalla relazione

$$\langle \lambda_n, f\rangle = \int_{\{T_1>n\}} f\circ X_n \,\mathrm{d}P \quad \text{per ogni funzione positiva } f \text{ su } E, \tag{7.5}$$

o anche dalla relazione

$$\lambda_n\{x\} = P\{X_n = x,\, T_1 > n\} \quad \text{per ogni elemento } x \text{ di } E. \tag{7.6}$$

Il lemma seguente, che riguarda le trasformate delle misure λ_n mediante il nucleo N, sarà utile per dimostrare che la misura m è invariante.

(7.7) Lemma. *Per ogni intero positivo n, si ha*

$$\lambda_n N\{x\} = \lambda_{n+1}\{x\} \quad \text{per } x \neq a, \tag{7.8}$$

$$\lambda_n N\{a\} = P\{T_1 = n+1\}. \tag{7.9}$$

Dimostrazione. Per ogni funzione positiva f su E, tenendo conto di (7.5), si può scrivere

$$\langle \lambda_n N,\, f\rangle = \langle \lambda_n,\, Nf\rangle = \int_{\{T_1>n\}} Nf\circ X_n \,\mathrm{d}P = \int_{\{T_1>n\}} f\circ X_{n+1}\,\mathrm{d}P$$

(dove l'eguaglianza finale è dovuta alla proprietà di Markov). In particolare, prendendo f eguale alla funzione indicatrice del singoletto $\{x\}$, con $x \neq a$, si trova

$$\lambda_n N\{x\} = P\{X_{n+1} = x,\, T_1 > n\} = P\{X_{n+1} = x,\, T_1 > n+1\} = \lambda_{n+1}\{x\}.$$

Prendendo invece f eguale alla funzione indicatrice del singoletto $\{a\}$, si trova

$$\lambda_n N\{a\} = P\{X_{n+1} = a,\, T_1 > n\} = P\{T_1 = n+1\}.$$

Il lemma è così dimostrato. □

Siamo ora in grado di dimostrare il teorema seguente, che riguarda le proprietà fondamentali della misura m.

(7.10) Teorema. *La misura m sopra definita possiede le proprietà seguenti:*

(a) m è invariante.

(b) $m(E) = P[T_1]$.

(c) $m\{x\} \neq 0$ se e solo se $a \hookrightarrow x$.

(d) $m\{x\} < \infty$ per ogni x.

– *Dimostrazione.* Per provare che m è invariante, basta provare che, per ogni elemento x di E, si ha $mN\{x\} = m\{x\}$. Supponiamo dapprima $x \neq a$. Allora, sommando le relazioni (7.8), si trova

$$mN\{x\} = \sum_{n\geq 0} \lambda_n N\{x\} = \sum_{n\geq 0} \lambda_{n+1}\{x\} = m\{x\}$$

(dove l'eguaglianza finale è dovuta al fatto che λ_0 è nulla su $\{x\}$). Supponiamo ora $x = a$. Allora, sommando le relazioni (7.9), si trova

$$mN\{a\} = \sum_{n\geq 0} \lambda_n N\{a\} = \sum_{n\geq 0} P\{T_1 = n+1\} = P\{T_1 > 0\} = 1 = m\{a\}.$$

È così provato che la misura m è invariante.

– Per provare la proprietà (b), basta osservare che la massa totale di m è la somma delle masse totali delle λ_n, sicché risulta

$$m(E) = \sum_{n\geq 0} P\{T_1 > n\} = P\big[T_1\big].$$

– Per provare la proprietà (c), fissiamo lo stato x e poniamo, per ogni intero k strettamente positivo,

$$V_k = \sum_{n\geq 0} I_{\{X_n = x,\ T_{k-1} \leq n < T_k\}}$$

(con $T_0 = 0$). Le variabili aleatorie V_k così definite sono isonome (come facilmente si riconosce applicando (3.1)), e la loro comune speranza è eguale a $m\{x\}$. Dunque, affinché risulti $m\{x\} = 0$, occorre e basta che sia trascurabile la variabile aleatoria

$$\sum_{k\geq 1} V_k = \sum_{n\geq 0} I_{\{X_n = x\}},$$

cioè che lo stato x sia inaccessibile da a.

– Infine, per provare la proprietà (d), fissiamo lo stato x, che potremo supporre *diverso* da a, e osserviamo che $m\{x\}$ coincide allora con la speranza della variabile aleatoria

$$Z = \sum_{n\geq 0} I_{\{X_{n+1} = x,\ T_1 > n+1\}}.$$

Si tratta dunque di dimostrare che questa variabile aleatoria è integrabile. Consideriamo la catena $Y = [X_{n+1}]_{n\geq 0}$. La variabile aleatoria $S = T_1 - 1$ è il primo istante in cui Y visita a, mentre Z rappresenta il numero totale delle visite allo stato x da parte del processo $Y^{|S}$ ottenuto arrestando Y all'istante S. Questo processo è una catena di Markov compatibile col nucleo M così definito:

$$M(z, \cdot\,) = \begin{cases} N(z, \cdot\,) & \text{se } z \neq a, \\ \epsilon_a & \text{altrimenti.} \end{cases}$$

Dunque la variabile aleatoria Z, essendo quasi certamente finita (perché maggiorata da T_1), è necessariamente integrabile (si veda (5.4)). □

(7.11) Definizione. La misura m considerata nel teorema precedente sarà detta la misura invariante *associata* allo stato ricorrente a. Quando sarà necessario metterne in risalto la dipendenza da a, essa sarà denotata con m_a. Ricordiamo che la sua massa totale è eguale al valor medio della variabile aleatoria T_1 (istante del primo ritorno in a per una catena X uscente da a). Questo valor medio non dipende dalla particolare catena X, ma solo dallo stato a e dal nucleo N. Lo chiameremo il *tempo medio di ritorno* in a (relativo al nucleo N).

(7.12) Proposizione. *Sia a uno stato ricorrente. Allora la misura invariante m ad esso associata è la minima misura eccessiva che maggiori la misura di Dirac ϵ_a.*

Dimostrazione. Sia ν una misura eccessiva, con $\nu \geq \epsilon_a$. Si ha allora

$$\nu\{a\} \geq \epsilon_a\{a\} = 1 = m\{a\}.$$

Per provare che ν maggiora m, rimane dunque solo da provare che, per ogni elemento x di E, con $x \neq a$, ha luogo la relazione

$$\nu\{x\} \geq \lambda_0\{x\} + \cdots + \lambda_n\{x\} \quad \text{per } n \geq 0.$$

Ciò si ottiene immediatamente per induzione, sfruttando il carattere eccessivo di ν e il Lemma (7.7). □

Dato lo stato ricorrente a, poniamo

$$H_a \doteq \{x \in E : a \hookrightarrow x\}, \qquad K_a \doteq \{x \in E : x \hookrightarrow a\}. \tag{7.13}$$

Si ha allora, grazie a (5.10), $H \subset K$. Inoltre, come risulta da (7.10) (c), la misura invariante m_a è concentrata su H (e quindi, a maggior ragione, su K). Allo scopo di fornire un'ulteriore caratterizzazione della misura m_a, conviene premettere il lemma seguente.

(7.14) Lemma. *Sia a un fissato stato. Allora ogni misura eccessiva che si annulli su $\{a\}$ si annulla sull'insieme K definito in (7.13).*

Dimostrazione. Sia δ una misura eccessiva che si annulli su $\{a\}$, e sia x un elemento di K, ossia un elemento di E tale che esista un intero positivo k con $N^k(x,a) > 0$. Si ha allora

$$\delta\{x\}\, N^k(x,a) \leq \sum_{y \in E} \delta\{y\}\, N^k(y,a) = (\delta N^k)\{a\} \leq \delta\{a\} = 0.$$

Ne segue $\delta\{x\} = 0$. □

Ed ecco l'annuciata caratterizzazione di m_a.

(7.15) Proposizione. *La misura invariante m_a associata allo stato ricorrente a è l'unica misura eccessiva che assuma il valore 1 su $\{a\}$ e che sia concentrata sull'insieme K definito in (7.13).*

Dimostrazione. Sia ν una misura eccessiva che assuma il valore 1 su $\{a\}$ e che sia concentrata su K. Grazie a (7.12), si ha $\nu \geq m_a$. Perciò ν si può mettere nella forma $\nu = m_a + \delta$, con δ misura positiva; questa misura si annulla su $\{a\}$ e, come facilmente si riconosce, è eccessiva e concentrata su K. Grazie al lemma precedente, essa è dunque nulla su K (e quindi identicamente nulla). □

(7.16) Definizione. Secondo la felice terminologia proposta da N. Pintacuda, uno stato ricorrente a si dice *velocemente ricorrente* (o, più brevemente, *veloce*) se la massa totale della misura invariante m_a ad esso associata è finita. (Ricordiamo che questa massa totale coincide col tempo medio di ritorno in a.) In caso contrario, lo stato ricorrente a si dice *lentamente ricorrente* (o *lento*).

Dalla Proposizione (7.15) discende il seguente corollario.

(7.17) Corollario. *Siano a, b due stati ricorrenti, accessibili l'uno dall'altro. Allora la misura invariante m_b associata a b è proporzionale a m_a. (Pertanto, affinché lo stato a sia veloce, occorre e basta che tale sia b.)*

Dimostrazione. Poiché i due stati a, b sono accessibili l'uno dall'altro, l'insieme H definito in (7.13) è identico all'insieme $\{x \in E : b \hookrightarrow x\}$. Perciò si ha $m_b\{x\} > 0$ per ogni x di H. La misura $(m_b\{a\})^{-1}\, m_b$ è invariante, è concentrata su H e assume il valore 1 su $\{a\}$. Grazie alla Proposizione (7.15), essa coincide dunque con m_a. □

8. Esistenza di una legge invariante

Ci proponiamo ora di indagare in quali condizioni il nucleo N ammetta una *legge* invariante (ossia una misura invariante che sia normalizzata). Cominciamo col provare il seguente semplice risultato.

(8.1) Proposizione. *Ogni legge invariante è concentrata sull'insieme degli stati ricorrenti.*

Dimostrazione. Sia μ una legge invariante. Allora, per ogni n, la legge di ξ_n secondo la misura μQ è eguale a μN^n, dunque a μ. In altri termini, per ogni elemento x di E, si ha $\mu Q\{\xi_n = x\} = \mu\{x\}$. Se x è transitorio, ne segue, grazie al lemma di Fatou,

$$\mu\{x\} = \limsup_n \mu Q\{\xi_n = x\} \leq \mu Q\big(\limsup_n \{\xi_n = x\}\big) = 0. \qquad \square$$

(8.2) Corollario. *Se, per il nucleo N, gli stati sono tutti transitori, allora non esiste alcuna legge invariante.*

La proposizione che segue riguarda il caso di un nucleo irriducibile ricorrente.

(8.3) Proposizione. *Si supponga che il nucleo N sia irriducibile ricorrente. Allora ogni misura eccessiva che, su almeno un singoletto, assuma un valore finito e non nullo (in particolare, ogni legge invariante) è proporzionale a una misura invariante della forma m_a.*

Dimostrazione. Sia μ una misura eccessiva e sia a un elemento di E tale che si abbia $0 < \mu\{a\} < \infty$. Essendo N irriducibile, gli insiemi $H(a,K)$ definiti da (7.13) coincidono con l'intero spazio E. Da (7.15) discende dunque che la misura $(\mu\{a\})^{-1}\mu$ coincide con m_a. □

Sfruttando il risultato precedente e il Corollario (7.17), si ottiene infine il risultato seguente.

(8.4) Teorema. *Si supponga che il nucleo N sia irriducibile ricorrente. Sono allora possibili i due soli casi seguenti:*

(a) Gli stati sono tutti lentamente ricorrenti.

(b) Gli stati sono tutti velocemente ricorrenti.

Nel caso (a), non esiste alcuna legge invariante. Nel caso (b), esiste un'unica legge invariante. Il valore di questa legge sul generico singoletto $\{x\}$ è eguale a $1/m_x(E)$, ossia al reciproco del tempo medio di ritorno in x relativo al nucleo N.

Dimostrazione. Che siano possibili soltanto i casi (a) e (b) discende da (7.17).

Nel caso (a), non può esistere una legge invariante perché una tal legge, grazie a (8.3), dovrebbe essere proporzionale a una misura della forma m_a, dunque a una misura di massa totale infinita.

Nel caso (b), le misure della forma m_a sono tutte di massa totale finita. Ogni misura ottenuta normalizzando una di esse è una legge invariante. Ma, grazie a (7.17), le leggi ottenute in questo modo sono tutte eguali a una medesima legge π.

D'altra parte, grazie a (8.3), ogni legge invariante è proporzionale a una delle misure m_a, e dunque è identica a π.

Infine, per ogni elemento x di E, poiché la legge π si può vedere come ottenuta normalizzando m_x, si ha

$$\pi\{x\} = m_x\{x\}/m_x(E) = 1/m_x(E).$$

Il teorema è così dimostrato. □

(8.5) Definizione. Se, per un nucleo irriducibile ricorrente, è verificato il caso (a) (risp. (b)) descritto nel teorema precedente, allora, in accordo con la terminologia introdotta in (7.16), il nucleo è detto *lentamente ricorrente* o *lento* (risp. *velocemente ricorrente* o *veloce*).

9. Un risultato di ergodicità*

L'importanza del fatto che un nucleo irriducibile e velocemente ricorrente ammetta una legge invariante è messa ben in luce dal teorema seguente.

(9.1) Teorema. *Supposto che il nucleo N sia irriducibile e velocemente ricorrente, si denoti con π l'unica legge per esso invariante. Su uno spazio probabilizzato $(\Omega, \mathcal{A}, P)$ sia X una catena avente π come legge iniziale. Valgono allora le affermazioni seguenti:*

(a) La successione $(X_n)_{n\geq 0}$ è stazionaria.

