1.13 Speranza condizionale: caso generale e legami con legge condizionale

È opportuno, ora che abbiamo gli strumenti per farlo, dare la definizione generale di speranza condizionale ed una costruzione per una generica σ-algebra $\mathcal{B} \subseteq \mathcal{A}$.

Supponiamo che sullo spazio $(\Omega, \mathcal{A}, \mathbb{P})$ ci sia la σ-algebra $\mathcal{B}$, con $\mathcal{B} \subset \mathcal{A}$, e di avere una variabile aleatoria reale X che sia $\mathcal{A}$-misurabile. Come già visto, un possibile modo di approssimare X con una variabile aleatoria Y che sia solo $\mathcal{B}$-misurabile è quello di imporre che valga la proprietà (1.2), che per comodità del lettore ripetiamo qui:

$$E[Y\mathbf{1}_A] = \int_A Y \, d\mathbb{P} = \int_A X \, d\mathbb{P} = E[X\mathbf{1}_A] \qquad \forall A \in \mathcal{B} \tag{1.6}$$

Definizione 1.73 *Una variabile aleatoria Y che gode della proprietà (1.6), viene chiamata* **speranza condizionale** *(o* **valore atteso condizionato***) di X rispetto a $\mathcal{B}$, e si indica con $Y = E[X|\mathcal{B}]$.*

La relazione (1.2) è simile alla definizione di misura definita da una densità. Non è quindi una sorpresa che i risultati di esistenza della speranza condizionale siano legati alla teoria riguardante misure definite da una densità. In effetti, grazie al teorema di Radon-Nykodim possiamo dimostrare che la Y esiste ed è unica nei due casi $X \in L^1$ e $X \in L^+$.

Teorema 1.74 *Se $X \in L^1(\mathbb{P})$ (risp. $X \in L^+(\mathbb{P})$), allora esiste un'unica $Y = E[X|\mathcal{B}] \in L^1(\mathbb{P})$ (risp. L^+).*

Dimostrazione. Partiamo dal caso $X \in L^1$, e supponiamo inizialmente che $X \geq 0$. Consideriamo la restrizione $\bar{\mathbb{P}} = \mathbb{P}|_{\mathcal{B}}$ e definiamo la misura $\mathbb{Q}$ tramite $\frac{d\mathbb{Q}}{d\mathbb{P}} = X$. Allora ovviamente $\mathbb{Q} << \mathbb{P}$. Se ora prendiamo la restrizione $\bar{\mathbb{Q}} = \mathbb{Q}|_{\mathcal{B}}$, abbiamo ovviamente che $\bar{\mathbb{Q}} << \bar{\mathbb{P}}$. Per il teorema di Radon-Nykodim, esiste allora un'unica $Y \in L^+(\Omega, \mathcal{B}, \bar{\mathbb{P}})$ tale che $Y = \frac{d\bar{\mathbb{Q}}}{d\bar{\mathbb{P}}}$. Questo significa che per ogni $A \in \mathcal{B}$ si ha:

$$\int_A Y \, d\mathbb{P} = \int_A Y \, d\bar{\mathbb{P}} = \bar{\mathbb{Q}}(A) = \mathbb{Q}(A) = \int_A X \, d\mathbb{P}$$

e possiamo quindi definire $E[X|\mathcal{B}]$ uguale a questa Y trovata.

Se X non è non negativa, si può decomporre in $X = X^+ - X^-$. Possiamo allora definire $E[X|\mathcal{B}] = E[X^+|\mathcal{B}] - E[X^-|\mathcal{B}]$.

Vediamo una traccia di cosa succede se $X \in L^+$ e non è integrabile. Se $X \in L^+$, esiste sicuramente una successione $(X_n)_n \subseteq L^1$ tale che $X_n \nearrow X$ (ad esempio possiamo prendere le $(X_n)_n$ semplici). Possiamo allora definire $E[X|\mathcal{B}] = \lim_{n\to\infty} E[X_n|\mathcal{B}]$. Ovviamente bisogna dimostrare che questo limite esiste ed è indipendente dalla successione, ma questi risultati seguono dalle proprietà della speranza condizionale. □

Supponiamo ora di avere due variabili aleatorie X e Y, a valori rispettivamente negli spazi $(E, \mathcal{E})$ ed $(F, \mathcal{F})$, e che $(\nu_x)_{x\in D}$ sia una famiglia di misure di probabilità su $(F, \mathcal{F})$, con $D \in \mathcal{E}$ tale che $\mathbb{P}\{X \in D\} = 1$.

Teorema 1.75 *Con le ipotesi sopra, sono equivalenti:*

a) $E[h(Y)|X](\omega) = \int_F h(y)\ d\nu_{X(\omega)}(y)$ *q.c. per ogni* $h \in L^+(F)$*;*

b) $(\nu_x)_{x\in D}$ *è una versione della legge condizionale di* Y *rispetto a* X*.*

Dimostrazione. L'affermazione a) è equivalente a dire che per ogni $A \in \mathcal{E}$,

$$E[\mathbf{1}_{\{X\in A\}}h(Y)] = E\left[\mathbf{1}_{\{X\in A\}}\int_F h(y)\ d\nu_X(y)\right] = \int_D\int_F \mathbf{1}_A(x)h(y)\ d\nu_x(y)\ d\mu(x)$$

Questo è equivalente a dire che

$$\mathbb{P}\{X \in A, Y \in B\} = \int_D\int_F \mathbf{1}_A(x)\mathbf{1}_B(y)\ d\nu_x(y)\ d\mu(x) = \int_A \nu_x(B)\ d\mu(x)$$

per ogni $A \in \mathcal{E}, B \in \mathcal{F}$; ma questa è una caratterizzazione della legge condizionale. □

A volte si scrive anche $E[h(Y)|X] = \varphi(X)$, dove φ è la funzione definita da $\varphi(x) := \int_F h(y)\ d\nu_x(y)$. Per questo motivo, si può anche trovare la notazione "impropria" $\varphi(x) := E[h(Y)|X = x]$. Questa notazione è pienamente giustificata se $\mathbb{P}\{X = x\} > 0$ e se si interpreta $E[\ \cdot\ |X = x]$ come la speranza rispetto alla probabilità $\mathbb{Q}_x := \mathbb{P}\{\cdot|X = x\}$. Se tuttavia $\mathbb{P}\{X = x\} = 0$, questa interpretazione si perde e bisogna ricorrere al significato attribuito dal precedente teorema.

1.14 Funzione caratteristica

Se μ è una misura di probabilità su $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, definiamo la sua **funzione caratteristica** $\varphi_\mu : \mathbb{R} \to \mathbb{C}$ come

$$\varphi_\mu(t) := \int_{\mathbb{R}} e^{itx}\ d\mu(x) \qquad \forall t \in \mathbb{R}$$

Se poi X è una variabile aleatoria reale di legge μ, la **funzione caratteristica** di X è data da

$$\varphi_X(t) = \varphi_\mu(t) := E[e^{itX}] \qquad \forall t \in \mathbb{R}$$

La funzione caratteristica, come ci si può aspettare dal nome, caratterizza la legge di una variabile aleatoria. Ciò è conseguenza del seguente risultato.

Teorema 1.76 (formula di inversione) *Se* $a, b \in \mathbb{R}$ *tali che* $a < b$ *e* $\mu(\{a\}) = \mu(\{b\}) = 0$*, allora*

$$\mu((a, b]) = \lim_{T\to+\infty} \frac{1}{2\pi}\int_{-T}^{T} \frac{e^{-ita} - e^{-itb}}{it}\varphi_\mu(t)\ dt$$

Nella dimostrazione daremo per noto il limite notevole

$$\lim_{T\to+\infty}\int_0^T \frac{\sin\theta t}{t}\ dt = \lim_{T\to+\infty}\int_0^{\theta T} \frac{\sin u}{u}\ du = \frac{\pi}{2}\operatorname{sgn}\theta \tag{1.7}$$

Notiamo che, sia nell'equazione precedente che nell'enunciato del teorema, l'integrale non può essere esteso a $T = +\infty$ secondo la teoria di Lebesgue, poichè le funzioni integrande non sono integrabili secondo Lebesgue: bisogna quindi ricorrere all'enunciato con il limite per $T \to +\infty$.

Dimostrazione. Innanzitutto abbiamo che

$$\int_{-T}^{T} \frac{e^{-ita} - e^{-itb}}{it} \varphi_\mu(t)\ dt = \int_{-T}^{T} \frac{e^{-ita} - e^{-itb}}{it} \int_{\mathbb{R}} e^{itx}\ d\mu(x)\ dt$$

Per semplificare questo integrale si può applicare il teorema di Fubini. A questo proposito notiamo che per ogni $x < y$ si ha

$$|\cos y - \cos x| \leq |y - x||\sin \xi| \leq |y - x|$$

per un opportuno $\xi \in (x, y)$, e allo stesso modo $|\sin y - \sin x| \leq |y - x|$. Allora

$$\begin{aligned} \left| \frac{e^{-ita} - e^{-itb}}{it} \cdot e^{itx} \right| &= \left| \frac{e^{-ita} - e^{-itb}}{it} \right| \leq \left| \frac{\cos(-ta) - \cos(-tb) + i\sin(-ta) - i\sin(-tb)}{it} \right| \leq \\ &\leq \left| \frac{t|b-a| + it|b-a|}{it} \right| \leq 2|b-a| \end{aligned}$$

Siccome $|b - a| \in L^1([-T, T] \times \mathbb{R}, \mathcal{B}([-T, T] \times \mathbb{R}), \lambda_1 \otimes \mu)$, possiamo applicare il teorema di Fubini ottenendo

$$\int_{-T}^{T} \frac{e^{-ita} - e^{-itb}}{it} \varphi_\mu(t)\ dt = \int_{\mathbb{R}} \psi_{a,b}^T(x)\ d\mu(x)$$

dove

$$\psi_{a,b}^T(x) := \int_{-T}^{T} \frac{e^{it(x-a)} - e^{it(x-b)}}{it}\ dt = 2\int_0^T \frac{\sin t(x-a) - \sin t(x-b)}{t}\ dt$$

Notiamo ora che per ogni $x \in \mathbb{R}$ si ha

$$\lim_{T\to+\infty} \psi_{a,b}^T(x) = \psi_{a,b}(x) := \pi(\text{sgn}\ (x-a) - \text{sgn}\ (x-b)) = 2\pi \begin{cases} 0 & \text{se } x < a, \\ \frac{1}{2} & \text{se } x = a, \\ 1 & \text{se } a < x < b, \\ \frac{1}{2} & \text{se } x = b, \\ 0 & \text{se } x > b. \end{cases}$$

Siccome $\mu(\{a, b\}) = 0$, si ha che $\psi_{a,b} = 2\pi \mathbf{1}_{(a,b]}$ μ-q.c.; inoltre, dall'Equazione (1.7) segue che per ogni θ la funzione $T \to \int_0^{\theta T} \frac{\sin u}{u}\ du$ è limitata, quindi esiste una costante $M > 0$ tale che $|\psi_{a,b}^T(x)| \leq M$ per ogni T, x; allora la famiglia $(\psi_{a,b}^T)_T$ è dominata in L^1, quindi

$$\begin{aligned} \lim_{T\to+\infty} \frac{1}{2\pi} \int_{-T}^{T} \frac{e^{-ita} - e^{-itb}}{it} \varphi_\mu(t)\ dt &= \lim_{T\to+\infty} \frac{1}{2\pi} \int_{\mathbb{R}} \psi_{a,b}^T(x)\ d\mu(x) = \\ &= \frac{1}{2\pi} \int_{\mathbb{R}} 2\pi \mathbf{1}_{(a,b]}\ d\mu(x) = \mu((a, b]) \end{aligned}$$

□

Corollario 1.77 *Se μ e μ' sono due probabilità su $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ tali che $\varphi_\mu = \varphi_{\mu'}$, allora $\mu = \mu'$ su $\mathcal{B}(\mathbb{R})$.*

Dimostrazione. Definiamo

$$D := \{x \in \mathbb{R} \mid \mu(\{x\}) = \mu'(\{x\}) = 0\} \qquad \text{ed} \qquad \mathcal{I} := \{(a, b] \mid a, b \in D\}$$

Per il precedente teorema, l'ipotesi implica che $\mu = \mu'$ su $\mathcal{I}$. Siccome D è denso in $\mathbb{R}$, si ha che $\mathcal{I}$ è una base di $\mathcal{B}(\mathbb{R})$, e quindi $\mu = \mu'$ su $\sigma(\mathcal{I}) = \mathcal{B}(\mathbb{R})$. □

Vediamo ora altre proprietà della funzione caratteristica.

Proposizione 1.78 *Sia X una variabile aleatoria reale e φ_X la sua funzione caratteristica. Allora valgono le seguenti:*

1. *$\varphi_X \in C^0$ e $\varphi_X(0) = 1$;*
2. *per ogni $t \in \mathbb{R}$ si ha $|\varphi_X(t)| \leq 1$ e $\varphi_X(-t) = \overline{\varphi_X(t)}$;*
3. *per ogni $c \in \mathbb{R}$, $\varphi_{cX}(t) = \varphi_X(ct)$;*
4. *φ_X è una funzione reale se e solo se X ha legge simmetrica;*
5. *se $X \in L^1(\Omega, \mathcal{A}, \mathbb{P})$, allora $\varphi_X \in C^1$ e $\varphi'_X(t) = iE[Xe^{itX}]$;*
6. *se $X \in L^2(\Omega, \mathcal{A}, \mathbb{P})$, allora $\varphi_X \in C^2$ e $\varphi'_X(t) = -E[X^2 e^{itX}]$.*

Dimostrazione.

1. per ogni $t \in \mathbb{R}$ si ha che $\lim_{u \to t} e^{iuX} = e^{itX}$ $\mathbb{P}$-quasi certamente; inoltre $|e^{iuX}| \leq 1 \in L^1(\Omega, \mathcal{A}, \mathbb{P})$, quindi applicando il teorema della convergenza dominata si ha

   $$\lim_{u \to t} \varphi_X(u) = \lim_{u \to t} E[e^{iuX}] = E\left[\lim_{u \to t} e^{iuX}\right] = E[e^{itX}] = \varphi_X(t)$$

   Inoltre $\varphi_X(0) = \mathcal{E}[e^0] = 1$.

2. Per ogni $t \in \mathbb{R}$ si ha

   $$|\varphi_X(t)| = |E[e^{itX}]| \leq E[|e^{itX}|] = E[1] = 1$$

   e

   $$\varphi_X(-t) = E[e^{-itX}] = E\left[\overline{e^{itX}}\right] = \overline{\varphi_X(t)}$$

3. per ogni $c, t \in \mathbb{R}$ si ha

   $$\varphi_{cX}(t) = E[e^{itcX}] = \varphi_X(ct)$$

4. φ_X è una funzione reale se e solo se $\varphi_X(t) = \overline{\varphi_X(t)}$; ma per il punto 2., questo significa che $\varphi_X(t) = \varphi_X(-t)$; il secondo membro, per il punto 3., è uguale a $\varphi_{-X}(t)$, quindi φ_X è una funzione reale se e solo se $\varphi_X(t) = \varphi_{-X}(t)$; siccome la funzione caratteristica caratterizza la legge, si ha che φ_X è una funzione reale se e solo se X e $-X$ sono identicamente distribuite.

5. In modo analogo al punto 1., si può dimostrare che se $X \in L^1$ è possibile derivare sotto il segno di integrale, e quindi si ha

$$\varphi'_X(t) = E\left[\frac{d}{dt}e^{itX}\right] = iE[Xe^{itX}]$$

6. In modo analogo al punto precedente, si può dimostrare che se $X \in L^2$, allora è possibile derivare sotto il segno di integrale, e quindi si ha

$$\varphi''_X(t) = E\left[\frac{d^2}{dt^2}e^{itX}\right] = -E[X^2e^{itX}]$$

□

Esiste poi un'importante proprietà che rende molto utile lo strumento della funzione caratteristica nel caso di somme di variabili aleatorie indipendenti.

Lemma 1.79 *Se X e Y sono variabili aleatorie reali indipendenti, allora $\varphi_{X+Y} = \varphi_X\varphi_Y$.*

Dimostrazione. Per ogni $t \in \mathbb{R}$ si ha che

$$\varphi_{X+Y}(t) = E[e^{it(X+Y)}] = E[e^{itX}e^{itY}] = E[e^{itX}]E[e^{itY}] = \varphi_X(t)\varphi_Y(t)$$

□

Il calcolo delle funzioni caratteristiche si può effettuare ad esempio tramite i seguenti metodi:

- calcolo diretto;
- metodo dell'equazione differenziale;
- metodo dei residui;
- metodo della forma differenziale.

Diamo un esempio del secondo metodo nel calcolo della funzione caratteristica di una normale.

Esempio 1.80 Se $X \sim N(0,1)$, allora X ha legge simmetrica, e quindi φ_X è una funzione reale. Siccome poi $X \in L^1$, abbiamo che $\varphi_X \in C^1$, e anche la sua derivata è una funzione reale; si ha quindi che

$$\begin{aligned}\varphi'_X(t) &= iE[Xe^{itX}] = -\int_{\mathbb{R}} \frac{\sin tx}{\sqrt{2\pi}} xe^{-\frac{1}{2}x^2}\, dx = \left[\frac{\sin tx}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}\right]_{-\infty}^{+\infty} - \int_{\mathbb{R}} t\frac{\cos tx}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}\, dx = \\ &= -tE[e^{itX}] = -t\varphi_X(t)\end{aligned}$$

e quindi φ_X soddisfa un'equazione differenziale lineare, con condizione iniziale $\varphi_X(0) = 1$; l'equazione ha come unica soluzione $\varphi_X(t) = e^{-\frac{1}{2}t^2}$.