(b) Ogni funzione f appartenente a $\mathcal{L}^1(\pi)$, e tale che la successione $(f \circ X_n)_{n\geq 0}$ converga in $\mathcal{L}^1(P)$, è costante.

(c) La tribù degli eventi invarianti relativa alla successione $(X_n)_{n\geq 0}$ è degenere (ossia la successione $(X_n)_{n\geq 0}$ è "ergodica").

(d) Per ogni funzione g appartenente a $\mathcal{L}^1(\pi)$, la successione

$$\left(n^{-1}\sum_{j=0}^{n-1} g \circ X_j\right)_{n\geq 1} \tag{9.2}$$

converge quasi certamente e in $\mathcal{L}^1(P)$ verso la costante (π, g).

Dimostrazione. (a) Il processo traslato $[X_{n+1}]_{n\geq 0}$ è una catena (compatibile con N) avente come legge iniziale πN, ossia π. Esso ha dunque la stessa legge πQ del processo X. In altri termini, la successione $(X_n)_{n\geq 0}$ è stazionaria.

(b) Ricordiamo che si ha $\pi\{x\} \neq 0$ per ogni elemento x di E. Sia f un elemento di $\mathcal{L}^1(\pi)$ tale che la successione $(f \circ X_n)_{n\geq 0}$ converga in $\mathcal{L}^1(P)$, e proviamo che f è costante. A questo scopo, fissata una coppia (x, y) di elementi di E, scegliamo un intero positivo h tale che si abbia $N^h(x, y) \neq 0$. Allora, poiché π è la legge di X_0, e N^h è una versione della legge condizionale di X_h rispetto a X_0, la legge del blocco $[X_0, X_h]$ è $\pi \otimes N^h$, sicché risulta

$$P\{X_0 = x,\, X_h = y\} = \pi \otimes N^h\big(\{x\} \times \{y\}\big) = \pi\{x\}N^h(x, y) \neq 0. \tag{9.3}$$

Inoltre, poiché la successione $(X_n)_{n\geq 0}$ è stazionaria, si ha, per ogni intero positivo n,

$$\begin{aligned}\int |f \circ X_n - f \circ X_{n+h}|\, \mathrm{d}P &= \int |f \circ X_0 - f \circ X_h|\, \mathrm{d}P \\ &\geq P\{X_0 = x,\, X_h = y\}\, |f(x) - f(y)|.\end{aligned}$$

Da questa relazione, nella quale il primo membro converge verso zero al tendere di n all'infinito, si deduce (tenendo conto di (9.3)) $f(x) = f(y)$. Per l'arbitrarietà della coppia (x, y), è così provata la costanza di f.

(c) Si tratta di provare che, se u è una funzione limitata e misurabile sullo spazio canonico, la quale sia invariante per l'operatore θ di slittamento, allora la variabile aleatoria reale $u \circ X$ è equivalente a una costante. A questo scopo, indicando con $\mathcal{F}$

la filtrazione naturale del processo X, osserviamo che, per ogni intero positivo n, una versione della speranza condizionale

$$P\big[u \circ X \mid \mathcal{F}_n\big] = P\big[u \circ [X_{n+h}]_{h\geq 0} \mid \mathcal{F}_n\big]$$

è la variabile aleatoria $Qu \circ X_n$ (si veda (3.2)). Ciò significa che $(Qu \circ X_n)_{n\geq 0}$ è, rispetto alla filtrazione $\mathcal{F}$, una martingala chiusa da $u \circ X$, dunque convergente quasi certamente e in $\mathcal{L}^1(P)$ verso $u \circ X$. Sfruttando l'affermazione (b) (nella quale si prenda $f = Qu$), si vede che la funzione Qu è eguale a una costante reale c. Di conseguenza, la variabile aleatoria $u \circ X$, come limite in $\mathcal{L}^1(P)$ di una successione i cui termini sono tutti eguali alla costante c, è equivalente, secondo P, a questa stessa costante.

(d) Sia g un elemento di $\mathcal{L}^1(\pi)$. Poiché, la successione $(X_n)_{n\geq 0}$ è stazionaria, tale è anche la successione di variabili aleatorie reali $(g \circ X_n)_{n\geq 0}$. Applicando ad essa la legge forte dei grandi numeri per successioni stazionarie, si trova che la successione (9.2) converge quasi certamente e in $\mathcal{L}^1(P)$ verso una versione della speranza condizionale $P\big[g \circ X_0 \mid \mathcal{J}\big]$, dove $\mathcal{J}$ denota la tribù degli eventi invarianti relativa alla successione $(X_n)_{n\geq 0}$. Poiché si è già provato che questa tribù è degenere, ciò è come dire che la successione (9.2) converge quasi certamente e in $\mathcal{L}^1(P)$ verso la costante

$$P\big[g \circ X_0\big] = \langle \pi, g \rangle. \qquad \square$$

Le passeggiate aleatorie sulla retta come speciali catene di Markov

1. Nuclei di convoluzione sulla retta

(1.1) Proposizione. *Data una misura di probabilità ν sulla tribù boreliana di $\mathbb{R}$, esiste un nucleo markoviano N, relativo allo spazio misurabile $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, caratterizzato dalla proprietà seguente: per ogni numero reale x, la misura $N(x, \cdot\,)$ è l'immagine di ν mediante la traslazione $y \mapsto x + y$.*

Dimostrazione. Si tratta di provare che, per ogni funzione g boreliana e limitata su $\mathbb{R}$, la funzione

$$x \mapsto \langle N(x, \cdot\,), g\rangle \tag{1.2}$$

è boreliana su $\mathbb{R}$. Ma ciò è immediato: infatti, essendo

$$\langle N(x, \cdot\,), g\rangle = \int \nu(\mathrm{d}y)\, g(x + y),$$

la funzione (1.2) è boreliana in virtù del teorema di Fubini. □

(1.3) Proposizione. *Nelle ipotesi della proposizione precedente, il nucleo N opera così sulle misure: esso trasforma la generica misura di probabilità μ su $\mathcal{B}(\mathbb{R})$ nel prodotto di convoluzione di μ con ν:*

$$\mu N = \mu \star \nu.$$

Dimostrazione. Infatti, per ogni misura di probabilità μ su $\mathcal{B}(\mathbb{R})$ ed ogni funzione g limitata e boreliana su $\mathbb{R}$, si ha

$$\langle \mu N, g\rangle = \langle \mu,\, Ng\rangle = \int \mu(\mathrm{d}x) \int \nu(\mathrm{d}y)\, g(x + y) = \langle \mu \star \nu,\, g\rangle. \qquad \square$$

Il risultato appena dimostrato giustifica la definizione seguente.

(1.4) Definizione. Per ogni misura di probabilità ν su $\mathcal{B}(\mathbb{R})$, il nucleo markoviano N caratterizzato nella Proposizione (1.1) si chiama il *nucleo di convoluzione* associato alla misura ν.

2. Passeggiate aleatorie come catene di Markov

(2.1) Definizione. Su uno spazio probabilizzato $(\Omega, \mathcal{A}, P)$ sia $(Y_n)_{n\geq 1}$ una successione di variabili aleatorie reali indipendenti, dotate di una medesima legge ν. Allora il processo reale $X = [X_n]_{n\geq 0}$ definito dalle relazioni

$$X_0 = 0, \qquad X_n = Y_1 + \cdots + Y_n \quad \text{per } n \geq 1$$

si chiama una *passeggiata aleatoria* su $\mathbb{R}$. Più precisamente, esso si chiama la passeggiata aleatoria su $\mathbb{R}$ *uscente dall'origine* e avente $(Y_n)_{n\geq 1}$ come *successione dei passi*.

(2.2) Teorema. *Nelle ipotesi della definizione precedente, la passeggiata aleatoria X è una catena di Markov compatibile col nucleo di convoluzione associato a ν.*

Dimostrazione. Denotiamo con N il nucleo di convoluzione associato a ν. Inoltre, fissato un intero positivo n, consideriamo il blocco Z così definito:

$$Z = \big[[X_0, \dots, X_n],\, Y_{n+1}\big]. \tag{2.3}$$

La relazione

$$X_{n+1} = X_n + Y_{n+1}$$

permette di scrivere X_{n+1} nella forma

$$X_{n+1} = h \circ Z = h \circ \big[[X_0, \dots, X_n],\, Y_{n+1}\big], \tag{2.4}$$

pur di denotare con h la funzione reale così definita sullo spazio d'arrivo di Z:

$$h\big((x_0, \dots, x_n),\, y\big) = x_n + y. \tag{2.5}$$

D'altra parte, la variabile aleatoria Y_{n+1} è indipendente dal blocco $[X_0, \dots, X_n]$ ed ha come legge ν: essa ammette dunque, come versione della propria legge condizionale rispetto a $[X_0, \dots, X_n]$, il nucleo costante K definito da

$$K\big((x_0, \dots, x_n),\, \cdot\,\big) = \nu.$$

Grazie a (2.4), ne segue che una versione della legge condizionale di X_{n+1} rispetto al blocco $[X_0, \dots, X_n]$ è il nucleo L_n caratterizzato da questa proprietà: per ogni elemento $(x_0, \dots, x_n)$ dello spazio d'arrivo di $[X_0, \dots, X_n]$, la misura $L_n\big((x_0, \dots, x_n),\, \cdot\,\big)$ è l'immagine di ν mediante l'applicazione

$$y \mapsto h\big((x_0, \dots, x_n),\, y\big).$$

Ma questa applicazione, grazie a (2.5), è semplicemente la traslazione $y \mapsto x_n + y$. Si vede dunque che L_n è il nucleo definito ponendo

$$L_n\big((x_0, \dots, x_n),\, \cdot\,\big) = \epsilon_{x_n} \star \nu = N(x_n,\, \cdot\,).$$

Poiché questa conclusione vale per ogni n, è così provato che X è una catena di Markov compatibile col nucleo N. □

Le passeggiate aleatorie sulla retta come speciali catene di Markov

1. Nuclei di convoluzione sulla retta

(1.1) Proposizione. *Data una misura di probabilità ν sulla tribù boreliana di $\mathbb{R}$, esiste un nucleo markoviano N, relativo allo spazio misurabile $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, caratterizzato dalla proprietà seguente: per ogni numero reale x, la misura $N(x, \cdot)$ è l'immagine di ν mediante la traslazione $y \mapsto x + y$.*

Dimostrazione. Si tratta di provare che, per ogni funzione g boreliana e limitata su $\mathbb{R}$, la funzione

$$x \mapsto \langle N(x, \cdot), g \rangle \tag{1.2}$$

è boreliana su $\mathbb{R}$. Ma ciò è immediato: infatti, essendo

$$\langle N(x, \cdot), g \rangle = \int \nu(\mathrm{d}y)\, g(x + y),$$

la funzione (1.2) è boreliana in virtù del teorema di Fubini. □

(1.3) Proposizione. *Nelle ipotesi della proposizione precedente, il nucleo N opera così sulle misure: esso trasforma la generica misura di probabilità μ su $\mathcal{B}(\mathbb{R})$ nel prodotto di convoluzione di μ con ν:*

$$\mu N = \mu \star \nu.$$

Dimostrazione. Infatti, per ogni misura di probabilità μ su $\mathcal{B}(\mathbb{R})$ ed ogni funzione g limitata e boreliana su $\mathbb{R}$, si ha

$$\langle \mu N, g \rangle = \langle \mu, Ng \rangle = \int \mu(\mathrm{d}x) \int \nu(\mathrm{d}y)\, g(x + y) = \langle \mu \star \nu, g \rangle. \qquad \square$$

Il risultato appena dimostrato giustifica la definizione seguente.

(1.4) Definizione. Per ogni misura di probabilità ν su $\mathcal{B}(\mathbb{R})$, il nucleo markoviano N caratterizzato nella Proposizione (1.1) si chiama il *nucleo di convoluzione* associato alla misura ν.

2. Passeggiate aleatorie come catene di Markov

(2.1) Definizione. Su uno spazio probabilizzato $(\Omega, \mathcal{A}, P)$ sia $(Y_n)_{n \geq 1}$ una successione di variabili aleatorie reali indipendenti e isonome, e sia X_0 una variabile reale indipendente dal blocco delle Y_n. Si ponga

$$X_n = X_0 + Y_1 + \cdots + Y_n \quad \text{per } n \geq 1.$$

Allora il processo reale $[X_n]_{n \geq 0}$ si chiama una *passeggiata aleatoria* su $\mathbb{R}$. Inoltre, per questo processo, la legge di X_0 si chiama la *legge iniziale*, mentre la legge comune delle Y_n si chiama la *legge del singolo passo*.

(2.2) Osservazione. Comunque si assegni una coppia (μ, ν) di misure di probabilità su $\mathcal{B}(\mathbb{R})$, è sempre possibile (usando lo schema delle prove indipendenti) costruire una passeggiata aleatoria su $\mathbb{R}$ ammettente μ come legge iniziale e ν come legge del singolo passo. Inoltre è chiaro che due siffatte passeggiate aleatorie, se considerate come variabili aleatorie a valori nello spazio canonico $\mathbb{R}^{\mathbb{N}}$, hanno leggi identiche: queste leggi si ottengono infatti prendendo l'immagine della misura prodotto $\mu \otimes \nu \otimes \nu \otimes \cdots$ mediante una medesima applicazione misurabile dello spazio canonico nello spazio canonico.

Il teorema che segue ribadisce l'osservazione precedente, rendendola più esplicita. Esso permette inoltre di ricondurre il concetto di passeggiata aleatoria su $\mathbb{R}$ al concetto, più generale, di catena di Markov.