Esempio 1.81 Se $X \sim N(m, \sigma^2)$, allora sappiamo che $X = \sigma Y + m$, con $Y := \frac{X-m}{\sigma} \sim N(0,1)$. Allora per ogni $t \in \mathbb{R}$ si ha

$$\varphi_X(t) = \varphi_{\sigma Y+m}(t) = E[e^{it(\sigma Y+m)}] = e^{imt}E[e^{it\sigma Y}] = e^{imt}\varphi_Y(\sigma t) = \exp\left(imt - \frac{1}{2}\sigma^2 t^2\right)$$

1.15 Spazi L^p, varianza e covarianza

Definizione 1.82 *Per ogni $p \geq 1$ definiamo $L^p = L^p(\Omega, \mathcal{A}, \mathbb{P})$ l'insieme delle classi di equivalenza delle variabili aleatorie su Ω con momento p-esimo integrabile rispetto a $\mathbb{P}$ (cioè tali che $E[|X|^p]$), dove $X \equiv Y$ se e solo se $X = Y$ $\mathbb{P}$-q.o.*

Questi spazi sono spazi vettoriali reali, e sono anche spazi di Banach rispetto alle norme $\|X-Y\|_{L^p} = (E[|X-Y|^p])^{1/p}$. Inoltre, L^2 è uno spazio di Hilbert rispetto al prodotto scalare $\langle X, Y\rangle_{L^2} = E[XY]$. Pur essendo gli L^p spazi definiti come classi di equivalenza di variabili aleatorie, come in analisi si usa la notazione $X \in L^1$ per indicare che X è una variabile aleatoria con momento p-esimo integrabile definita q.o. Questo leggero abuso di notazione é universalmente accettato e non causa quasi mai ambiguità.

Si può facilmente dimostrare che gli spazi L^p sono contenuti uno nell'altro, grazie alla seguente disuguaglianza.

Proposizione 1.83 (Disuguaglianza di Jensen) *Sia X una variabile aleatoria reale, $\varphi : \mathbb{R} \to \mathbb{R}$ una funzione convessa e $\mathcal{B}$ una σ-algebra contenuta in $\mathcal{A}$. Si assuma che le variabili aleatorie X e $\varphi(X)$ ammettano valor medio. Allora*

$$\varphi(E[X|\mathcal{B}]) \leq E[\varphi(X)|\mathcal{B}].$$

Dimostrazione. Il fatto che φ sia convessa è equivalente ad affermare che, per ogni $x_0 \in \mathbb{R}$, esiste $l(x_0) \in \mathbb{R}$ tale che per ogni $x \in \mathbb{R}$

$$\varphi(x) \geq \varphi(x_0) + l(x_0)(x - x_0) \tag{1.8}$$

(il lettore verifichi quest'ultima affermazione). Posto $x = X$ e $x_0 = E[X|\mathcal{B}]$ in (1.8) otteniamo quasi ovunque

$$\varphi(X) \geq \varphi(E[X|\mathcal{B}]) + l(E[X|\mathcal{B}])(X - E[X|\mathcal{B}]). \tag{1.9}$$

Prendendo la speranza condizionale rispetto a $\mathcal{B}$ dei due membri di (1.9), la conclusione segue subito. □

Corollario 1.84 *Sia $X \in L^p$ e sia $0 < q \leq p$. Allora $\|X\|_{L^q} \leq \|X\|_{L^p}$.*

Dimostrazione. È sufficiente applicare la disuguaglianza di Jensen alla variabile aleatoria $|X|^q$ e alla funzione convessa $\varphi(x) = |x|^{p/q}$. □

In particolare, quindi, tutte le variabili aleatorie di ogni spazio L^p, $p \geq 1$, sono dotate di media.

La quantità $E[|X|^p] = \|X\|_{L^p}^p$ si chiama **momento di ordine** p della variabile aleatoria X, ed è finito se e solo se $X \in L^p$. Poichè le costanti appartengono a L^p per ogni $p \geq 1$, X ammette momento di ordine p se e solo se $X - c$ ammette momento di ordine p per ogni $c \in \mathbb{R}$. Se X ammette momento di ordine $p \geq 1$, allora ha senso considerare il momento di ordine p di $X - E[X]$, cioè $E[(X - E[X])^p]$, che viene detto **momento centrato di ordine** p. Si vede subito che il momento centrato di ordine 1 vale zero. Il momento centrato di ordine 2 si dice **varianza**, ed è denotato con Var $[\cdot]$:

$$\text{Var}\,[X] = E[(X - E[X])^2].$$

Si osservi che la varianza non è un operatore lineare. Infatti, per $a, b \in \mathbb{R}$,

$$\text{Var}\,[aX] = a^2 \text{Var}\,[X],$$

e

$$\text{Var}\,[X + b] = \text{Var}\,[X].$$

La verifica di tali identità è semplice, ed è lasciata al lettore.

La varianza si può interpretare come una misura della "dispersione" dei valori della variabile aleatoria X attorno alla sua media. In particolare valgono i seguenti risultati.

Proposizione 1.85 *Sia $X \in L^2(\Omega, \mathcal{A}, \mathbb{P})$.*

i) Var $[X] = 0$ *se e solo se $X = c$ q.c. per qualche costante $c \in \mathbb{R}$.*

ii) Per ogni $c \in \mathbb{R}$, Var $[X] \leq E[|X - c|^2] = \|X - c\|_{L^2}^2$*, e vale l'uguaglianza se e solo se* $c =$ Var $[X]$.

Dimostrazione.

i) Supponiamo Var $[X] = 0$. Questo significa che $E[|X - E[X]|^2] = 0$, e quindi, per la Proposizione 1.34, $|X - E[X]|^2 = 0$ q.c., cioè $X = E[X]$ q.c., e abbiamo la tesi con $c = E[X]$. Viceversa, se supponiamo $X = c$ q.c., chiaramente $E[X] = c$, e quindi $|X - E[X]| = 0$ q.c., e dunque Var $[X] = 0$.

ii) Si consideri il polinomio in c

$$p(c) = \|X - c\|_{L^2}^2 = E[(X - c)^2] = c^2 - 2E[X]c + E[X^2].$$

Si vede subito che $p(c)$ ha un unico minimo assoluto per $c = E[X]$, e $p(E[X]) =$ Var $[X]$.

□

Nella precedente proposizione, il punto (i.) afferma che le variabili aleatorie con varianza zero sono costanti a meno di insiemi di probabilità zero, mentre il punto (ii.) dà

una caratterizzazione "variazionale" del valor medio, affermando che esso è la costante che realizza la distanza minima da X.

Come in molti altri settori della matematica, in probabilità le disuguaglianze giocano un ruolo fondamentale. La seconda che vediamo qui di seguito chiarisce ulteriormente il significato della varianza come indice della dispersione dei valori di una variabile aleatoria: la probabilità di una deviazione dalla media di una variabile aleatoria X si può stimare dall'alto con la sua varianza.

Proposizione 1.86 *Sia X una variabile aleatoria reale su $(\Omega, \mathcal{A}, \mathbb{P})$.*

i) **Disuguaglianza di Markov.** *Se $X \in L^1(\Omega, \mathcal{A}, \mathbb{P})$ ed è positiva, allora, per ogni $\varepsilon > 0$,*

$$P\{X \geq \varepsilon\} \leq \frac{E[X]}{\varepsilon}.$$

ii) **Disuguaglianza di Chebichev.** *Se $X \in L^2(\Omega, \mathcal{A}, P)$, allora, per ogni $\varepsilon > 0$,*

$$P\{|X - E[X]| > \varepsilon\} \leq \frac{\text{Var }[X]}{\varepsilon^2}.$$

Dimostrazione.

i) Tenuto conto che $X \geq \varepsilon \mathbf{1}_{\{X \geq \varepsilon\}}$, abbiamo

$$E[X] \geq E[\varepsilon \mathbf{1}_{\{X \geq \varepsilon\}}] = \varepsilon \mathbb{P}\{X \geq \varepsilon\}$$

da cui segue la tesi.

ii) Poichè

$$\mathbb{P}\{|X - E[X]| > \varepsilon\} = \mathbb{P}\{(X - E[X])^2 > \varepsilon^2\},$$

è sufficiente applicare quanto dimostrato in (i.) a $(X - E[X])^2$ e ad ε^2.

□

Proposizione 1.87 (Disuguaglianze di Chebichev unilatere) *Sia $X \in L^2(\Omega, \mathcal{A}, \mathbb{P})$ con media m e varianza σ^2. Allora, per ogni $a > 0$,*

$$\mathbb{P}\{X > m + a\} \leq \frac{\sigma^2}{\sigma^2 + a^2}, \qquad \mathbb{P}\{X < m - a\} \leq \frac{\sigma^2}{\sigma^2 + a^2}$$

Dimostrazione. Supponiamo all'inizio che $m = 0$. Allora per ogni $b > 0$ si ha

$$\mathbb{P}\{X > a\} = \mathbb{P}\{X + b > a + b\} \leq \mathbb{P}\{(X+b)^2 > (a+b)^2\} \leq \frac{E[(X+b)^2]}{(a+b)^2} = \frac{E[X^2] + b^2}{(a+b)^2} = \frac{\sigma^2 + b^2}{(a+b)^2}$$

dove la prima disuguaglianza segue dal fatto che $a + b > 0$, e la seconda dalla disuguaglianza di Markov. Siccome questa disuguaglianza vale per ogni $b > 0$, deve valere anche per il minimo di $f(b) := \frac{\sigma^2 + b^2}{(a+b)^2}$. Derivando questa funzione si ottiene

$$f'(b) = \frac{2(ab - \sigma^2)}{(a+b)^3} \geq 0 \quad \text{ se e solo se } b \geq \frac{\sigma^2}{a}$$

quindi

$$\mathbb{P}\{X > a\} \leq f\left(\frac{\sigma^2}{a}\right) = \frac{\sigma^2}{\sigma^2 + a^2}$$

Se m è un generico numero reale, in generale le variabili aleatorie $X - m$ e $m - X$ hanno media nulla, e quindi applicando quanto appena dimostrato si ha

$$P\{X > m+a\} = P\{X-m > a\} \leq \frac{\sigma^2}{\sigma^2 + a^2}, \qquad P\{X < m-a\} = P\{m-X > a\} \leq \frac{\sigma^2}{\sigma^2 + a^2}$$

□

Ricordiamo infine la disuguaglianza di Cauchy-Schwartz.

Proposizione 1.88 (Disuguaglianza di Cauchy-Schwartz) *Siano* $X, Y \in L^2(\Omega, \mathcal{A}, \mathbb{P})$. *Allora* XY *ammette valor medio, e*

$$|E[XY]| \leq \sqrt{E[X^2]E[Y^2]}. \tag{1.10}$$

L'uguaglianza in (1.10) vale se e solo se esiste $c \in \mathbb{R}$ *tale che* $X = cY$ *q.c.*

Dimostrazione. Innanzitutto, per mostrare che XY ammette valor medio, si osservi che

$$XY = \frac{1}{2}((X + Y)^2 - X^2 - Y^2)$$

Siccome L^2 è uno spazio vettoriale, il secondo membro ammette valor medio, e quindi anche il primo.

Passiamo ora a dimostrare la disuguaglianza. Non è restrittivo supporre $P\{X = 0\} < 1$, e $P\{Y = 0\} < 1$, il che è equivalente a $E[X^2] > 0$, $E[Y^2] > 0$ (infatti in caso contrario la (1.10) è banalmente verificata). Poniamo allora

$$X_* = \frac{X}{\sqrt{E[X^2]}}, \quad Y_* = \frac{Y}{\sqrt{E[Y^2]}}.$$

Si noti che

$$0 \quad \leq \quad E[(X_* \pm Y_*)^2] = E[X_*^2] + E[Y_*^2] \pm 2E[X_*Y_*] = 2 \pm 2\frac{E[XY]}{\sqrt{E[X^2]E[Y^2]}},$$

da cui (1.10) segue.

Supponiamo ora che (1.10) valga come uguaglianza. Allora, per quanto appena visto, $E[(X_* - Y_*)^2] = 0$ che, per la Proposizione 1.35, equivale a $X_* = Y_*$ q.c., cioè a $X = cY$ q.c. con $c = \frac{\sqrt{E[X^2]}}{\sqrt{E[Y^2]}}$.

Viceversa, se $X = cY$ q.c. per qualche $c \in \mathbb{R}$, allora $E[XY] = cE[Y^2]$, $E[X^2] = c^2E[Y^2]$, da cui si verifica che la (1.10) vale come uguaglianza. □

Una applicazione immediata è la seguente. Siano X, Y due variabili aleatorie reali definite su $(\Omega, \mathcal{A}, \mathbb{P})$, che ammettono valor medio. Se la variabile aleatoria $(X - E[X])(Y - E[Y])$ ammette media, la quantità

$$\text{Cov}\,(X, Y) = E[(X - E[X])(Y - E[Y])]$$

si dice **covarianza** tra X e Y.

Si noti che ogni qual volta X, Y e XY ammettono valor medio, allora la covarianza è ben definita, e

$$\text{Cov}\,(X, Y) = E[XY] - E[X]E[Y].$$

Proposizione 1.89 *Siano $X, Y \in L^2(\Omega, \mathcal{A}, \mathbb{P})$. Allora* Cov (X, Y) *è ben definita. Inoltre*

$$|\text{Cov}\,(X, Y)| \leq \sqrt{\text{Var}\,[X]\text{Var}\,[Y]}. \tag{1.11}$$

Infine, assumendo Var $[X] > 0$, Var $[Y] > 0$, *si ha che* Cov $(X, Y) = \sqrt{\text{Var}\,[X]\text{Var}\,[Y]}$ *(risp.* Cov $(X, Y) = -\sqrt{\text{Var}\,[X]\text{Var}\,[Y]}$*) se e solo se esistono costanti $a > 0$ (risp. $a < 0$) e $b \in \mathbb{R}$ tali che $X = aY + b$ q.c.*

Dimostrazione. Si applica la Proposizione 1.88 alle variabili aleatorie $\bar{X} = X - E[X]$ e $\bar{Y} = Y - E[Y]$. Si ottiene immediatamente l'esistenza della covarianza, e la disuguaglianza (1.11) segue da (1.10). Inoltre, se (1.11) vale come uguaglianza, allora (1.10) vale come uguaglianza per $\bar{X}$ e $\bar{Y}$, e quindi esiste $c \in \mathbb{R}$ per cui $P\{\bar{X} = c\bar{Y}\} = 1$. Notare che dev'essere $c \neq 0$, essendo $E[\bar{X}^2] = \text{Var}\,[X] > 0$. Ma allora $X = cY + E[X] - cE[Y]$ q.c., cioè la tesi vale con $a = c$ e $b = E[X] - cE[Y]$. Infine, un calcolo diretto simile a quello fatto nella Proposizione 1.88 mostra che se $X = aY + b$ q.c. con $a \neq 0$ allora

$$\text{Cov}\,(X, Y) = \frac{a}{|a|}\sqrt{\text{Var}\,[X]\text{Var}\,[Y]},$$

e la dimostrazione è conclusa. □

Sia ora $X = (X_1, \ldots, X_d)$ un vettore aleatorio d-dimensionale. Se $X_i \in L^1$ per ogni $i = 1, \ldots, d$, si definisce **vettore delle medie** il vettore $m = (m_1, \ldots, m_d)$ dove $m_i := E[X_i]$. Se $X_i \in L^2$ per ogni $i = 1, \ldots, d$, si definisce **matrice delle covarianze** (o anche **matrice di varianza-covarianza**) la matrice $\Sigma = (\sigma_{ij})_{i,j=1,\ldots,d}$ dove $\sigma_{ij} := \text{Cov}\,(X_i, X_j)$. È facile vedere che $m \in \mathbb{R}^d$, ed inoltre per ogni $m \in \mathbb{R}^d$ può esistere un vettore aleatorio avente m come vettore delle medie. Sulla matrice delle covarianze, invece, ci sono alcune restrizioni.

Lemma 1.90 *Se $X_i \in L^2$ per ogni $i = 1, \ldots, d$ con vettore delle medie m e matrice delle covarianze Σ, allora Σ è simmetrica e semidefinita positiva, cioè per ogni $y \in \mathbb{R}^d$ si ha $\langle \Sigma y, y \rangle \geq 0$. Inoltre, $\langle \Sigma y, y \rangle = 0$ se e solo se $\langle y, X \rangle$ è quasi certamente uguale a una costante.*

Dimostrazione. Ovviamente $\sigma_{ij} = \text{Cov}\ (X_i, X_j) = \text{Cov}\ (X_j, X_i) = \sigma_{ji}$. Siccome poi $X_i \in L^2$ per ogni $i = 1, \ldots, d$, allora per ogni $y \in \mathbb{R}^d$ si ha che $\langle y, X \rangle \in L^2$, e quindi

$$\begin{aligned} 0 \ &\leq \ \text{Var}\ [\langle y, X \rangle] = \text{Var}\ \left[\sum_{i=1}^{d} y_i X_i\right] = \sum_{i=1}^{d} y_i^2 \text{Var}\ [X_i] + \sum_{j \neq i} y_i y_j \text{Cov}\ (X_i, X_j) = \\ &= \ \sum_{i=1}^{d} y_i^2 \sigma_{ii} + \sum_{j \neq i} y_i y_j \sigma_{ij} = \langle \Sigma y, y \rangle \end{aligned}$$

Infine, $\langle \Sigma y, y \rangle = \text{Var}\ [\langle y, X \rangle] = 0$ se e solo se $\langle y, X \rangle$ è quasi certamente uguale a una costante. □

Capitolo 2

Successioni di variabili aleatorie

2.1 Convergenze di variabili aleatorie

Sia $(X_n)_n$ una successione di variabili aleatorie reali. Diciamo che le $(X_n)_n$ convergono a X:

- **quasi certamente** (e scriviamo $X_n \xrightarrow{\text{q.c.}} X$) se $\mathbb{P}\{\lim_{n\to\infty} X_n = X\} = 1$;
- **in probabilità** (e scriviamo $X_n \xrightarrow{\mathbb{P}} X$) se per ogni $\delta > 0$ si ha $\lim_{n\to\infty} \mathbb{P}\{|X_n - X| > \delta\} = 0$;
- **in L^p** con $p \geq 1$ (e scriviamo $X_n \xrightarrow{L^p} X$) se $\lim_{n\to\infty} \|X_n - X\|_p = 0$;
- **in legge** (e scriviamo $X_n \rightharpoonup X$) se $\lim_{n\to\infty} E[g(X_n)] = E[g(X)]$ per ogni $g : \mathbb{R} \to \mathbb{R}$ misurabile, limitata e tale che l'insieme $D_g := \{x | g \text{ è discontinua in } x\}$ sia trascurabile rispetto alla legge di X.