(2.3) Teorema. *Siano μ, ν due misure di probabilità sulla tribù boreliana di $\mathbb{R}$. Si denoti con N il nucleo di convoluzione associato a ν e con Q la realizzazione canonica di N. Allora ogni passeggiata aleatoria su $\mathbb{R}$ ammettente μ come legge iniziale e ν come legge del singolo passo è una catena di Markov compatibile col nucleo N (ossia ammette come legge μQ).*

*Dimostrazione.** Su uno spazio probabilizzato $(\Omega, \mathcal{A}, P)$ sia $[X_n]_{n\geq 0}$ una passeggiata aleatoria ammettente μ come legge iniziale e ν come legge del singolo passo. Indicando con $[\xi_n]_{n\geq 0}$ il processo canonico sullo spazio canonico $\mathbb{R}^{\mathbb{N}}$, basterà provare che, per ogni intero positivo n, ha luogo la seguente eguaglianza tra leggi:

$$[X_0, \ldots, X_n](P) = [\xi_0, \ldots, \xi_n](\mu Q). \tag{2.4}$$

Questa eguaglianza è banalmente vera per $n = 0$ (riducendosi alla forma $X_0(P) = \mu$). Supponiamola vera per un certo intero positivo n. Per provare che essa vale anche per $n+1$, basterà provare che esiste un nucleo markoviano che sia, nello stesso tempo, una versione della legge condizionale di X_{n+1} rispetto a $[X_0, \ldots, X_n]$ secondo P e una versione della legge condizionale di ξ_{n+1} rispetto a $[\xi_0, \ldots, \xi_n]$ secondo μQ. A questo scopo, poniamo $Y_{n+1} = X_{n+1} - X_n$ e consideriamo il blocco Z_n così definito:

$$Z_n = [[X_0, \ldots, X_n],\, Y_{n+1}]. \tag{2.5}$$

La relazione

$$X_{n+1} = X_n + Y_{n+1}$$

permette di scrivere X_{n+1} nella forma

$$X_{n+1} = h_n \circ Z_n = h_n \circ [[X_0, \ldots, X_n],\, Y_{n+1}], \tag{2.6}$$

dove h_n è la funzione reale così definita sullo spazio d'arrivo di Z_n:

$$h_n\big((x_0, \ldots, x_n),\, y\big) = x_n + y. \tag{2.7}$$

D'altra parte, la variabile aleatoria Y_{n+1} è indipendente dal blocco $[X_0, \ldots, X_n]$ ed ha come legge ν: essa ammette dunque, come versione della propria legge condizionale rispetto a $[X_0, \ldots, X_n]$, il nucleo costante K_n definito da

$$K_n\big((x_0, \ldots, x_n),\ \cdot\ \big) = \nu.$$

Grazie a (2.6), ne segue che una versione della legge condizionale di X_{n+1} rispetto al blocco $[X_0, \ldots, X_n]$ è il nucleo L_n caratterizzato da questa proprietà: per ogni elemento $(x_0, \ldots, x_n)$ dello spazio d'arrivo di $[X_0, \ldots, X_n]$, la misura $L_n\big((x_0, \ldots, x_n), \cdot\big)$ è l'immagine di ν mediante l'applicazione

$$y \mapsto h_n\big((x_0, \ldots, x_n),\, y\big).$$

Ma questa applicazione, grazie a (2.7), è semplicemente la traslazione $y \mapsto x_n + y$. Si vede dunque che L_n è il nucleo definito ponendo

$$L_n\big((x_0, \ldots, x_n),\ \cdot\ \big) = \epsilon_{x_n} \star \nu = N(x_n,\ \cdot\ ).$$

Per la definizione stessa di catena di Markov, il nucleo L_n è, secondo μQ, una versione della legge condizionale di ξ_{n+1} rispetto al blocco $\big[\xi_0, \ldots, \xi_n\big]$. Il teorema è così dimostrato. □

Le passeggiate aleatorie sulla retta come speciali catene di Markov

1. Nuclei di convoluzione sulla retta

(1.1) Proposizione. *Data una misura di probabilità ν sulla tribù borciana di $\mathbb{R}$, esiste un nucleo markoviano N, relativo allo spazio misurabile $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, caratterizzato dalla proprietà seguente: per ogni numero reale x, la misura $N(x, \cdot)$ è l'immagine di ν mediante la traslazione $y \mapsto x + y$.*

Dimostrazione. Si tratta di provare che, per ogni funzione g boreliana e limitata su $\mathbb{R}$, la funzione

$$x \mapsto \langle N(x, \cdot), g \rangle \tag{1.2}$$

è boreliana su $\mathbb{R}$. Ma ciò è immediato: infatti, essendo

$$\langle N(x, \cdot), g \rangle = \int \nu(\mathrm{d}y)\, g(x + y),$$

la funzione (1.2) è boreliana in virtù del teorema di Fubini. □

(1.3) Proposizione. *Nelle ipotesi della proposizione precedente, il nucleo N opera così sulle misure: esso trasforma la generica misura di probabilità μ su $\mathcal{B}(\mathbb{R})$ nel prodotto di convoluzione di μ con ν:*

$$\mu N = \mu \star \nu.$$

Dimostrazione. Infatti, per ogni misura di probabilità μ su $\mathcal{B}(\mathbb{R})$ ed ogni funzione g limitata e boreliana su $\mathbb{R}$, si ha

$$\langle \mu N, g \rangle = \langle \mu, Ng \rangle = \int \mu(\mathrm{d}x) \int \nu(\mathrm{d}y)\, g(x + y) = \langle \mu \star \nu, g \rangle. \qquad \square$$

Il risultato appena dimostrato giustifica la definizione seguente.

(1.4) Definizione. Per ogni misura di probabilità ν su $\mathcal{B}(\mathbb{R})$, il nucleo markoviano N caratterizzato nella Proposizione (1.1) si chiama il *nucleo di convoluzione* associato alla misura ν.

2. Passeggiate aleatorie come catene di Markov

(2.1) Definizione. Su uno spazio probabilizzato $(\Omega, \mathcal{A}, P)$ sia $(Y_n)_{n \geq 1}$ una successione di variabili aleatorie reali indipendenti e isonome, e sia X_0 una variabile reale indipendente dal blocco delle Y_n. Si ponga

$$X_n = X_0 + Y_1 + \cdots + Y_n \quad \text{per } n \geq 1.$$

Allora il processo reale $[X_n]_{n \geq 0}$ si chiama una *passeggiata aleatoria* su $\mathbb{R}$. Inoltre, per questo processo, la legge di X_0 si chiama la *legge iniziale*, mentre la legge comune delle Y_n si chiama la *legge del singolo passo*.

(2.2) Osservazione. Comunque si assegni una coppia (μ, ν) di misure di probabilità su $\mathcal{B}(\mathbb{R})$, è sempre possibile (usando lo schema delle prove indipendenti) costruire una passeggiata aleatoria su $\mathbb{R}$ ammettente μ come legge iniziale e ν come legge del singolo passo. Inoltre è chiaro che due siffatte passeggiate aleatorie, se considerate come variabili aleatorie a valori nello spazio canonico $\mathbb{R}^{\mathbb{N}}$, hanno leggi identiche: queste leggi si ottengono infatti prendendo l'immagine della misura prodotto $\mu \otimes \nu \otimes \nu \otimes \cdots$ mediante una medesima applicazione misurabile dello spazio canonico nello spazio canonico.

Il teorema che segue ribadisce l'osservazione precedente, rendendola più esplicita. Esso permette inoltre di ricondurre il concetto di passeggiata aleatoria su $\mathbb{R}$ al concetto, più generale, di catena di Markov.

(2.3) Teorema. *Siano μ, ν due misure di probabilità sulla tribù boreliana di $\mathbb{R}$. Si denoti con N il nucleo di convoluzione associato a ν e con Q la realizzazione canonica di N. Allora ogni passeggiata aleatoria su $\mathbb{R}$ ammettente μ come legge iniziale e ν come legge del singolo passo è una catena di Markov compatibile col nucleo N (ossia ammette come legge μQ).*

Dimostrazione.[*] Su uno spazio probabilizzato $(\Omega, \mathcal{A}, P)$ sia $[X_n]_{n\geq 0}$ una passeggiata aleatoria ammettente μ come legge iniziale e ν come legge del singolo passo. Per ogni intero positivo n, poniamo

$$Y_{n+1} = X_{n+1} - X_n, \qquad Z_n = \big[[X_0, \ldots, X_n],\, Y_{n+1}\big]. \tag{2.4}$$

Si ha allora $X_{n+1} = X_n + Y_{n+1}$ e quindi

$$X_{n+1} = h_n \circ Z_n = h_n \circ \big[[X_0, \ldots, X_n],\, Y_{n+1}\big], \tag{2.5}$$

dove h_n è la funzione reale così definita sullo spazio d'arrivo di Z_n:

$$h_n\big((x_0, \ldots, x_n),\, y\big) = x_n + y. \tag{2.6}$$

D'altra parte, la variabile aleatoria Y_{n+1} è indipendente dal blocco $[X_0, \ldots, X_n]$ ed ha come legge ν: essa ammette dunque, come versione della propria legge condizionale rispetto a $[X_0, \ldots, X_n]$, il nucleo costante K_n definito da

$$K_n\big((x_0, \ldots, x_n),\, \cdot\,\big) = \nu.$$

Grazie a (2.5), ne segue che una versione della legge condizionale di X_{n+1} rispetto al blocco $[X_0, \ldots, X_n]$ è il nucleo L_n caratterizzato da questa proprietà: per ogni elemento $(x_0, \ldots, x_n)$ dello spazio d'arrivo di $[X_0, \ldots, X_n]$, la misura $L_n\big((x_0, \ldots, x_n), \cdot\big)$ è l'immagine di ν mediante l'applicazione

$$y \mapsto h_n\big((x_0, \ldots, x_n),\, y\big).$$

Ma questa applicazione, grazie a (2.6), è semplicemente la traslazione $y \mapsto x_n + y$. Si vede dunque che L_n è il nucleo definito ponendo

$$L_n\big((x_0, \ldots, x_n),\, \cdot\,\big) = \epsilon_{x_n} \star \nu = N(x_n,\, \cdot\,).$$

Poiché questa conclusione vale per ogni n, è così provato che $[X_n]_{n\geq 0}$ è una catena di Markov compatibile col nucleo N. □

Arresto di una catena di Markov

Teorema. *Assegnato un nucleo markoviano N, relativo a uno spazio misurabile $(E, \mathcal{E})$, si consideri, su uno spazio probabilizzato $(\Omega, \mathcal{A}, P)$, una catena di Markov X con esso compatibile. Inoltre, fissato un elemento A di $\mathcal{E}$, si denoti con T il tempo d'arresto, relativo alla filtrazione naturale di X, così definito:*

$$T(\omega) := \inf\{n \in \mathbb{N} : X_n(\omega) \in A\}.$$

Allora il processo $X^{|T}$ ottenuto arrestando X al tempo T è una catena di Markov compatibile col nucleo M così definito:

$$M(x, \cdot) := \begin{cases} \epsilon_x & \text{se } x \in A, \\ N(x, \cdot) & \text{altrimenti.} \end{cases}$$

Dimostrazione. Denoteremo con $\mathcal{F}$ la filtrazione naturale di X, e con Y il processo arrestato $X^{|T}$. Fissiamo un intero positivo n e un elemento H di $\mathcal{F}_n$. Si ha allora

$$\{n \leq T\} \subset \{Y_{n+1} = X_{n+1}\} \cap \{Y_n = X_n \notin A\} \subset \{Nf \circ X_n = Mf \circ Y_n\}.$$

Di qui, poiché l'insieme $H \cap \{n < T\}$ appartiene alla tribù $\mathcal{F}_n$, si deduce (utilizzando la proprietà di Markov del processo X)

$$\begin{aligned} \int_{H\cap\{n<T\}} f \circ Y_{n+1}\, dP &= \int_{H\cap\{n<T\}} f \circ X_{n+1}\, dP \\ &= \int_{H\cap\{n<T\}} Nf \circ X_n\, dP \\ &= \int_{H\cap\{n<T\}} Mf \circ Y_n\, dP. \end{aligned} \tag{1}$$

Si ha inoltre

$$\{T \leq n\} \subset \{Y_{n+1} = Y_n = Y_T = X_T \in A\} \subset \{f \circ X_T = Mf \circ X_T\}$$

e quindi

$$\begin{aligned} \int_{H\cap\{T\leq n\}} f \circ Y_{n+1}\, dP &= \int_{H\cap\{T\leq n\}} Mf \circ X_T\, dP \\ &= \int_{H\cap\{T\leq n\}} Mf \circ Y_n\, dP. \end{aligned} \tag{2}$$

Sommando le relazioni (1) e (2), si ottiene infine

$$\int_H f \circ Y_{n+1}\, dP = \int_H Mf \circ Y_n\, dP.$$

Per l'arbitrarietà di n e di H, ciò prova che Y è una catena di Markov compatibile col nucleo M. □

Alcuni risultati sui processi additivi

1. Definizione di processo additivo

Tutti i processi considerati saranno tacitamente supposti *reali*. Inoltre tutte le filtrazioni e tutti i processi saranno supposti avere $\mathbb{R}_+$ come insieme dei tempi. Gli elementi di $\mathbb{R}_+$ saranno anche chiamati *istanti*.

È facile riconoscere che, se su uno spazio probabilizzabile $(\Omega, \mathcal{A})$ è dato un processo X le cui traiettorie siano continue a destra, allora X, pensato come funzione reale sullo spazio misurabile

$$\Big(\mathbb{R}_+ \times \Omega,\ \mathcal{B}(\mathbb{R}_+) \otimes \mathcal{A}\Big),$$

è misurabile. A questo scopo, supponendo per semplicità $X_0 = 0$, basta osservare che X è limite puntuale della successione $(X^{(n)})_{n\geq 0}$ così definita:

$$X^{(n)} := \sum_{k\geq 1} I_{](k-1)2^{-n},\, k2^{-n}]} \otimes X_{k2^{-n}}.$$

(1.1) Definizione Un processo X si dice *additivo* rispetto a una filtrazione $\mathcal{F}$ se, oltre ad essere adattato a $\mathcal{F}$, è tale che, per ogni istante s, il blocco $[X_t - X_s]_{t>s}$ sia indipendente da $\mathcal{F}_s$.