Notiamo che, a differenza delle prime tre, la convergenza in legge è di natura analitica piuttosto che probabilistica: difatti, se $X_n \rightharpoonup X$ e Y ha la stessa legge di X, allora si ha anche $X_n \rightharpoonup Y$. Per questo motivo si può anche dire che $X_n \rightharpoonup \mu$, dove μ è la legge di X; ovviamente si ha allora $\lim_{n\to\infty} E[g(X_n)] = \int_{\mathbb{R}} g \, d\mu$ per ogni g con $\mu(D_g) = 0$.

Esempio 2.1 La legge dei grandi numeri nel suo enunciato più elementare si può rienunciare come: se $(X_n)_n$ sono i.i.d. e contenute in L^2 e definiamo $S_n := \sum_{i=1}^n X_i$, allora $\frac{S_n}{n} \xrightarrow{\mathbb{P}} E[X_1]$.

Esempio 2.2 Il teorema limite centrale nel suo enunciato più elementare si può rienunciare come: se $(X_n)_n$ sono i.i.d. e contenute in L^2 con $m := E[X_n]$, $\sigma^2 = \text{Var}\,[X_n]$ e definiamo $S_n := \sum_{i=1}^n X_i$, allora

$$\frac{S_n - nm}{\sigma\sqrt{n}} \rightharpoonup N(0,1)$$

La definizione di convergenza in legge non è particolarmente operativa. Può essere quindi utile conoscere due criteri equivalenti, uno rispetto alla funzione di ripartizione, uno rispetto alla funzione caratteristica.

Proposizione 2.3 (caratterizzazione della convergenza in legge) *Dette F_n (rispettivamente F) le funzioni di ripartizione delle variabili aleatorie reali X_n (risp. X) F_n^{-1} (risp. F^{-1}) le loro pseudoinverse, sono equivalenti:*

1. *$X_n \rightharpoonup X$;*
2. *$F_n(t) \to F(t)$ per ogni t in cui F sia continua;*
3. *$F_n(t) \to F(t)$ per ogni $t \in D$ con D denso in $\mathbb{R}$;*
4. *$F_n^{-1}(t) \to F^{-1}(t)$ per ogni t in cui F^{-1} sia continua;*

Dimostrazione. Vedi [4, Sezione 15.3]. □

Proposizione 2.4 (teorema di Paul Levy) *Dette φ_n (rispettivamente φ) le funzioni caratteristiche delle variabili aleatorie reali X_n (risp. X), si ha che $X_n \rightharpoonup X$ se e solo se $\varphi_n \to \varphi$ puntualmente.*

Dimostrazione. Vedi [4, Sezione 15.4]. □

Le convergenze sono poi legate tra di loro dalle seguenti implicazioni.

Lemma 2.5 *Se $X_n \xrightarrow{\text{q.c.}} X$, allora $X_n \xrightarrow{\mathbb{P}} X$.*

Dimostrazione. Per ogni $\delta > 0$, le funzioni $(\mathbf{1}_{\{|X_n - X| > \delta\}})_n$ sono dominate dalla costante $1 \in L^1(\Omega, \mathcal{A}, \mathbb{P})$ e tali che $\mathbf{1}_{\{|X_n - X| > \delta\}} \xrightarrow{\text{q.c.}} 0$. Allora, per il teorema della convergenza dominata, si ha

$$\lim_{n\to\infty} \mathbb{P}\{|X_n - X| > \delta\} = \lim_{n\to\infty} E[\mathbf{1}_{\{|X_n - X| > \delta\}}] = E[0] = 0$$

□

Questa implicazione ha anche una inversa parziale.

Lemma 2.6 *Se $X_n \xrightarrow{\mathbb{P}} X$, allora esiste una sottosuccessione $(n_k)_k \subseteq \mathbb{N}$ tale che $X_{n_k} \xrightarrow{\text{q.c.}} X$*

Dimostrazione. Vedi [2, Sezione 2.5]. □

Lemma 2.7 *Se $X_n \xrightarrow{L^p} X$ con $p \geq 1$, allora $X_n \xrightarrow{\mathbb{P}} X$.*

Dimostrazione. Per ogni $n \geq 1$, $\delta > 0$ definiamo l'evento $E_{n,\delta} := \{|X_n - X| > \delta\}$. Allora per la disuguaglianza di Markov si ha

$$\mathbb{P}(E_{n,\delta}) = \mathbb{P}\{|X_n - X|^p > \delta^p\} \leq \frac{E[|X_n - X|^p]}{\delta^p} = \frac{\|X_n - X\|_p^p}{\delta^p}$$

e quindi $\lim_{n\to\infty} \mathbb{P}(E_{n,\delta}) \leq \lim_{n\to\infty} \frac{\|X_n - X\|_p^p}{\delta^p} = 0$. □

Lemma 2.8 *Se $X_n \xrightarrow{\mathbb{P}} X$, allora $X_n \rightharpoonup X$.*

Dimostrazione. Fissiamo t punto di continuità di F_X. Allora per ogni $\delta > 0$ si ha

$$\begin{aligned} F_{X_n}(t) &= \mathbb{P}\{X_n \le t\} = \mathbb{P}(\{X_n \le t, |X - X_n| > \delta\}) + \mathbb{P}(\{X_n \le t, |X - X_n| \le \delta\}) \le \\ &\le \mathbb{P}\{|X - X_n| > \delta\} + \mathbb{P}\{X \le t + \delta\} = F_X(t + \delta) + \mathbb{P}\{|X - X_n| > \delta\} \end{aligned}$$

e anche

$$\begin{aligned} 1 - F_{X_n}(t) &= \mathbb{P}\{X_n > t\} = \mathbb{P}(\{X_n > t, |X - X_n| > \delta\}) + \mathbb{P}(\{X_n > t, |X - X_n| \le \delta\}) \le \\ &\le \mathbb{P}\{|X - X_n| > \delta\} + \mathbb{P}\{X > t - \delta\} = 1 - F_X(t - \delta) + \mathbb{P}\{|X - X_n| > \delta\} \end{aligned}$$

Allora per ogni $n \ge 1$, $\delta > 0$ si ha

$$F_X(t - \delta) - \mathbb{P}\{|X - X_n| > \delta\} \le F_{X_n}(t) \le F_X(t + \delta) + \mathbb{P}\{|X - X_n| > \delta\}$$

Fissiamo ora δ e facciamo il limite per $n \to \infty$, ottenendo

$$F_X(t - \delta) \le \liminf_{n\to\infty} F_{X_n}(t) \le \limsup_{n\to\infty} F_{X_n}(t) \le F_X(t + \delta)$$

Siccome però F_X è continua in t, facendo il limite per $\delta \to 0$ si ottiene che $F_X(t) \le \liminf_{n\to\infty} F_{X_n}(t) \le \limsup_{n\to\infty} F_{X_n}(t) \le F_X(t)$, e quindi che $F_X(t) \le \lim_{n\to\infty} F_{X_n}(t)$. □

Anche questa implicazione è invertibile, ma solo nel caso particolare in cui X è una costante.

Lemma 2.9 *Se $X_n \rightharpoonup c$, con $c \in \mathbb{R}$, allora $X_n \xrightarrow{\mathbb{P}} c$.*

Dimostrazione. Per ogni $\delta > 0$ abbiamo che

$$\begin{aligned} \mathbb{P}\{|X_n - c| > \delta\} &= \mathbb{P}\{X_n - c > \delta\} + \mathbb{P}\{-X_n + c > \delta\} = \\ &= \mathbb{P}\{X_n > c + \delta\} + \mathbb{P}\{X_n < c - \delta\} = 1 - F_{X_n}(c + \delta) + F_{X_n}((c - \delta)^-) \end{aligned}$$

La funzione di ripartizione della variabile aleatoria costante c è $F(t) = \mathbf{1}_{[c,+\infty)}(t)$, ed è discontinua solo per $t = c$, quindi poichè $X_n \rightharpoonup c$, per ogni $t \ne c$ si ha $\lim_{n\to\infty} F_{X_n}(t) = F(t) = \mathbf{1}_{[c,+\infty)}(t)$, e quindi

$$\lim_{n\to\infty} \mathbb{P}\{|X_n - c| > \delta\} = \lim_{n\to\infty} [1 - F_{X_n}(c + \delta) + F_{X_n}((c - \delta)^-)] = 1 - 1 + 0 = 0$$

□

2.2 Teoremi limite per somme di variabili aleatorie i.i.d.

I teoremi limite classici nel calcolo delle probabilità sono quelli che riguardano le somme di variabili aleatorie i.i.d., e i due più noti sono la legge dei grandi numeri e il teorema limite centrale.

Data una successione di variabili aleatorie $(X_n)_n$ contenuta in L^1, diciamo che la successione soddisfa una **legge forte dei grandi numeri** se

$$\frac{X_n - E[X_n]}{n} \xrightarrow{\text{q.c.}} 0$$

e che la successione soddisfa una **legge debole dei grandi numeri** se

$$\frac{X_n - E[X_n]}{n} \xrightarrow{\mathbb{P}} 0$$

Ovviamente la legge forte implica quella debole, ma è tipicamente molto più impegnativa da dimostrare.

Consideriamo ora una successione di variabili aleatorie $(X_n)_n$ contenuta in L^1, e definiamo

$$S_n := \sum_{i=1}^{n} X_i \qquad \text{per ogni } n \geq 1$$

Abbiamo i seguenti tre risultati sulla successione $(S_n)_n$. I primi due sono due diverse leggi forti dei grandi numeri, che valgono in ipotesi più generali del classico caso $(X_n)_n$ i.i.d. e contenute in L^2. Vedremo che in questo caso vale il terzo risultato, cioè il teorema limite centrale.

Teorema 2.10 (legge forte dei grandi numeri di Rajchmann) *Se le $(X_n)_n$ sono contenute in L^2, scorrelate e tali che $\sup_n \text{Var}\,[X_n] \leq c < +\infty$, allora la successione $(S_n)_n$ soddisfa la legge forte dei grandi numeri.*

Dimostrazione. Vedi [4, Sezione 15.8] □

Teorema 2.11 (legge forte dei grandi numeri di Kolmogorov-Khintchine) *Se le $(X_n)_n$ sono i.i.d. e contenute in L^1, allora la successione $(S_n)_n$ soddisfa la legge forte dei grandi numeri.*

Dimostrazione. Vedi [4, Sezione 15.9] □

Teorema 2.12 (teorema limite centrale, versione di Lindeberg-Levy) *Se le $(X_n)_n$ sono i.i.d. e contenute in L^2, e chiamiamo $m := E[X_n]$, $\sigma^2 := \text{Var}\,[X_n]$, allora*

$$\frac{S_n - nm}{\sigma\sqrt{n}} \rightharpoonup N(0,1)$$

Dimostrazione. Vedi [4, Sezione 15.10] □

2.3 Disuguaglianza di Berry-Esseen

Si può in realtà dimostrare un risultato più forte del teorema limite centrale: difatti se, con le stesse notazioni del teorema, chiamiamo F_n la funzione di ripartizione di $\frac{S_n - nm}{\sigma\sqrt{n}}$ e Φ la funzione di ripartizione della legge $N(0,1)$, il teorema limite centrale ci dice che $F_n \to \Phi$ puntualmente. Si può in realtà dimostrare che, posta

$$\rho_n := \sup_{x \in \mathbb{R}} |F_n(x) - \Phi(x)| = \|F_n(x) - \Phi(x)\|_\infty$$

allora si ha che $\rho_n \to 0$, cioè $F_n \to \Phi$ uniformemente.

Nel caso che le $(X_n)_n$ siano anche contenute in L^3, si può dimostrare un risultato più forte: detto infatti $c := E[\|X - m\|^3]$ il momento centrato di ordine 3 delle $(X_n)_n$, si ha che

$$\rho_n \leq A \frac{c}{\sigma^3 \sqrt{n}}$$

dove A è una costante universale (che cioè non dipende dalla particolare legge delle $(X_n)_n$) minore di 1. Questa maggiorazione è nota come **disuguaglianza di Berry-Esseen**.

Una possibile applicazione è vedere come mai, nell'approssimazione normale di variabili binomiali, si usi un criterio del tipo $np(1-p) > 5$.

Esempio 2.13 Consideriamo $(X_n)_n$ i.i.d. di legge $Be(p)$: ovviamente $(X_n)_n \subset L^3$, e quindi le ipotesi della disuguaglianza di Berry-Esseen sono soddisfatte (oltre che ovviamente quelle del teorema limite centrale). Abbiamo allora $m = E[X_n] = p$, $\sigma^2 = \text{Var}\,[X_n] = p(1-p)$, e

$$c = E[\|X - m\|^3] = p(1-p)^3 + (1-p)(0-p)^3 = p(1-p)(p^2 + (1-p)^2)$$

Allora

$$\rho_n \leq A \frac{c}{\sigma^3 \sqrt{n}} \leq \frac{p(1-p)(p^2 + (1-p)^2)}{p^{3/2}(1-p)^{3/2}\sqrt{n}} = \frac{p^2 + (1-p)^2}{\sqrt{np(1-p)}}$$

È facile vedere che $\frac{1}{2} \leq p^2 + (1-p)^2 \leq 1$ per ogni $p \in [0,1]$. Allora, se per un dato $\varepsilon > 0$ si vuole che n sia abbastanza grande da avere $\rho_n < \varepsilon$, basta imporre

$$\frac{p^2 + (1-p)^2}{\sqrt{np(1-p)}} < \varepsilon$$

e quindi

$$np(1-p) > \left(\frac{p^2 + (1-p)^2}{\varepsilon} \right)^2$$

Se ora scegliamo $\varepsilon := 5^{-1/2}$, allora $\left(\frac{p^2 + (1-p)^2}{\varepsilon} \right)^2 \leq 5$, e quindi imporre $np(1-p) > 5$ ci assicura che $\rho_n < 5^{-1/2} (= 0.45)$. Questa in effetti sembra una pessima approssimazione, ma bisogna ricordarsi che A è una costante universale adatta ad una qualunque variabile aleatoria in L^3: siccome le variabili aleatorie di Bernoulli sono addirittura in L^∞, l'approssimazione è in realtà molto migliore di questa.

2.4 Altri possibili comportamenti asintotici

Abbiamo visto che per somme di variabili aleatorie i.i.d. che siano in L^2 valgono legge dei grandi numeri e teorema limite centrale. Il loro spirito è il seguente: se si vuole rinormalizzare la somma centrata di n variabili aleatorie i.i.d. con una funzione di n in modo da avere una convergenza, allora una rinormalizzazione (= divisione) per n dà una convergenza (quasi certa) verso la costante 0, mentre se si vuole un limite verso una legge non banale bisogna rinormalizzare per $\sqrt{n}$, ottenendo come limite in legge una normale standard.

È naturale chiedersi cosa succede se rilassiamo l'ipotesi che le $(X_n)_n$ siano in L^2: abbiamo visto che, finche $(X_n)_n \subseteq L^1$ vale ancora una legge dei grandi numeri, mentre non si può dire niente sul teorema limite centrale. Se invece $(X_n)_n \nsubseteq L^1$, anche la legge dei grandi numeri potrebbe non valere più.

Il seguente esempio è significativo per capire cosa può succedere. In questo esempio consideriamo una legge per le $(X_n)_n$, detta **legge di Cauchy**, per cui $(X_n)_n \nsubseteq L^1$. Si ha allora che per avere un limite non banale non basta rinormalizzare la somma per $\sqrt{n}$, ma occorre rinormalizzarla con n, e come limite in legge non si ottiene una normale, ma ancora una Cauchy.

Esempio 2.14 Diciamo che la variabile aleatoria reale X è una variabile aleatoria **di Cauchy** di parametro $a > 0$, scritto $X \sim Ca(a)$, se ha densità

$$f_X(x) = \frac{a}{\pi(a^2 + x^2)}$$

rispetto alla misura di Lebesgue. È chiaro che, se $X \sim Ca(a)$, allora $X \notin L^1$: difatti

$$E[|X|] = 2\int_0^{+\infty} \frac{ax}{\pi(a^2 + x^2)}\, dx = +\infty$$

Si può dimostrare che la sua funzione caratteristica è $\varphi_X(t) = e^{-a|t|}$.

Consideriamo ora una successione $(X_n)_n$ i.i.d. di legge $Ca(1)$: se come al solito $S_n := \sum_{i=1}^n X_i$, allora la sua funzione caratteristica è

$$\varphi_{S_n}(t) = \prod_{i=1}^n e^{-a|t|} = e^{-na|t|}$$

e quindi $S_n \sim Ca(na)$. Inoltre, $\varphi_{S_n/n}(t) = \varphi_{S_n}(\frac{t}{n}) = e^{-a|t|}$, e quindi $\frac{S_n}{n} \sim Ca(a)$. È allora chiaro che $\frac{S_n}{n} \nrightarrow Ca(a)$, mentre se invece avessimo avuto $(X_n)_n \subseteq L^1$ avremmo avuto $\frac{S_n}{n} \xrightarrow{\mathbb{P}} 0$ e quindi $\frac{S_n}{n} \rightharpoonup 0$.

Se invece consideriamo la successione $(\frac{S_n}{\sqrt{n}})_n$, abbiamo $\varphi_{S_n/\sqrt{n}}(t) = \varphi_{S_n}(\frac{t}{\sqrt{n}}) = e^{-a\sqrt{n}|t|}$, che per $n \to \infty$ converge puntualmente verso la funzione $\mathbf{1}_{\{0\}}$, che non essendo continua non può essere una funzione caratteristica: per il teorema di Paul Levy, questo significa che la successione $(\frac{S_n}{\sqrt{n}})_n$ non converge in legge.

2.5 Introduzione ai processi stocastici

Un **processo stocastico** è una famiglia di variabili aleatorie $X = (X_t)_{t\in I}$, definite sullo spazio probabilizzato $(\Omega, \mathcal{A}, \mathbb{P})$, tutte a valori nello stesso spazio $(E, \mathcal{E})$, dove I è un sottoinsieme di $\mathbb{R}$. L'interpretazione più comune è quella di un fenomeno aleatorio che evolve nel tempo; per ogni $t \in I$, X_t è la variabile di stato che interessa studiare, all'istante di tempo t. Per questa ragione $(E, \mathcal{E})$ viene chiamato **spazio degli stati** e I viene chiamato **insieme dei tempi**. Inoltre, per ogni $\omega \in \Omega$ fissato, la funzione $I \ni t \to X_t(\omega)$ si chiama **traiettoria** del processo X.