(1.2) Teorema. *Sia X un processo adattato alla filtrazione $\mathcal{F}$. Affinché X sia additivo rispetto a $\mathcal{F}$, è (necessario e) sufficiente che, per ogni coppia s, t d'istanti, con $s < t$, la variabile aleatoria $X_t - X_s$ sia indipendente da $\mathcal{F}_s$.*

Dimostrazione. Supposta soddisfatta la condizione dell'enunciato, fissiamo un istante s e un elemento H di $\mathcal{F}_s$. Basterà dimostrare che I_H è indipendente da ogni vettore aleatorio della forma

$$[X_{t_1} - X_s,\ X_{t_2} - X_s,\ \ldots,\ X_{t_n} - X_s], \tag{1.3}$$

con $s < t_1 < t_2 < \cdots < t_n$. A questo scopo, osserviamo che le componenti del vettore aleatorio

$$[I_H,\ X_{t_1} - X_s,\ X_{t_2} - X_{t_1},\ \ldots,\ X_{t_n} - X_{t_{n-1}}]$$

sono tra loro indipendenti: infatti ciascuna componente è indipendente dalla tribù generata dalle componenti di indice più piccolo. Ne segue che la prima componente è indipendente dal blocco delle altre, ossia dal vettore aleatorio

$$[X_{t_1} - X_s,\ X_{t_2} - X_{t_1},\ \ldots,\ X_{t_n} - X_{t_{n-1}}], \tag{1.4}$$

e quindi anche dal vettore aleatorio (1.3) (le cui componenti sono le somme parziali di quelle del vettore aleatorio (1.4)). L'asserzione è così dimostrata. □

(1.5) Definizione. Un processo si dice *additivo* (o anche *intrinsecamente additivo*) se esso è additivo rispetto alla propria filtrazione naturale.

(1.6) Osservazione. È facile riconoscere che, affinché un processo X sia additivo, occorre e basta che esso sia *a incrementi indipendenti*, nel senso che, per ogni successione finita $(t_j)_{0\le j\le n}$ d'istanti, con $0 = t_0 < t_1 < \cdots < t_n$, le variabili aleatorie $X_{t_0}, X_{t_1} - X_{t_0}, \ldots, X_{t_n} - X_{t_{n-1}}$ siano indipendenti.

2. Filtrazioni continue a destra

(2.1) Definizione. Ad ogni filtrazione $\mathcal{F}$ si può associare la filtrazione $\mathcal{G}$ definita ponendo, per ogni istante t,

$$\mathcal{G}_t = \bigcap_{\epsilon>0} \mathcal{F}_{t+\epsilon}. \tag{2.2}$$

Quando $\mathcal{G}$ coincida con $\mathcal{F}$, si dice che $\mathcal{F}$ è *continua a destra*. È facile verificare che, qualunque sia la filtrazione $\mathcal{F}$, la filtrazione $\mathcal{G}$ definita da (2.2) è sempre continua a destra: essa sarà chiamata, nel seguito, la filtrazione continua a destra *associata* a $\mathcal{F}$. Talvolta la si denoterà col simbolo $\mathcal{F}^+$.

(2.3) Definizione. Siano $\mathcal{F}$ una filtrazione e T una variabile aleatoria a valori in $[0,\infty]$. Si dirà che T è, rispetto a $\mathcal{F}$, un *tempo d'arresto* se, per ogni istante t, si ha $\{T \le t\} \in \mathcal{F}_t$. Si dirà invece che T è, rispetto a $\mathcal{F}$, un *tempo d'arresto in senso largo* (o, più brevemente, un *tempo d'arresto largo*) se, per ogni istante t, si ha $\{T < t\} \in \mathcal{F}_t$. Questa condizione, come facilmente si verifica, equivale al fatto che T sia un tempo d'arresto per la filtrazione $\mathcal{F}^+$.

(2.4) Teorema. *Sia X un processo additivo rispetto alla filtrazione $\mathcal{F}$ e avente traiettorie continue a destra. Allora X è additivo anche rispetto alla filtrazione continua a destra associata a $\mathcal{F}$.*

Dimostrazione. Denoteremo con $\mathcal{G}$ la filtrazione continua a destra associata a $\mathcal{F}$. Grazie a (1.2), basterà provare che, per ogni coppia (s,t) d'istanti, con $s < t$, la variabile aleatoria $X_t - X_s$ è indipendente dalla tribù $\mathcal{G}_s$. A questo scopo, osserviamo che l'istante s è limite di una successione decrescente $(s_n)_{n\ge1}$ d'istanti, con $s < s_n < t$. In corrispondenza, $X_t - X_s$ è limite puntuale della successione $(X_t - X_{s_n})_{n\ge1}$. Inoltre, per ogni n, la variabile aleatoria $X_t - X_{s_n}$ è indipendente dalla tribù $\mathcal{F}_{s_n}$ e quindi, a maggior ragione, dalla tribù $\mathcal{G}_s$ (che è contenuta in $\mathcal{F}_{s_n}$). Dunque, per ogni evento H, non trascurabile, appartenente a $\mathcal{G}_s$, la legge di $X_t - X_{s_n}$ secondo P_H è identica a quella secondo P. Passando al limite nel senso della convergenza debole delle leggi, si vede che la stessa proprietà vale per la variabile aleatoria $X_t - X_s$. Per l'arbitrarietà di H, ciò prova che $X_t - X_s$ è indipendente da $\mathcal{G}_s$. □

(2.5) Definizione. Per ogni intero positivo n, denoteremo con D_n l'insieme costituito dai numeri diadici di stadio n–esimo, cioè dai numeri della forma $k\,2^{-n}$, con $k \in \mathbb{N}$. Un tempo d'arresto T sarà detto *diadico* se esiste un intero positivo n tale che l'insieme dei valori finiti di T sia contenuto in D_n.

(2.6) Proposizione. *Su uno spazio probabilizzabile $(\Omega, \mathcal{A})$ siano date una filtrazione $\mathcal{F}$ e un tempo d'arresto largo T ad essa relativo. Allora esiste una successione decrescente $(T_n)_{n\geq 0}$ di tempi d'arresto diadici (relativi a $\mathcal{F}$) tale che si abbia*

$$\inf_n T_n = T, \qquad \{T_n < \infty\} = \{T < \infty\} \quad \text{per ogni } n. \tag{2.7}$$

Dimostrazione. Per ogni intero positivo n ed ogni elemento ω di Ω, poniamo (con la solita convenzione $\inf \emptyset = \infty$)

$$T_n(\omega) = \inf\{s \in D_n : T(\omega) < s\}.$$

Si ha allora, per ogni istante t,

$$\{T_n \leq t\} = \bigcup_{s \in D_n,\, s \leq t} \{T < s\} \in \mathcal{F}_t,$$

e ciò prova che T_n è un tempo d'arresto. Si ha inoltre

$$\{T < \infty\} = \left\{T < T_n \leq T + 2^{-n}\right\}.$$

La successione $(T_n)_{n\geq 0}$ possiede dunque le desiderate proprietà. □

3. La legge 0-1 di Blumenthal

(3.1) Teorema. *Su uno spazio probabilizzato $(\Omega, \mathcal{A}, P)$ sia X un processo intrinsecamente additivo, con traiettorie continue a destra. Si denoti con $\mathcal{F}$ la filtrazione naturale di X e con $\mathcal{G}$ la filtrazione continua a destra associata a $\mathcal{F}$.*

(a) Se la variabile aleatoria X_0 è degenere, tale è la tribù $\mathcal{G}_0$ ("Legge 0-1 di Blumenthal").

(b) Se, per giunta, si suppone che X_0 coincida quasi certamente con la costante 0, che ciascuna traiettoria del processo X sia continua in tutto un intorno dello zero, e che ciascuna delle X_t con $t > 0$ abbia legge simmetrica e diffusa, allora, per quasi ogni ω, lo zero è punto di accumulazione per ciascuno dei tre insiemi seguenti

$$\{t > 0 : X_t(\omega) > 0\}, \quad \{t > 0 : X_t(\omega) < 0\}, \quad \{t > 0 : X_t(\omega) = 0\}.$$

Dimostrazione. (a) Si supponga che X_0 coincida quasi certamente con la costante x. Grazie al Teorema (2.4), la tribù $\mathcal{G}_0$ è indipendente dal blocco $[X_t - X_0]_{t>0}$, ossia dal blocco $[X_t - x]_{t>0}$. Poiché la tribù generata da quest'ultimo blocco contiene $\mathcal{G}_0$, ne segue che la tribù $\mathcal{G}_0$ è indipendente da sé stessa, ossia degenere. □

(b) Mettiamoci ora nelle ipotesi di (b). Basterà provare che, per quasi ogni ω, lo zero è punto di accumulazione per *il primo* dei tre insiemi sopra considerati.

Ora, ciò si ottiene osservando che l'insieme degli ω provvisti della desiderata proprietà contiene l'evento

$$A = \limsup_n \{X_{1/n} > 0\},$$

il quale appartiene alla tribù degenere $\mathcal{G}_0$ e non può essere trascurabile, come risulta dalla relazione

$$P(A) \geq \limsup_n P\{X_{1/n} > 0\} = \tfrac{1}{2}. \qquad \square$$

(3.2) Teorema. *Su uno spazio probabilizzato (Ω, A, P) sia X un processo additivo, con traiettorie continue a destra. Si denoti con $\mathcal{F}$ la filtrazione naturale di X e con $\mathcal{G}$ la filtrazione continua a destra associata a $\mathcal{F}$. Allora, per ciascun istante t, ogni elemento di $\mathcal{G}_t$ è equivalente (secondo P) a un elemento di $\mathcal{F}_t$.*

Dimostrazione. Si denoti con $\mathcal{H}_t$ la tribù generata dal blocco $[X_u - X_t]_{u>t}$. Si ha allora

$$\mathcal{F}_t \subset \mathcal{G}_t \subset \mathcal{F}_t \vee \mathcal{H}_t. \tag{3.3}$$

Inoltre, grazie al Teorema (2.4), $\mathcal{H}_t$ è indipendente da $\mathcal{G}_t$. Ne segue che, per ogni elemento A di $\mathcal{F}_t$ ed ogni elemento B di $\mathcal{H}_t$, una versione della speranza condizionale $P\big[I_A I_B \,|\, \mathcal{G}_t\big]$ è la variabile aleatoria $I_A P(B)$ (misurabile rispetto a $\mathcal{F}_t$). Il teorema delle classi monotone assicura allora che ogni variabile aleatoria limitata e misurabile rispetto alla tribù $\mathcal{F}_t \vee \mathcal{H}_t$ ammette come versione della propria speranza condizionale rispetto a $\mathcal{G}_t$ una variabile aleatoria misurabile rispetto a $\mathcal{F}_t$. Ciò vale in particolare (grazie a (3.3)) per la funzione indicatrice di un elemento di $\mathcal{G}_t$. Dato dunque un elemento C di $\mathcal{G}_t$, la speranza condizionale $P\big[I_C \,|\, \mathcal{G}_t\big]$, una cui versione è, banalmente, la stessa I_C, ammette anche una versione V misurabile rispetto a $\mathcal{F}_t$: pertanto C risulta equivalente, secondo P, all'evento $\{V = 1\}$, che è un elemento di $\mathcal{F}_t$. □

(3.4) Corollario. *Sia $\mathcal{F}$ la filtrazione naturale di un processo intrinsecamente additivo X, con traiettorie continue a destra. Per ogni istante t, si denoti con $\widetilde{\mathcal{F}}_t$ la tribù costituita dagli eventi che sono equivalenti (secondo P) a un elemento di $\mathcal{F}_t$. Allora la famiglia $\big(\widetilde{\mathcal{F}}_t\big)_{t\geq 0}$ è una filtrazione continua a destra.*

Dimostrazione. Denoteremo con $\mathcal{G}$ la filtrazione continua a destra associata a $\mathcal{F}$. Sia A un elemento di $\bigcap_{\epsilon>0} \widetilde{\mathcal{F}}_{t+\epsilon}$, e mostriamo che esso appartiene a $\widetilde{\mathcal{F}}_t$. Per ogni intero n strettamente positivo, A è un elemento di $\widetilde{\mathcal{F}}_{t+1/n}$, ossia è equivalente a un elemento B_n di $\mathcal{F}_{t+1/n}$. Ne segue che A è equivalente all'evento $\liminf_n B_n$. Questo, a sua volta, appartenendo a $\mathcal{G}_t$, è equivalente, in virtù di (3.2), a un elemento di $\mathcal{F}_t$. Dunque, per transitività, A è equivalente a un elemento di $\mathcal{F}_t$, ossia appartiene a $\widetilde{\mathcal{F}}_t$. □

4. Processi additivi omogenei

(4.1) Definizione. Un processo X si dice *omogeneo* se, per ogni coppia (s,t) di numeri reali positivi, la variabile aleatoria $X_{s+t} - X_s$ è isonoma a $X_t - X_0$. In termini meno formali: X si dice omogeneo se, per ogni coppia d'intervalli temporali di eguale ampiezza, i corrispondenti incrementi di X sono tra loro isonomi.

(4.2) Teorema. *Sia X un processo omogeneo e intrinsecamente additivo. Allora, per ogni istante s, il processo $[X_{s+t} - X_s]_{t\geq 0}$ è isonomo al processo $[X_t - X_0]_{t\geq 0}$.*

Dimostrazione. Senza ledere la generalità, si può supporre $X_0 = 0$. Fissato l'istante s, poniamo

$$Y = [X_{s+t} - X_s]_{t\geq 0}.$$

Allora l'incremento di Y relativo a un fissato intervallo temporale coincide con l'incremento di X relativo a un intervallo temporale di eguale ampiezza. Di conseguenza, per ogni successione finita $(t_j)_{0\leq j\leq n}$ d'istanti, con $0 = t_0 < t_1 < \cdots < t_n$, i due vettori aleatori

$$[X_{t_j} - X_{t_{j-1}}]_{1\leq j\leq n}\,, \qquad [Y_{t_j} - Y_{t_{j-1}}]_{1\leq j\leq n} \tag{4.3}$$

sono tra loro isonomi. (Infatti ciascuno di essi ha componenti indipendenti, e inoltre ciascuna componente del primo vettore è isonoma alla componente di egual indice del secondo.) Ne segue che anche i due vettori aleatori

$$[X_{t_j}]_{1\leq j\leq n}\,, \qquad [Y_{t_j}]_{1\leq j\leq n}$$

sono tra loro isonomi. Infatti il primo (risp. il secondo) ha come componenti le somme parziali delle componenti del primo (risp. del secondo) dei due vettori aleatori (4.3). Per l'arbitrarietà della n–upla $(t_j)_{1\leq j\leq n}$, ciò basta per concludere che i due processi X, Y sono tra loro isonomi. □

(4.4) Teorema.* *Su uno spazio probabilizzato $(\Omega, \mathcal{A}, P)$ sia X un processo additivo per la filtrazione $\mathcal{F}$, omogeneo e con traiettorie continue a destra. Sia inoltre T un tempo d'arresto largo per $\mathcal{F}$, non equivalente alla costante $+\infty$. Si ponga*

$$Q = P\big(\,\cdot \mid \{T < \infty\}\big)$$

e si denoti con Y il processo che è nullo su $\{T = \infty\}$ e che è definito altrove da $Y_t(\omega) = X_{T+t}(\omega) - X_T(\omega)$.

Allora la legge del processo Y secondo Q coincide con la legge del processo $[X_t - X_0]_{t\geq 0}$ secondo P.

Dimostrazione. Senza ledere la generalità, supporremo $X_0 = 0$. Si tratta allora di dimostrare che, comunque si fissi una parte H di $\mathbb{R}_+$, finita e non vuota, risulta

$$\textstyle\int_{\{T<\infty\}} f\left([Y_t]_{t\in H}\right) dP = P\{T < \infty\} \int f\left([X_t]_{t\in H}\right) dP$$

per ogni funzione f continua e limitata sullo spazio d'arrivo del blocco $[X_t]_{t\in H}$.

Grazie a (2.6), ci si può limitare al caso particolare in cui T sia un tempo d'arresto diadico (nel senso di (2.5)). Infatti nel caso generale, scelta una successione $(T_n)_{n\geq 0}$ di tempi d'arresto diadici come in (2.6), e considerato, per ogni n, il processo Y^n definito come Y, ma partendo da T_n anziché da T, si vede che le traiettorie di Y^n convergono puntualmente verso le corrispondenti traiettorie di Y.

Supponiamo dunque che T sia un tempo d'arresto diadico. Basterà allora provare che, se s è uno qualsiasi dei possibili valori finiti di T, si ha

$$\textstyle\int_{\{T=s\}} f\left([Y_t]_{t\in H}\right) dP = P\{T = s\} \int f\left([X_t]_{t\in H}\right) dP.$$

Ora ciò discende dal fatto che, sull'evento $\{T = s\}$ (appartenente a $\mathcal{F}_s$), il processo Y coincide col processo $[X_{s+t} - X_s]_{t\geq 0}$, il quale è indipendente da $\mathcal{F}_s$ e, grazie a (4.2), isonomo a X. □

(4.5) Corollario. *Nelle stesse ipotesi del teorema precedente, si denoti con $\mathcal{G}$ la filtrazione continua a destra associata a $\mathcal{F}$. Allora il processo Y è, secondo Q, additivo rispetto alla filtrazione $(\mathcal{G}_{T+t})_{t\geq 0}$.*

Dimostrazione. Si tratta di dimostrare che, per ogni coppia s, t di numeri reali positivi, la variabile aleatoria $Y_{s+t} - Y_s$ è, secondo Q, indipendente da $\mathcal{G}_{T+s}$. Senza ledere la generalità, si può supporre $s = 0$. (Il caso generale si può infatti ottenere applicando il caso particolare al tempo d'arresto largo $T+s$.) Fissiamo dunque un elemento H di $\mathcal{G}_T$, non trascurabile secondo Q, e proviamo che la variabile aleatoria $Y_t - Y_0 = Y_t$ ha la stessa legge secondo Q e secondo

$$Q_H = P(\bullet \mid H \cap \{T < \infty\}).$$

A questo scopo, consideriamo il tempo d'arresto largo T_H che coincide con T su H e con ∞ altrove, e denotiamo con Y' il processo ottenuto come Y, ma partendo da T_H anziché da T. Si ha allora

$$H \cap \{T < \infty\} = \{T_H < \infty\},$$

sicché, grazie al teorema precedente, la legge di Y'_t secondo Q_H coincide con la legge di $X_t - X_0$ secondo P. Inoltre, secondo la misura Q_H, la quale è concentrata sull'insieme $H \cap \{T < \infty\}$, le due variabili aleatorie Y_t, Y'_t sono equivalenti, e quindi isonome. Si ha dunque, in conclusione,

$$Y_t(Q_H) = Y'_t(Q_H) = (X_t - X_0)(P) = Y_t(Q)$$

(dove l'ultima eguaglianza discende dal teorema precedente). L'asserzione è così dimostrata. □

Il criterio di hölderianità di Kolmogorov

Denoteremo con D l'insieme dei numeri diadici e con D_n la parte di D costituita dai numeri diadici di stadio n–esimo (cioè della forma $k\,2^{-n}$, con k intero positivo). Porremo inoltre $D^* = D \setminus \{0\}$, $D_n^* = D_n \setminus \{0\}$.

Lemma. *Siano m un intero strettamente positivo, f una funzione reale definita su $D \cap]0,m]$ e γ un numero reale strettamente positivo. Si supponga che esista un intero positivo n_0 tale che, per ogni intero n superiore o eguale a n_0, e ogni coppia s,t di elementi consecutivi di $D_n \cap]0,m]$, si abbia $|f(s) - f(t)| \le 2^{-n\gamma}$.*

Allora f è hölderiana di esponente γ (ossia prolungabile in una funzione hölderiana di esponente γ su $[0,m]$).

Dimostrazione. Mostriamo dapprima che esiste una costante reale C tale che, comunque si prendano un intero n superiore o eguale a n_0, una coppia a,b di elementi consecutivi di $D_n \cap]0,m]$ e un elemento x di D tra di essi compreso, risulti

$$|f(x) - f(a)| \le C\,2^{-n\gamma}, \qquad |f(x) - f(b)| \le C\,2^{-n\gamma}.$$

A questo scopo, osserviamo che si ha

$$x = a + \textstyle\sum_{h\ge 1} \epsilon_h\, 2^{-(n+h)},$$

con $(\epsilon_h)_{h\ge 1}$ successione di elementi di $\{0,1\}$, definitivamente nulla.

Pertanto, ponendo $s_0 = a$ e, per ogni $h \ge 1$,

$$s_h = s_{h-1} + \epsilon_h\, 2^{-(n+h)},$$

si definisce una successione crescente $(s_h)_{h\ge 0}$ i cui termini coincidono definitivamente con x. Inoltre, per ogni $h \ge 1$, i due termini s_{h-1}, s_h o coincidono o sono elementi consecutivi di D_{n+h}. Si ha pertanto

$$|f(x) - f(a)| \le \textstyle\sum_{h\ge 1} |f(s_h) - f(s_{h-1})| \le \sum_{h\ge 1} 2^{-(n+h)\gamma} = C\,2^{-n\gamma},$$

con C costante. Un ragionamento analogo si applica a $|f(x) - f(b)|$.

Presi ora due punti x, y di $D \cap]0,m]$, con

$$x < y, \qquad |x - y| < 2^{-n_0},$$

si consideri l'intero n (superiore o eguale a n_0) caratterizzato dalla relazione

$$2^{-(n+1)} \le |x - y| < 2^{-n}.$$

Esistono allora in $D_n \cap]0,m]$ due punti consecutivi a,b, con $a \le x \le y \le b$, oppure tre punti consecutivi a,b,c, con $a \le x \le b \le y \le c$. In entrambi i casi, risulta

$$\begin{aligned} |f(x) - f(y)| &\le |f(x) - f(b)| + |f(b) - f(y)| \\ &\le 2C\,2^{-n\gamma} = K\,2^{-(n+1)\gamma} \le K\,|x - y|^{\gamma} \end{aligned}$$

(con K opportuna costante). La funzione f è dunque uniformemente continua, ossia univocamente prolungabile in una funzione g continua su $[0,m]$. Inoltre questo prolungamento è una funzione hölderiana, di esponente γ, perché verifica la relazione

$$|g(x) - g(y)| \le K|x-y|^\gamma$$

per ogni coppia x, y di elementi di $[0,m]$ con $|x-y| < 2^{-n_0}$. □

Teorema. *Su uno spazio probabilizzato $(\Omega, \mathcal{A}, P)$ sia dato un processo reale X avente D^* come insieme dei tempi. Siano poi α, δ, C costanti reali strettamente positive, tali che, qualunque sia l'intero positivo n, la relazione*

$$P\big[|X_s - X_t|^\alpha\big] \le C\,|s-t|^{1+\delta}$$

abbia luogo per ogni coppia s,t di elementi consecutivi di D_n^*.

Per ogni intero $m \ge 1$ e ogni numero reale $\gamma > 0$, si denoti con $A_{m,\gamma}$ l'evento costituito dagli ω tali che la restrizione della traiettoria $X_\bullet(\omega)$ all'insieme $D \cap]0,m]$ sia hölderiana di esponente γ.

Allora l'evento $A_{m,\gamma}$ è quasi-certo se γ è minore di δ/α.

Dimostrazione. Denotiamo con Z_n l'inviluppo superiore di tutte le variabili aleatorie della forma $|X_s - X_t|$, con s,t coppia di elementi consecutivi dell'insieme $D_n \cap]0,m]$. Grazie al lemma precedente, si ha

$$\liminf_n \big\{ Z_n \le 2^{-n\gamma} \big\} \subset A_{m,\gamma}.$$

Basta perciò provare che, supposto $\gamma < \delta/\alpha$, il primo membro della precedente inclusione è quasi certo, ossia l'evento complementare

$$\limsup_n \big\{ Z_n > 2^{-n\gamma} \big\}$$

è trascurabile. Per questo, osserviamo che, per ogni coppia s,t di elementi consecutivi di $D_n \cap]0,m]$, si ha (grazie alla diseguaglianza di Markov e all'ipotesi)

$$P\big\{|X_s - X_t| > 2^{-n\gamma}\big\} \le 2^{n\gamma\alpha} P\big[|X_s - X_t|^\alpha\big] \le 2^{n\gamma\alpha} C\, 2^{-n(1+\delta)}.$$

Ne segue

$$P\big\{Z_n > 2^{-n\gamma}\big\} \le m\,2^n \cdot 2^{n\gamma\alpha} C\, 2^{-n(1+\delta)} = m\,C\,2^{n(\gamma\alpha - \delta)}.$$

Ciò prova l'asserzione, grazie al primo lemma di Borel–Cantelli. □

Corollario. *Su uno spazio probabilizzato $(\Omega, \mathcal{A}, P)$ sia dato un processo reale X avente D^* come insieme dei tempi e tale che, per ogni intero positivo n e ogni coppia s,t di elementi di D_n^* con $t - s = 2^{-n}$, l'incremento $X_t - X_s$ abbia legge normale centrata, di varianza $t-s$.*

Allora (con le stesse notazioni del teorema precedente) l'evento $A_{m,\gamma}$ è quasi certo se γ è minore di $\frac{1}{2}$.

Dimostrazione. Fissiamo un numero reale α strettamente positivo, un intero positivo n e una coppia s,t di elementi di $D_n \cap]0,m]$ con $t - s = 2^{-n}$. Si ha allora

$$P\big[|X_t - X_s|^\alpha\big] = C_\alpha\,|t-s|^{\alpha/2},$$

dove C_α denota il momento assoluto di ordine α della legge normale $\mathcal{N}(0,1)$. Perciò, supposto $\frac{\alpha}{2} > 1$, cioè $\alpha > 2$, se si applica il teorema precedente (con $\delta = \frac{\alpha}{2} - 1$), si vede che, per ogni numero reale γ verificante la relazione

$$0 < \gamma < \tfrac{1}{2} - \tfrac{1}{\alpha},$$

l'evento $A_{m,\gamma}$ è quasi certo. Basta allora osservare che i numeri γ tali che esista un α maggiore di 2 per il quale valga la relazione precedente sono esattamente i numeri γ strettamente compresi tra 0 e $\frac{1}{2}$. □

Processi di Wiener e misura di Wiener

1. Definizioni, notazioni, risultati preliminari

Si chiama *processo di Wiener* (o anche *moto browniano*) un processo reale X ammettente $\mathbb{R}_+$ come insieme dei tempi, avente traiettorie continue, additivo (cioè con incrementi indipendenti) e tale che, per ogni coppia (t,u) di elementi di $\mathbb{R}_+$ con $t<u$, l'incremento $X_u - X_t$ abbia legge normale $\mathcal{N}(0, u-t)$.

Denoteremo con D l'insieme costituito dai numeri diadici, cioè dai numeri della forma $k2^{-m}$, con k e m interi positivi. Per ogni intero n strettamente positivo, denoteremo con S_n l'insieme costituito dai punti s di $\mathbb{R}^n$ le cui coordinate verifichino la relazione $0 \le s_1 < \ldots < s_n$. Porremo inoltre $S = \bigcup_{n\ge 1} S_n$. Un elemento di S si dirà *diadico* se le sue coordinate appartengono all'insieme D. Se s è un elemento di S_n, la partizione $(J_k)_{1\le k\le n}$ dell'intervallo $]0, s_n]$ definita da

$$J_1 =]0, s_1], \quad J_2 =]s_1, s_2], \ \ldots, \ J_n =]s_{n-1}, s_n] \tag{1.1}$$

sarà chiamata la *partizione associata* a s. Si denoterà inoltre con ν_s la legge su $\mathbb{R}^n$ definita da

$$\nu_s = \bigotimes_{k=1}^{n} \mathcal{N}(0, \lambda(J_k)) \tag{1.2}$$

(dove λ denota la misura di Lebesgue).