Esempio 2.15 (processo di Bernoulli) L'esempio più facile di processo stocastico è questo: poniamo $I = \mathbb{N}^*$, $E = \{0, 1\}$, $\mathcal{E} = \mathcal{P}(E)$, e imponiamo che le $(X_n)_n$ siano indipendenti e identicamente distribuite (ovviamente avranno quindi legge di Bernoulli con il medesimo parametro $p \in (0, 1)$). In questo caso X prende il nome di **processo di Bernoulli** di parametro p.

Esempio 2.16 (processo di Poisson) Consideriamo ora $(X_n)_n$ i.i.d. di legge $Exp(\lambda)$, con $\lambda > 0$, e definiamo il processo $N = (N_t)_{t\geq 0}$ come segue: per ogni $t \geq 0$,

$$N_t := \sum_{n=1}^{\infty} \mathbf{1}_{\{S_n \leq t\}}, \qquad \text{con } S_n := \sum_{i=1}^{n} X_i$$

Il processo N si chiama **processo di Poisson** di parametro λ. In questo caso, quindi, abbiamo $I = \mathbb{R}^+$ e $E = \mathbb{N}$. L'interpretazione più comune è che le $(X_n)_n$ rappresentino dei tempi di attesa, indipendenti tra di loro, tra due avvenimenti successivi. S_n è allora il tempo totale da attendere prima dell'n-esimo avvenimento, e N_t è quindi il numero di avvenimenti accaduti fino all'istante t. A causa di questa interpretazione, i processi di Poisson sono molto utilizzati in quella che si chiama *teoria delle code*. Si può dimostrare che, per ogni $t > 0$, $N_t \sim Po(\lambda t)$ (esercizio), e che le $(N_t)_t$ non sono indipendenti tra di loro.

Consideriamo ora una famiglia di σ-algebre $(\mathcal{F}_t)_{t\in I}$, tutte contenute in $\mathcal{A}$. Essa si dice **filtrazione** se $\mathcal{F}_s \subseteq \mathcal{F}_t$ per ogni $s < t$.

Dati un processo $(X_t)_{t\in I}$ e una filtrazione $(\mathcal{F}_t)_{t\in I}$, diciamo che X è **adattato** a $(\mathcal{F}_t)_t$ se X_t è $\mathcal{F}_t$-misurabile per ogni $t \in I$. Possiamo poi costruire una filtrazione ponendo $\mathcal{F}_t^X := \sigma(X_s | s \in I, s \leq t)$ per ogni $t \in I$: questa filtrazione si dice **filtrazione naturale** di X, ed è la più piccola filtrazione a cui X è adattato. L'interpretazione più comune è che $\mathcal{F}_t$ rappresenta, per ogni istante di tempo t, tutta l'informazione a disposizione fino all'istante t.

Nella teoria dei processi stocastici si fa la grande distinzione tra **processi a tempo discreto**, per cui I è un insieme discreto (finito o infinito), e **processi a tempo continuo**, per cui I è un intervallo reale (limitato o illimitato). Nel seguito considereremo solo processi a tempo discreto, per cui la teoria è molto più semplice. In particolare supporremo sempre $I \subseteq \mathbb{Z}$.

2.6 Tempi di arresto

Data una filtrazione $(\mathcal{F}_i)_{i\in I}$, una variabile aleatoria τ, definita su $I\cup\{+\infty\}$, si dice **tempo di arresto** rispetto a $(\mathcal{F}_i)_i$ se $\{\tau \le i\} \in \mathcal{F}_i$ per ogni $i \in I$. L'interpretazione è che noi possiamo decidere se arrestarci o meno solo sulla base dell'informazione posseduta fino all'istante i.

Come annunciato nella sezione precedente, da ora in poi considereremo solo in caso $I \subseteq \mathbb{Z}$.

Nota 2.17 Se $I \subseteq \mathbb{Z}$, allora una definizione equivalente è la seguente: τ è un tempo di arresto se $\{\tau = n\} \in \mathcal{F}_n$ per ogni $n \in I$. Difatti, se τ è un tempo di arresto rispetto alla definizione originale, allora per ogni $n \in I$, $\{\tau = n\} = \{\tau \le n\} \setminus \{\tau \le n-1\}$; il primo dei due eventi appartiene a $\mathcal{F}_n$, mentre il secondo a $\mathcal{F}_{n-1}$, che è contenuta in $\mathcal{F}_n$, e quindi $\{\tau = n\} \in \mathcal{F}_n$.

Viceversa, supponiamo che τ sia tale che $\{\tau = n\} \in \mathcal{F}_n$ per ogni $n \in I$. Allora, per ogni $n \in I$,

$$\{\tau \le n\} = \bigcup_{k\le n, k\in I} \{\tau = k\}$$

Ognuno di questi eventi appartiene a $\mathcal{F}_k$, e quindi a $\mathcal{F}_n$, e quindi abbiamo un'unione al più numerabile di eventi in $\mathcal{F}_n$, e quindi $\{\tau \le n\} \in \mathcal{F}_n$. Questa definizione equivalente vale ovviamente solo a tempi discreti!

Esempio 2.18 La variabile aleatoria $\tau \equiv k$ è un tempo di arresto per ogni $k \in I$ e per ogni filtrazione $(\mathcal{F}_n)_n$.

Esempio 2.19 Se $(X_n)_{n\in I}$ è un processo a valori in $(E, \mathcal{E})$ adattato a $(\mathcal{F}_n)_n$, con $I \subseteq \mathbb{Z}$, e $\Gamma \in \mathcal{E}$, allora il tempo

$$\tau := \inf\{n \in I \mid X_n \in \Gamma\}$$

è un tempo di arresto, chiamato **tempo di entrata** di X in Γ. Che τ sia un tempo di arresto è evidente: difatti per ogni $n \in I$ si ha

$$\{\tau \le n\}^c = \{\tau > n\} = \left(\bigcap_{m\le n} \{X_m \notin \Gamma\}\right)^c$$

Siccome $\{X_m \notin \Gamma\} \in \mathcal{F}_m \subseteq \mathcal{F}_n$, allora abbiamo $\{\tau \le n\} \in \mathcal{F}_n$.

Un particolare esempio lo abbiamo già visto in una definizione della variabile aleatoria geometrica: in quel caso X è un processo di Bernoulli di parametro $p \in (0,1)$, $\Gamma = \{1\}$ e τ ha legge geometrica di parametro p.

Se nella definizione si sostituisce $\in$ con $\notin$, si parla invece di **tempo di uscita** di X in Γ.

Dato un tempo di arresto τ rispetto a una filtrazione $(\mathcal{F}_n)_n$, definiamo la **σ-algebra del passato di τ** come

$$\mathcal{F}_\tau := \{A \in \mathcal{F}_\infty \mid A \cap \{\tau \le n\} \in \mathcal{F}_n \quad \forall n \in I\}$$

dove $\mathcal{F}_\infty = \sigma(\mathcal{F}_n | n \in I)$ (che a priori potrebbe essere contenuta strettamente in $\mathcal{A}$). In generale, $\mathcal{F}_\tau$ è una σ-algebra che contiene $\sigma(\tau)$ (esercizio). La sua interpretazione piú comune é che appunto, come $\mathcal{F}_n$ rappresenta l'informazione fino al tempo (deterministico) n, $\mathcal{F}_\tau$ rappresenta l'informazione fino al tempo (aleatorio) τ.

Dato un processo $(X_n)_n$ adattato a $(\mathcal{F}_n)_n$ ed un tempo di arresto τ, possiamo definire il **processo arrestato** $X^{|\tau}$ in questo modo: per ogni $n \in I$,

$$X_n^{|\tau}(\omega) := X_{n \wedge \tau(\omega)}(\omega)$$

L'interpretazione è che possiamo far evolvere il fenomeno rappresentato dal processo X fino all'istante aleatorio τ (che, essendo un tempo di arresto, è determinato dall'informazione rappresentata da $(\mathcal{F}_n)_n$): in quel momento, arrestiamo il sistema, che continua ad avere sempre lo stesso valore $X_n^{|\tau}(\omega)$ per ogni $n \geq \tau(\omega)$.

2.7 Martingale

Un processo stocastico $(X_n)_{n \in I}$ adattato ad una filtrazione $(\mathcal{F}_n)_{n \in I}$ si dice **(super-)(sub-)martingala** se $(X_n)_n \subseteq L^1$ e per ogni $n, m \in I$, $n \leq m$, si ha

$$E[X_m \mid \mathcal{F}_n](\leq)(\geq) = X_n$$

Nota 2.20 Se $I \subseteq \mathbb{Z}$, allora la condizione precedente è equivalente a $E[X_{n+1} \mid \mathcal{F}_n](\leq)(\geq) = X_n$ per ogni n tale che $n, n+1 \in I$. Inoltre, se $(X_n)_n$ è una (super-)(sub-)martingala, allora la funzione $n \to E[X_n]$ è costante (decrescente)(crescente).