Se X è un processo reale, avente come insieme dei tempi una parte I di $\mathbb{R}_+$, allora, per ogni intero n strettamente positivo ed ogni elemento s di S_n le cui coordinate appartengano a I, denoteremo con $\Delta_s X$ il vettore aleatorio (a valori in $\mathbb{R}^n$) così definito:

$$\Delta_s X = \left[X_{s_1}, X_{s_2} - X_{s_1}, \ \ldots, \ X_{s_n} - X_{s_{n-1}}\right]. \tag{1.3}$$

Usando le notazioni appena introdotte, è possibile caratterizzare un processo di Wiener *uscente dall'origine* (cioè avente ϵ_0 come legge iniziale) nel modo seguente.

(1.4) Proposizione. *Su uno spazio probabilizzato $(\Omega, \mathcal{A}, P)$ sia X un processo reale, ammettente $\mathbb{R}_+$ come insieme dei tempi e avente traiettorie continue. Le condizioni che seguono sono equivalenti:*

(a) X è un processo di Wiener uscente dall'origine.

(b) Per ogni elemento s di S, si ha

$$\Delta_s X(P) = \nu_s. \tag{1.5}$$

Dimostrazione. Osserviamo innanzitutto che, se è verificata la condizione (b), allora, in particolare, si ha $X_t(P) = \mathcal{N}(0,t)$ per ogni numero reale t strettamente positivo, sicché la variabile aleatoria X_0, essendo limite puntuale della successione $(X_{1/n})_{n\ge 1}$,

ammette come legge ϵ_0. Grazie a questa osservazione, si può supporre *a priori* che X sia un processo uscente dall'origine. Allora, per ogni intero n strettamente positivo ed ogni elemento s di S_n, poiché il vettore aleatorio $\Delta_s X$ definito da (1.3) coincide quasi certamente col vettore aleatorio

$$\big[X_{s_1} - X_0,\, X_{s_2} - X_{s_1},\, \ldots,\, X_{s_n} - X_{s_{n-1}}\big], \tag{1.6}$$

l'eguaglianza (1.5) equivale al fatto che le componenti di quest'ultimo vettore aleatorio siano indipendenti e ammettano come leggi

$$\mathcal{N}(0, s_1),\ \mathcal{N}(0,\, s_2 - s_1),\ \ldots,\ \mathcal{N}(0,\, s_n - s_{n-1}).$$

Di conseguenza, il fatto che l'eguaglianza (1.5) abbia luogo per ogni elemento s di S equivale al fatto che X sia un processo di Wiener. □

Il lemma che segue permette di indebolire la condizione (b) della proposizione precedente.

(1.7) Lemma. *Su uno spazio probabilizzato $(\Omega, \mathcal{A}, P)$ sia X un processo reale, ammettente $\mathbb{R}_+$ come insieme dei tempi e avente traiettorie continue. Affinché X sia un processo di Wiener uscente dall'origine, è sufficiente che l'eguaglianza (1.5) abbia luogo per ogni elemento s di S che sia diadico.*

Dimostrazione. Cominciamo con una semplice osservazione: due leggi μ, μ' su $\mathbb{R}^n$, tali che si abbia $\int f\,\mathrm{d}\mu = \int f\,\mathrm{d}\mu'$ per ogni funzione f continua e limitata su $\mathbb{R}^n$, sono identiche, in quanto coincidono su ogni insieme aperto; infatti la funzione indicatrice di un tal insieme è l'inviluppo superiore di una successione crescente di funzioni continue a valori in $[0,1]$. Tenendo conto di questa osservazione, si vede che, per un arbitrario elemento s di S_n, la relazione (1.5) equivale al fatto che, per ogni funzione f continua e limitata su $\mathbb{R}^n$, abbia luogo l'eguaglianza $P\big[f(\Delta_s X)\big] = \int f\,\mathrm{d}\nu_s$, ossia

$$P\big[f(\Delta_s X)\big] = \textstyle\int_{\mathbb{R}^n} f(x)\, g_s(x)\,\mathrm{d}x, \tag{1.8}$$

dove g_s denota la densità di ν_s (o, meglio, la versione continua di questa densità). D'altra parte, le due funzioni

$$s \mapsto P\big[f(\Delta_s X)\big], \qquad s \mapsto \textstyle\int_{\mathbb{R}^n} f(x)\, g_s(x)\,\mathrm{d}x$$

sono continue su S_n: la prima per il teorema di Lebesgue sulla convergenza dominata (grazie alla continuità di f e delle traiettorie di X); la seconda per il teorema di Scheffé. Dunque, affinché l'eguaglianza (1.8) abbia luogo per ogni elemento s di S_n, è sufficiente che essa abbia luogo per ogni s che sia diadico: infatti gli elementi diadici di S_n formano un insieme denso in S_n. □

Il lemma che segue permette di preparare la costruzione di un processo di Wiener mediante la costruzione preliminare di un processo che si può considerare come l'analogo di un processo di Wiener nel quadro dei processi con tempi diadici.

(1.9) Lemma. *È possibile costruire uno spazio probabilizzato $(\Omega, \mathcal{A}, P)$ e, su di esso, un processo X, avente come insieme dei tempi l'insieme D^* dei numeri diadici strettamente positivi, in modo tale che, per ogni elemento diadico s di S, abbia luogo l'eguaglianza (1.5).*

Dimostrazione. Prendiamo come spazio Ω il prodotto $\mathbb{R}^{D^*}$, come tribù $\mathcal{A}$ la tribù prodotto $\mathcal{B}(\mathbb{R})^{\otimes D^*}$ e come processo X quello *canonico*, cioè quello per il quale X_t è la proiezione canonica di indice t di Ω su $\mathbb{R}$. Grazie a questa scelta, per ogni intero n strettamente positivo ed ogni elemento diadico s di S_n, si ha $[X_{s_1}, \dots, X_{s_n}](\Omega) = \mathbb{R}^n$ e quindi anche $\Delta_s X(\Omega) = \mathbb{R}^n$. Di conseguenza, se si pone

$$\mathcal{F}_s = \mathcal{T}(X_{s_1}, \dots, X_{s_n}) = \mathcal{T}(\Delta_s X), \tag{1.10}$$

si può affermare che il nucleo, da $(\Omega, \mathcal{F}_s)$ a $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ associato a $\Delta_s X$ (cioè quello che trasforma ogni misura di probabilità su $(\Omega, \mathcal{F}_s)$ nella sua immagine mediante $\Delta_s X$) è invertibile. Esiste dunque, sulla tribù $\mathcal{F}_s$, un'unica misura di probabilità P_s secondo la quale $\Delta_s X$ abbia come legge ν_s. È facile ora verificare che, se t è un elemento diadico di S_{n+1}, tale che si abbia

$$\{s_1, \dots, s_n\} \subset \{t_1, \dots, t_{n+1}\}$$

(e quindi $\mathcal{F}_s \subset \mathcal{F}_t$), allora la misura P_t prolunga la misura P_s. In altri termini, si tratta di verificare che le n componenti del vettore aleatorio $\Delta_s X$ sono indipendenti secondo P_t e che le loro leggi secondo P_t sono identiche a quelle secondo P_s. Ciò si otterrà impiegando la proprietà associativa dell'indipendenza e il fatto che la classe delle leggi normali centrate è stabile per l'operazione di convoluzione. Cominciamo con l'osservare che, nel caso in cui l'unico elemento dell'insieme

$$\{t_1, \dots, t_{n+1}\} \setminus \{s_1, \dots, s_n\} \tag{1.11}$$

sia t_{n+1}, si ha $s_j = t_j$ per ogni j compreso tra 1 e n, e quindi ognuna delle componenti del vettore aleatorio $\Delta_s X$ è anche una componente di $\Delta_t X$, sicché la tesi è ovvia. Si può dunque escludere questo caso banale. Si ha allora $s_n = t_{n+1}$ e l'unico elemento dell'insieme (1.11) cade in uno degli n intervalli (1.1) della partizione associata a s. Se k è l'indice di questo intervallo, allora le componenti di $\Delta_s X$ con indice diverso da k sono anche componenti di $\Delta_t X$ (e quindi le loro leggi secondo P_t sono identiche a quelle secondo P_s), mentre la componente di $\Delta_s X$ di indice k è somma di due componenti di $\Delta_t X$ che, secondo P_t, sono tra loro indipendenti e hanno leggi normali centrate, con somma delle varianze eguale a $\lambda(J_k)$. Dunque, anche per la componente di indice k di $\Delta_s X$, la legge secondo P_t è identica a quella secondo P_s. Inoltre, secondo P_t, questa componente di $\Delta_s X$ è indipendente dal blocco delle altre $n-1$ componenti, e queste sono tra loro indipendenti. In conclusione, si vede che le n componenti di $\Delta_s X$ sono indipendenti secondo P_t e che le loro leggi secondo P_t sono identiche a quelle secondo P_s. L'affermazione è così dimostrata.

Osserviamo ora che, se H è una parte di D^* finita e non vuota, e se n è il suo cardinale, allora esiste un unico elemento diadico s di S_n tale che H si possa mettere nella forma $H = \{s_1, \dots, s_n\}$. Porremo allora

$$Q_H = Q_{\{s_1,\dots,s_n\}} = P_s. \tag{1.12}$$

Si ottiene così una famiglia $(Q_H)_H$ di misure di probabilità, la quale costituisce un *sistema proiettivo*, nel senso che, per ogni coppia H, K di parti di D^* finite e non vuote, con $H \subset K$, la misura Q_K prolunga la misura Q_H. (Per provare questa affermazione, basta osservare che essa è vera, grazie a quanto sopra dimostrato, nel caso particolare in cui l'insieme $K \setminus H$ sia costituito da un sol elemento). Un risultato ben noto permette allora di affermare che esiste, sulla tribù $\mathcal{A}$, un'unica misura di probabilità P che prolunghi simultaneamente tutte le Q_H, ossia tutte le P_s. Poiché questa misura possiede la proprietà pretesa nell'enunciato, il lemma è dimostrato. □

2. Costruzione di un processo di Wiener

(2.1) Teorema. *Esiste un processo di Wiener uscente dall'origine.*

Dimostrazione. Prendiamo le mosse da uno spazio probabilizzato $(\Omega, \mathcal{A}, P)$ sul quale esista un processo X con le proprietà descritte nel Lemma (1.9). Fissati un numero reale m strettamente positivo ed un esponente γ con $0 < \gamma < \frac{1}{2}$, denotiamo con $A_{m,\gamma}$ l'insieme costituito dagli elementi ω di Ω tali che la traiettoria $t \mapsto X_t(\omega)$ sia γ–hölderiana su $]0, m] \cap D$, cioè poniamo

$$A_{m,\gamma} = \left\{\sup_{t,u\in]0,m]\cap D,\, t\neq u} |X_t - X_u|/|t-u|^\gamma < \infty\right\}. \tag{2.2}$$

Grazie al lemma di hölderianità di Kolmogorov, $A_{m,\gamma}$ è un evento quasi certo. Tale è dunque anche l'intersezione degli eventi $A_{m,\gamma}$ con $m \in \mathbb{N}^*$ e $\gamma \in \left]0, \frac{1}{2}\right[\cap \mathbb{Q}$. Denoteremo con A questa intersezione. Osserviamo che, per ogni elemento ω di A, la traiettoria $t \mapsto X_t(\omega)$, essendo una funzione reale su D^* la cui restrizione a ciascun insieme della forma $]0, m] \cap D$ (con $m \in \mathbb{N}^*$) è uniformemente continua, può essere univocamente prolungata in una funzione continua su $\mathbb{R}_+$. Se dunque si pone, per ogni u di $\mathbb{R}_+$,

$$Y_u = I_A \liminf_{t\to u,\, t\in D^*} X_t, \tag{2.3}$$

si ottiene un processo Y, ammettente $\mathbb{R}_+$ come insieme dei tempi e avente traiettorie continue, il quale verifica, per ogni elemento t di D^*, la relazione $Y_t = I_A X_t$. Allora, per ogni elemento s di S che sia diadico, si ha

$$\Delta_s Y = I_A \Delta_s X \sim \Delta_s X \pmod P,$$

e quindi

$$(\Delta_s Y)(P) = (\Delta_s X)(P) = \nu_s.$$

Grazie al Lemma (1.7), ciò basta per concludere che Y è un processo di Wiener uscente dall'origine. □

(2.4) Osservazione. Il processo di Wiener Y costruito nella dimostrazione del teorema precedente possiede la seguente proprietà: per ogni elemento γ di $]0, \frac{1}{2}[$ ed ogni elemento ω di Ω, la traiettoria $t \mapsto Y_t(\omega)$ è localmente γ–hölderiana (nel senso che è γ–hölderiana su ciascuna parte limitata di $\mathbb{R}_+$): infatti essa è identicamente nulla se ω non appartiene ad A, mentre, se ω appartiene ad A, essa è il prolungamento continuo a $\mathbb{R}_+$ di una funzione localmente γ–hölderiana su D^* (cioè γ–hölderiana su ciascuna parte limitata di D^*).

3. La misura di Wiener

È ben noto che un processo reale avente $\mathbb{R}_+$ come insieme dei tempi può essere definito in vari modi, formalmente diversi, ma sostanzialmente equivalenti. Uno di questi modi consiste nel dire che, se $(\Omega, \mathcal{A})$ è uno spazio probabilizzabile, allora un processo reale su $(\Omega, \mathcal{A})$, avente $\mathbb{R}_+$ come insieme dei tempi, è una variabile aleatoria su $(\Omega, \mathcal{A})$ a valori nello spazio prodotto

$$\left(\mathbb{R}^{\mathbb{R}_+},\ \mathcal{B}(\mathbb{R})^{\otimes \mathbb{R}_+}\right).$$

Se, per ogni elemento t di $\mathbb{R}_+$, si denota con π_t la proiezione canonica, di indice t, di $\mathbb{R}^{\mathbb{R}_+}$ su $\mathbb{R}$, si può dire che la tribù $\mathcal{B}(\mathbb{R})^{\otimes \mathbb{R}_+}$ è la tribù generata su $\mathbb{R}^{\mathbb{R}_+}$ dalle π_t:

$$\mathcal{B}(\mathbb{R})^{\otimes \mathbb{R}_+} = \mathcal{T}\left(\pi_t :\ t \in \mathbb{R}_+\right). \tag{3.1}$$

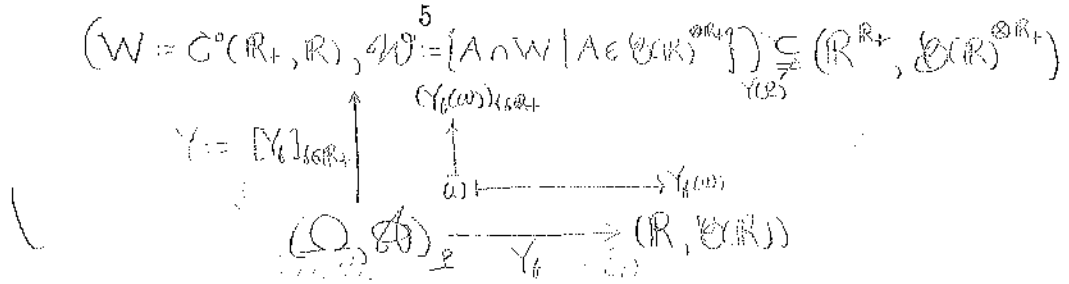
Essa è dunque l'unione di tutte le tribù della forma

$$\mathcal{T}\left(\pi_t :\ t \in H\right),\quad \text{con } H \text{ parte numerabile di } \mathbb{R}_+. \tag{3.2}$$

Denoteremo con W la parte di $\mathbb{R}^{\mathbb{R}_+}$ costituita dalle funzioni reali continue su $\mathbb{R}_+$. È facile verificare che l'insieme W non appartiene alla tribù (3.1) (cioè che esso non appartiene ad alcuna delle tribù della forma (3.2)). Denoteremo con $\mathcal{W}$ la *traccia* della tribù (3.1) su W, ossia la tribù su W costituita dalle intersezioni di W con i vari elementi della tribù (3.1). In modo equivalente, se si denota con ξ_t la restrizione di π_t a W, si può dire che $\mathcal{W}$ è la tribù generata su W dalle ξ_t:

$$\mathcal{W} = \mathcal{T}\left(\xi_t :\ t \in \mathbb{R}_+\right). \tag{3.3}$$

Lo spazio misurabile $(W, \mathcal{W})$ sarà chiamato lo *spazio di Wiener*. Un processo reale su $(\Omega, \mathcal{A})$, ammettente $\mathbb{R}_+$ come insieme dei tempi e avente traiettorie continue, si potrà allora considerare anche come una variabile aleatoria su $(\Omega, \mathcal{A})$ a valori nello spazio di Wiener: questo modo di concepire un tal processo sarà sistematicamente (e tacitamente) adottato nel seguito. Un particolarissimo processo del tipo appena descritto è il processo $\xi = [\xi_t]_{t \in \mathbb{R}_+}$: esso non è altro che l'applicazione identica di W. Lo chiameremo il *processo canonico* sullo spazio di Wiener.



(3.4) Definizione. Si chiama *misura di Wiener* una misura di probabilità sullo spazio di Wiener $(W, \mathcal{W})$, secondo la quale il processo canonico $[\xi_t]_{t\in\mathbb{R}_+}$ sia un processo di Wiener uscente dall'origine.

(3.5) Teorema. *Sullo spazio di Wiener esiste una ed una sola misura di Wiener. Inoltre, affinché un processo reale ammettente $\mathbb{R}_+$ come insieme dei tempi e avente traiettorie continue sia un processo di Wiener uscente dall'origine, occorre e basta che esso abbia come legge la misura di Wiener.*

Dimostrazione. Su uno spazio probabilizzato $(\Omega, \mathcal{A}, P)$ sia X un processo reale ammettente $\mathbb{R}_+$ come insieme dei tempi e avente traiettorie continue. Poiché, come sopra convenuto, il processo X va pensato come una variabile aleatoria a valori nello spazio di Wiener, ha senso considerare, per ogni elemento t di $\mathbb{R}_+$, la variabile aleatoria reale $\xi_t \circ X$. Inoltre questa coincide con X_t. Usando allora le notazioni introdotte all'inizio della Sezione 1, si può scrivere, per ogni elemento s di S,

$$\Delta_s X = (\Delta_s \xi) \circ X,$$

e quindi, indicando con μ la legge di X,

$$(\Delta_s X)(P) = (\Delta_s \xi)(\mu).$$

Di qui, grazie a (1.4), discende che, affinché X sia un processo di Wiener uscente dall'origine, occorre e basta che μ sia una misura di Wiener. Poiché, d'altra parte, un processo di Wiener uscente dall'origine esiste (si veda (2.1)), ne segue che esiste anche una misura di Wiener. Rimane solo da verificare che una tal misura è unica. A questo scopo, osserviamo che, per ogni intero n strettamente positivo ed ogni elemento s di S_n, due misure di Wiener danno la stessa legge al vettore aleatorio $\Delta_s \xi$, e quindi coincidono sulla tribù

$$\mathcal{T}(\Delta_s \xi) = \mathcal{T}(\xi_{s_1}, \ldots, \xi_{s_n}).$$

Poiché l'unione delle tribù di questo tipo è un'algebra che, grazie a (3.3), genera la tribù $\mathcal{W}$, ciò basta per concludere che le due misure sono identiche. □

(3.6) Corollario. *Si denoti con Γ l'insieme costituito da tutti gli elementi di W che siano localmente γ–hölderiani per ogni γ appartenente a $]0, \frac{1}{2}[$. Allora l'insieme Γ appartiene alla tribù $\mathcal{W}$ e ha misura eguale a 1 secondo la misura di Wiener.*

Dimostrazione. Per provare la prima affermazione, basta osservare che, a causa della continuità delle funzioni appartenenti a W, si ha

$$\Gamma = \bigcap_{m\in\mathbb{N}^*,\, \gamma\in]0,\frac{1}{2}[\cap\mathbb{Q}} \Big\{ \sup_{t,u\in]0,m[\cap D,\, t\neq u} |\xi_t - \xi_u|/|t-u|^\gamma < \infty \Big\}.$$

Per provare la seconda affermazione, basta osservare che, grazie a (2.4), esiste un processo di Wiener, uscente dall'origine, le cui traiettorie appartengano tutte a Γ. □

(3.7) Osservazione. Su uno spazio probabilizzato $(\Omega, \mathcal{A}, P)$ sia X un qualsiasi processo di Wiener uscente dall'origine. Poiché X ha come legge la misura di Wiener, risulta dal corollario precedente che l'evento $\{X \in \Gamma\}$ (costituito dagli elementi di Ω tali che la corrispondente traiettoria di X appartenga a Γ) è quasi certo.

Alcune proprietà delle traiettorie di un moto browniano

1. Non derivabilità

(1.1) Teorema. *Su uno spazio probabilizzato $(\Omega, \mathcal{A}, P)$ sia X un moto browniano. Allora, per ogni numero reale γ maggiore di $\frac{1}{2}$, l'insieme*

$$\bigcup_{s\in\mathbb{R}_+} \left\{\limsup_{t\to s,\, t>s} |X_t - X_s|/(t-s)^\gamma < \infty\right\}$$

è esternamente trascurabile (ossia contenuto in un evento trascurabile). [☆]

In particolare, per $\gamma = 1$, il precedente enunciato si riduce a dire che l'insieme costituito dagli elementi ω di Ω per i quali la traiettoria $X_\bullet(\omega)$ ammetta, in qualche punto, derivata destra inferiore e derivata destra superiore entrambe finite è esternamente trascurabile. Del risultato generale daremo una dimostrazione che è una naturale estensione (suggerita da Alessio Figalli) della dimostrazione impiegata da Dvorezki, Erdös e Kakutani nel caso $\gamma = 1$. Premetteremo un'osservazione e un lemma.

(1.2) Osservazione.* Se, sullo spazio $(\Omega, \mathcal{A}, P)$, è data una variabile aleatoria reale Z con legge normale $\mathcal{N}(0,1)$ e se, per ogni numero reale x maggiore di zero, si pone

$$F(x) = P\{|Z| \le x\} = 2\cdot(2\pi)^{-1/2}\int_0^x \exp\left(-z^2/2\right) \mathrm{d}z,$$

allora, al tendere di x a zero, si ha $F(x) \approx x$ (nel senso che il rapporto $F(x)/x$ ammette un limite finito e non nullo).

(1.3) Lemma.* *Nelle ipotesi del Teorema (1.1), poniamo, per ogni coppia n, k d'interi strettamente positivi,*

$$J_k^n = \left[(k-1)/n,\ k/n\right[, \qquad \Delta_k^n = \left|X_{k/n} - X_{(k-1)/n}\right|. \tag{1.4}$$

Inoltre, fissati una costante reale C maggiore di zero e un intero h tale che si abbia $h(\gamma - \frac{1}{2}) > 1$, poniamo

$$D_n = \bigcup_{k=1}^n \left\{\Delta_{k+1}^n \vee \cdots \vee \Delta_{k+h}^n \le Cn^{-\gamma}\right\}.$$

Allora l'evento $\liminf_n D_n$ è trascurabile.

Dimostrazione. Grazie al lemma di Fatou, basterà provare la relazione

$$\lim_n P(D_n) = 0. \tag{1.5}$$

A questo scopo, osserviamo che, per ogni coppia (n,k) d'interi strettamente positivi, $(\Delta_{k+j}^n)_{1\le j\le h}$ è una h–upla di variabili aleatorie indipendenti, tutte isonome a Δ_1^n. Si può dunque scrivere (tenendo anche conto dell'Osservazione (1.2))

$$\begin{aligned} P(D_n) &\le \textstyle\sum_{k=1}^n P\left\{\Delta_{k+1}^n \vee \cdots \vee \Delta_{k+h}^n \le Cn^{-\gamma}\right\} \\ &= n\left(P\left\{\Delta_1^n \le Cn^{-\gamma}\right\}\right)^h \\ &= n\left(P\left\{\sqrt{n}\Delta_1^n \le Cn^{-(\gamma-\frac{1}{2})}\right\}\right)^h \approx n^{1-h(\gamma-\frac{1}{2})}, \end{aligned}$$

e ciò prova la relazione (1.5). □

Mantenendo le notazioni del lemma appena dimostrato, veniamo ora alla dimostrazione del Teorema (1.1). Fissata una costante reale c maggiore di zero, e posto, per ogni elemento s di $\mathbb{R}_+$,

$$A_s = \left\{\limsup_{t\to s,\, t>s} |X_t - X_s|/(t-s)^\gamma < c\right\},$$

basterà provare che l'insieme $\bigcup_{s\in[0,1[} A_s$ è esternamente trascurabile. Infatti, per estendere il risultato ad un insieme della forma $\bigcup_{s\in[m,m+1[} A_s$, con $m \in \mathbb{N}$, basterà considerare il processo $[X_{m+t} - X_m]_{t\in\mathbb{R}_+}$ (che è ancora un moto browniano).

Fissato l'intero h come nel lemma precedente, poniamo, per ogni elemento s di $[0,1[$ ed ogni intero n strettamente positivo,

$$B_{s,n} = \bigcap_{t\in[s,\, s+(h+1)/n]} \left\{|X_t - X_s| \le c\,(t-s)^\gamma\right\}.$$

Allora, per ogni elemento s di $[0,1[$, si ha $A_s \subset \liminf_n B_{s,n}$. Ponendo

$$A = \bigcup_{s\in[0,1[} A_s, \qquad B_n = \bigcup_{s\in[0,1[} B_{s,n},$$

se ne deduce

$$A \subset \liminf_n B_n. \tag{1.6}$$

D'altra parte, l'ovvia inclusione

$$B_{s,n} \subset \bigcap_{t\in[s,\, s+(h+1)/n]} \left\{|X_t - X_s| \le c\big((h+1)/n\big)^\gamma\right\}$$

mostra che, per ogni elemento ω di $B_{s,n}$, la traiettoria $X_\bullet(\omega)$ ha, sull'intervallo

$$\big[s,\ s+(h+1)/n\big], \tag{1.7}$$

un'oscillazione maggiorata dalla costante $2c\big((h+1)/n\big)^\gamma$: costante che, più semplicemente, scriveremo nella forma $Cn^{-\gamma}$. In particolare, se s appartiene a J_k^n, allora, per ogni elemento ω di $B_{s,n}$, la medesima costante $Cn^{-\gamma}$ maggiora le oscillazioni della traiettoria $X_\bullet(\omega)$ sugli h intervalli $J_{k+1}^n, \ldots, J_{k+h}^n$: ciascuno di questi è infatti contenuto nell'intervallo (1.7). Si ha dunque

$$\bigcup_{s\in J_k^n} B_{s,n} \subset \left\{\Delta_{k+1}^n \vee \cdots \vee \Delta_{k+h}^n \le Cn^{-\gamma}\right\}.$$

Se ora, in quest'ultima relazione, si prendono le unioni dei due membri (al variare di k da 1 a n), si trova

$$B_n = \bigcup_{k=1}^n \bigcup_{s\in J_k^n} B_{s,n} \subset \bigcup_{k=1}^n \left\{\Delta_{k+1}^n \vee \cdots \vee \Delta_{k+h}^n \le Cn^{-\gamma}\right\}. \tag{1.8}$$

Dalle relazioni (1.6) e (1.8) si deduce infine, indicando con D_n l'ultimo membro di (1.8),

$$A \subset \liminf_n D_n.$$

Poiché, grazie al lemma precedente, l'evento $\liminf_n D_n$ è trascurabile, il teorema è dimostrato. □

2. Ulteriori proprietà

(2.1) Teorema. *Su uno spazio probabilizzato $(\Omega, \mathcal{A}, P)$ sia X un moto browniano. Allora:*

(a) Per quasi ogni ω, la traiettoria $X_\bullet(\omega)$ non è monotona su alcun intervallo proprio (e quindi possiede un insieme denso di punti di massimo locale).

(b) Per quasi ogni ω, la traiettoria $X_\bullet(\omega)$ assume massimi diversi su ogni coppia d'intervalli propri, compatti e disgiunti, con estremi razionali (e quindi non possiede alcun punto di massimo locale che non sia stretto).

Dimostrazione. L'affermazione (a) è una conseguenza del Teorema (1.1). Infatti, se una traiettoria di X è monotona su un intervallo proprio, allora, grazie al teorema di Lebesgue–Vitali, essa possiede derivata finita in quasi ogni punto di questo intervallo.

Per provare l'affermazione (b), basta provare che, dati due intervalli compatti $H = [a, b]$, $K = [c, d]$, con $0 \le a < b < c < d < \infty$, e posto

$$M_H = \sup_{t \in H} X_t, \qquad M_K = \sup_{t \in K} X_t,$$

l'evento $\{M_H = M_K\}$ è trascurabile. Per questo, basta scrivere l'evento in questione nella forma $\{M_H - X_b = M_K - X_b\}$ e osservare che la variabile aleatoria $M_K - X_b$ è indipendente dal blocco $[X_t]_{0 \le t \le b}$ (dunque da $M_H - X_b$) e dotata di densità, in quanto può esser messa nella forma

$$M_K - X_b = (X_c - X_b) + \sup_{c \le t \le d} (X_t - X_c). \qquad \square$$

(2.2) Proposizione. *Siano dati un moto browniano X, una funzione boreliana f su $\mathbb{R}$, trascurabile secondo la misura di Lebesgue, e una misura σ–finita μ sulla tribù boreliana di $\mathbb{R}_+$, con $\mu\{0\} = 0$.*

Allora, per quasi ogni ω, la funzione $t \mapsto f\big(X_t(\omega)\big)$ è trascurabile secondo μ.

Dimostrazione. Poiché la trascurabilità di una funzione equivale a quella del suo modulo, possiamo supporre f positiva. Si ha allora, denotando con ν_t la legge di X_t,

$$\begin{aligned} \int P(\mathrm{d}\omega) \int \mu(\mathrm{d}t)\, f\big(X_t(\omega)\big) &= \int \mu(\mathrm{d}t) \int P(\mathrm{d}\omega)\, f\big(X_t(\omega)\big) \\ &= \int \mu(\mathrm{d}t) \int f \,\mathrm{d}\nu_t = 0. \end{aligned}$$

Ciò basta per concludere. $\square$

(2.3) Corollario. *Siano dati un moto browniano X e un insieme boreliano B di $\mathbb{R}$, trascurabile secondo la misura di Lebesgue.*

Allora, per quasi ogni ω, l'insieme $\{t : X_t(\omega) \in B\}$ è trascurabile secondo la misura di Lebesgue (e quindi privo di punti interni).

Dimostrazione. Basta applicare la proposizione precedente, nella quale si prenda come funzione f l'indicatrice di B e come misura μ la misura di Lebesgue sulla tribù boreliana di $\mathbb{R}_+$. $\square$

(2.4) Teorema. *Dato, sullo spazio probabilizzato $(\Omega, \mathcal{A}, P)$, un moto browniano X con $X_0 = 0$, si ponga, per ogni ω,*

$$A(\omega) = \{t : X_t(\omega) > 0\}, \qquad B(\omega) = \{t : X_t(\omega) < 0\},$$
$$C(\omega) = \{t : X_t(\omega) = 0\}.$$

Allora, per quasi ogni ω, l'insieme chiuso $C(\omega)$ possiede le proprietà seguenti:

(a) è illimitato;

(b) è trascurabile secondo Lebesgue (dunque privo di punti interni);

(c) è contenuto sia nella chiusura di $A(\omega)$ sia in quella di $B(\omega)$ (e quindi è identico all'intersezione di queste due chiusure).

Dimostrazione. La successione $(X_n)_{n\in\mathbb{N}}$ è una passeggiata aleatoria su $\mathbb{R}$, con legge (di un singolo passo) eguale a $\mathcal{N}(0,1)$. Perciò le due variabili aleatorie

$$\liminf_n X_n, \qquad \limsup_n X_n$$

sono equivalenti alle costanti $-\infty$, $+\infty$ rispettivamente. Ciò prova che, per quasi ogni ω, i due insiemi $A(\omega)$, $B(\omega)$ (e quindi anche l'insieme $C(\omega)$) sono illimitati. Senza ledere la generalità, potremo anche supporre che questa proprietà valga *per ogni* ω (senza eccezioni).

Poiché il singoletto $\{0\}$ è trascurabile secondo la misura di Lebesgue, tale è, per quasi ogni ω, l'insieme $C(\omega)$.

Per completare la dimostrazione, denotiamo con $C_0(\omega)$ l'insieme costituito dai punti di $C(\omega)$ che sono *isolati a sinistra* (cioè dall'origine e dai punti che siano estremi destri di componenti connesse dell'insieme aperto $A(\omega)\cup B(\omega)$). Osserviamo che ogni punto di $C(\omega) \setminus C_0(\omega)$ è limite di una successione strettamente crescente di elementi di $C(\omega)$ e che tra due termini consecutivi di una tal successione è compresa una componente connessa di $A(\omega)\cup B(\omega)$, e quindi un elemento di $C_0(\omega)$. Ciò mostra che $C(\omega)$ coincide con la chiusura di $C_0(\omega)$. Dunque, per completare la dimostrazione, basterà provare che, per quasi ogni ω, si ha

$$C_0(\omega) \subset \overline{A(\omega)} \cap \overline{B(\omega)}.$$

A questo scopo, consideriamo, per ogni istante razionale r, il tempo d'arresto T_r (relativo alla filtrazione naturale di X) così definito:

$$T_r(\omega) = \inf\{t : t \geq r,\ X_t(\omega) = 0\}.$$

Osserviamo che T_r è finito e che, per ogni ω, risulta $X_{T_r}(\omega) = 0$, ossia $T_r(\omega) \in C(\omega)$. Inoltre, per ogni ω, si ha

$$C_0(\omega) \subset \{T_r(\omega) : r \in \mathbb{Q}_+\}.$$

Infatti è chiaro innanzitutto che l'origine appartiene al secondo membro (essendo eguale a $T_0(\omega)$); se poi t è un elemento di $C_0(\omega)$ con $t > 0$, allora, preso un suo intorno sinistro nel quale non cada alcun punto di $C(\omega)$ distinto da t, e scelto in questo intorno un istante razionale r, si vede che t coincide con $T_r(\omega)$.

Basterà dunque provare che, per quasi ogni ω, si ha

$$\{T_r(\omega) : r \in \mathbb{Q}_+\} \subset \overline{A(\omega)} \cap \overline{B(\omega)}.$$

Ora ciò si ottiene osservando che, qualunque sia r, il processo $\{X_{T_r+t}\}_{t\geq 0}$ è un moto browniano uscente dall'origine, sicché, grazie alla Legge 0-1 di Blumenthal, ciascuno dei due eventi

$$\limsup_n \{X_{T_r+1/n} > 0\}, \qquad \limsup_n \{X_{T_r+1/n} < 0\}$$

è quasi certo.

Il teorema è così dimostrato. □

Sui punti di massimo locale per le traiettorie di un processo

Teorema. *Su uno spazio probabilizzabile $(\Omega, \mathcal{A})$ sia X un processo reale, con traiettorie continue, avente come insieme dei tempi uno spazio metrico T localmente compatto. Allora l'insieme A costituito dagli elementi ω di Ω tali che la traiettoria $X(\cdot, \omega)$ possieda un punto di massimo locale non stretto appartiene alla tribù $\mathcal{A}$.*

Dimostrazione. Fissiamo, per la topologia di T, una base numerabile $\mathcal{U}$, costituita da insiemi relativamente compatti e non vuoti. Sia inoltre D un insieme numerabile denso in T. Per ogni elemento U di $\mathcal{U}$, denoteremo con M_U la variabile aleatoria reale così definita su $(\Omega, \mathcal{A})$:

$$M_U(\omega) = \sup_{t \in U \cap D} X_t(\omega) = \max_{t \in \overline{U}} X_t(\omega). \tag{1}$$

Sia ω un elemento dell'insieme A definito nell'enunciato del teorema. Ciò vuol dire che esiste un elemento t_0 di T che sia punto di massimo locale non stretto per la traiettoria $X(\cdot, \omega)$. Poiché, per questa traiettoria, t_0 è punto di massimo locale, esiste un elemento W di $\mathcal{U}$ con

$$t_0 \in W, \qquad X_{t_0}(\omega) = M_W(\omega). \tag{2}$$

Naturalmente, esiste anche un elemento V di $\mathcal{U}$ con

$$t_0 \in V \subset \overline{V} \subset W. \tag{3}$$

Sia ora $\mathcal{R}$ un arbitrario ricoprimento finito di $\overline{V}$ costituito da elementi di $\mathcal{U}$ contenuti in W. Allora t_0 appartiene ad uno almeno degli elementi di $\mathcal{R}$. Sia U un tal elemento. Poiché t_0 non è punto di massimo locale *in senso stretto* per la traiettoria $X(\cdot, \omega)$, esiste in U un punto t_1, distinto da t_0, con

$$X_{t_1}(\omega) = X_{t_0}(\omega) = M_W(\omega). \tag{4}$$

Si può dunque trovare una coppia (T_0, T_1) di elementi di $\mathcal{U}$ tale che si abbia

$$t_0 \in T_0, \qquad t_1 \in T_1, \qquad T_0 \cup T_1 \subset U, \qquad \overline{T_0} \cap \overline{T_1} = \emptyset \tag{5}$$

e quindi $M_{T_0}(\omega) = M_{T_1}(\omega) = M_W(\omega)$.

Affinché il risultato così ottenuto si lasci tradurre in una semplice formula, sarà necessario introdurre alcune notazioni. Cominciamo col porre

$$\mathcal{K} = \Big\{(V,W) \in \mathcal{U}\times\mathcal{U} :\ \overline{V} \subset W\Big\}. \tag{6}$$

Inoltre, per ogni elemento (V,W) di $\mathcal{K}$, denotiamo con $\rho(V,W)$ l'insieme (numerabile) dei ricoprimenti finiti di $\overline{V}$ costituiti da elementi di $\mathcal{U}$ contenuti in W. Poniamo poi, per ogni elemento U di $\mathcal{U}$,

$$\mathcal{T}(U) = \Big\{(T_0,T_1) \in \mathcal{U}\times\mathcal{U} :\ \ T_0 \cup T_1 \subset U,\ \ \overline{T_0} \cap \overline{T_1} = \emptyset\Big\}. \tag{7}$$

Poniamo infine

$$B = \bigcup_{(V,W)\in\mathcal{K}}\ \bigcap_{\mathcal{R}\in\rho(V,W)}\ \bigcup_{U\in\mathcal{R}}\ \bigcup_{(T_0,T_1)\in\mathcal{T}(U)} \big\{M_{T_0} = M_{T_1} = M_W\big\}. \tag{8}$$

Utilizzando queste notazioni, il risultato precedente si traduce nell'inclusione $A \subset B$. Poiché, come risulta dalla relazione (8), B è un elemento della tribù $\mathcal{A}$, basterà, per concludere la dimostrazione del teorema, provare l'inclusione opposta: $B \subset A$. Sia dunque ω un elemento di B. Ciò significa che esiste un elemento (V,W) di $\mathcal{K}$ con la proprietà seguente: comunque si scelga un ricoprimento $\mathcal{R}$ di $\overline{V}$ appartenente a $\rho(V,W)$, esistono un elemento U di $\mathcal{R}$ e due elementi T_0, T_1 di $\mathcal{U}$ tali che si abbia

$$T_0 \cup T_1 \subset U, \qquad \overline{T_0} \cap \overline{T_1} = \emptyset, \qquad M_{T_0}(\omega) = M_{T_1}(\omega) = M_W(\omega). \tag{9}$$

Scegliamo, per ogni intero n strettamente positivo, un ricoprimento $\mathcal{R}_n$ di $\overline{V}$, appartenente a $\rho(V,W)$, i cui elementi incontrino $\overline{V}$ e abbiano diametro minore di $1/n$. Denotiamo inoltre con $\mathcal{R}_n^*$ l'insieme costituito dagli elementi U di $\mathcal{R}_n$ tali che esistano due elementi T_0, T_1 di $\mathcal{U}$ verificanti le condizioni (9). Scegliamo infine, per ciascun n, un elemento U_n di $\mathcal{R}_n^*$ e un elemento s_n di $U_n \cap \overline{V}$. Poiché l'insieme $\overline{V}$ è compatto, esiste un suo elemento t che sia valore di aderenza per la successione $(s_n)_{n\geq 1}$. Allora ogni intorno di t contiene uno degli insiemi U_n e quindi contiene anche due elementi T_0, T_1 di $\mathcal{U}$ verificanti le condizioni

$$T_0 \cup T_1 \subset W, \qquad \overline{T_0} \cap \overline{T_1} = \emptyset, \qquad M_{T_0}(\omega) = M_{T_1}(\omega) = M_W(\omega). \tag{10}$$

Ciò prova che t è un punto di massimo locale non stretto per la traiettoria $X(\,\cdot\,,\omega)$. Ne segue che che ω è un elemento di A, e ciò conclude la dimostrazione. □