Nota 2.21 I segni che definiscono le super- e submartingale possono apparire fuorvianti, ma sono conseguenza della seguente interpretazione. Una martingala è la rappresentazione matematica del capitale di un giocatore in un cosiddetto gioco equo, un gioco in cui cioè ad ogni partita la speranza di guadagno futura è nulla, e quindi la miglior previsione futura del capitale del giocatore, in termini di speranza condizionale, è uguale al capitale presente. Se invece il gioco è sbilanciato contro il giocatore (condizione che abbiamo caratterizzato con una supermartingala), questo significa che per avere una somma futura uguale a quella che otterremmo se avessimo una martingala, bisognerà partire con un capitale presente maggiore, e quindi il processo che rappresenta il capitale sarà sempre maggiore di una martingala fino all'istante finale. Chiaramente per una submartingala si applica il ragionamento inverso.

Esempio 2.22 Se $(X_n)_n$ sono i.i.d. tali che $E[X_n] = 0$ per ogni $n \geq 1$ e definiamo $S_n := \sum_{i=1}^n X_i$ e $\mathcal{F}_n := \mathcal{F}_n^X$, allora $(S_n)_n$ è una martingala: difatti per ogni $n \geq 1$ si ha

$$E[S_{n+1}|\mathcal{F}_n] = E[S_n + X_{n+1}|\mathcal{F}_n] = S_n + E[X_{n+1}|\mathcal{F}_n] = S_n + E[X_{n+1}] = S_n$$

per le proprietà della speranza condizionale.

Esempio 2.23 Se $(X_n)_n$ sono i.i.d. tali che $E[X_n] = 1$ per ogni $n \geq 1$ e definiamo $P_n := \prod_{i=1}^n X_i$ e $\mathcal{F}_n := \mathcal{F}_n^X$, allora $(P_n)_n$ è una martingala: difatti per ogni $n \geq 1$ si ha

$$E[P_{n+1}|\mathcal{F}_n] = E[P_n X_{n+1}|\mathcal{F}_n] = P_n E[X_{n+1}|\mathcal{F}_n] = P_n E[X_{n+1}] = P_n$$

per le proprietà della speranza condizionale.

Esempio 2.24 Se $X \in L^1$ e $(\mathcal{F}_n)_n$ è una filtrazione, allora $X_n := E[X|\mathcal{F}_n]$ definisce una martingala.

Esempio 2.25 Se $(X_n)_n$ è una martingala rispetto a $(\mathcal{F}_n)_n$ e $\varphi : \mathbb{R} \to \mathbb{R}$ è una funzione convessa, allora $(\varphi(X_n))_n$ è una submartingala. Difatti per ogni n si ha

$$E[\varphi(X_{n+1})|\mathcal{F}_n] \geq \varphi(E[X_{n+1}|\mathcal{F}_n]) = \varphi(X_n)$$

Se $(X_n)_n$ è più in generale una super- o submartingala, per ottenere lo stesso risultato bisogna imporre che φ sia monotona nel verso opportuno.

Le proprietà delle martingale sono numerose, e ne elencheremo solo alcune. Una delle più significative è la seguente. Abbiamo detto che un'interpretazione comune è quella del capitale di un giocatore che giochi una serie di partite eque. A questo giocatore potrebbe venire in mente che, arrestandosi al momento opportuno (ad esempio quando sta vincendo molto), potrebbe trasformare il suo capitale in qualcosa che non sia una martingala ma una cosa più favorevole a lui: chiaramente è realistico fare questo usando non un tempo aleatorio qualunque, ma un tempo di arresto: se infatti (come al solito) $\mathcal{F}_n$ rappresenta l'informazione fino all'istante n, allora quando siamo al tempo n possiamo decidere se fermarci o continuare solo in base a quel che è successo fino a quell'istante, e non in base ad un futuro che non abbiamo ancora osservato. Detto τ il tempo di arresto in cui il giocatore smette di giocare, modellizziamo dunque il capitale del giocatore con $(X_n^{|\tau})_n$.

Teorema 2.26 (teorema di arresto) *Se $(X_n)_{n\in I}$ è una (super-)(sub-)martingala con $I \subseteq \mathbb{N}$ e τ_1, τ_2 sono tempi di arresto, allora:*

1. *$(X_n^{|\tau_1})_n$ è una (super-)(sub-)martingala;*
2. *se $\tau_1 \leq \tau_2 \leq M$, allora $E[X_{\tau_2}|\mathcal{F}_{\tau_1}](\leq)(\geq) = X_{\tau_1}$.*

Dimostrazione.

1. Per ogni $n \in I$, se $(X_n)_n$ è una martingala abbiamo

$$\begin{aligned} E[X_{n+1}^{|\tau_1}|\mathcal{F}_n] &= E[X_{\tau_1\wedge(n+1)}|\mathcal{F}_n] = E\left[\sum_{i\leq n} \mathbf{1}_{\{\tau_1=i\}} X_i + \mathbf{1}_{\{\tau_1>n\}} X_{n+1}|\mathcal{F}_n\right] = \\ &= \sum_{i\leq n} \mathbf{1}_{\{\tau_1=i\}} X_i + \mathbf{1}_{\{\tau_1>n\}} E[X_{n+1}|\mathcal{F}_n] = \\ &= \sum_{i\leq n} \mathbf{1}_{\{\tau_1=i\}} X_i + \mathbf{1}_{\{\tau_1>n\}} X_n = \sum_{i\leq n-1} \mathbf{1}_{\{\tau_1=i\}} X_i + \mathbf{1}_{\{\tau_1\geq n\}} X_n = X_{\tau_1\wedge n} \end{aligned}$$

   mentre invece, se $(X_n)_n$ è una super- o submartingala, nella terza riga il primo uguale è sostituito dall'opportuna disuguaglianza.

2. Presentiamo la dimostrazione solo per il caso in cui $(X_n)_n$ è una martingala, poichè il caso generale è un po' più complicato. Supponiamo all'inizio che $\tau_2 = M$ q.c. Allora, per ogni $A \in \mathcal{F}_{\tau_1}$ si ha che

$$E[X_{\tau_1}\mathbf{1}_A] = E\left[\sum_{i\leq M} \mathbf{1}_{\{\tau_1=i\}} X_i \mathbf{1}_A\right] = E\left[\sum_{i\leq M} \mathbf{1}_{\{\tau_1=i\}\cap A} X_i\right] = \sum_{i\leq M} E[\mathbf{1}_{\{\tau_1=i\}\cap A} X_i]$$

Siccome $\{\tau_1 = i\} \cap A \in \mathcal{F}_i$ e $(X_n)_n$ è una martingala, si ha $E[\mathbf{1}_{\{\tau_1=i\}\cap A} X_i] = E[\mathbf{1}_{\{\tau_1=i\}\cap A} X_M]$, e quindi

$$E[X_{\tau_1}\mathbf{1}_A] = \sum_{i\leq M} E[\mathbf{1}_{\{\tau_1=i\}\cap A} X_M] = E\left[\sum_{i\leq M} \mathbf{1}_{\{\tau_1=i\}\cap A} X_M\right] = E[\mathbf{1}_A X_M] = E[\mathbf{1}_A X_{\tau_2}]$$

Se in generale $\tau_2 \leq M$, allora poichè $\mathcal{F}_{\tau_1} \subseteq \mathcal{F}_{\tau_2}$, si ha che per ogni $A \in \mathcal{F}_{\tau_1}$

$$E[X_{\tau_1}\mathbf{1}_A] = E[\mathbf{1}_A X_M] = E[\mathbf{1}_A X_{\tau_2}]$$

□

Presentiamo infine, senza dimostrazione, un teorema di convergenza per le (super)martingale.

Teorema 2.27 (teorema di convergenza) *Se $(X_n)_n$ è una supermartingala tale che $\sup_n E[X_n^-] < +\infty$, allora esiste $X_\infty \in L^1$ tale che $X_n \to X_\infty$ q.c.*

Ci sono due casi particolari in cui è immediato applicare il teorema:

1. il caso in cui $(X_n)_n$ è una supermartingala positiva;
2. il caso in cui $(X_n)_n$ è una supermartingala limitata in L^1, cioè tale che $\sup_n E[|X_n|] < +\infty$.

2.8 Catene di Markov

I processi stocastici più semplici da analizzare sono quelli della forma $X = (X_n)_n$ con le $(X_n)_n$ indipendenti: infatti per questi processi qualunque problema che ne coinvolga la legge congiunta può essere ricondotto a calcoli effettuati con le leggi marginali.

Il discorso si fa più complesso quando si analizzano processi $X = (X_n)_n$ in cui le $(X_n)_n$ non sono indipendenti. Tra questi processi, le cosiddette "catene di Markov" hanno la forma più semplice di dipendenza tra le variabili $(X_n)_n$. Difatti, sono processi per cui la legge marginale di X rispetto al suo passato dipende solo dall'istante di tempo immediatamente precedente. Per dare una definizione matematica, dobbiamo tuttavia premettere alcuni concetti.

Dati due spazi misurabili $(E, \mathcal{E})$ ed $(F, \mathcal{F})$, chiamiamo **nucleo di transizione** una applicazione $N : E \times \mathcal{F} \to \mathbb{R}^*$ tale che

- per ogni $x \in E$, $N(x, \cdot)$ è una misura su $(F, \mathcal{F})$;
- per ogni $B \in \mathcal{F}$, $N(\cdot, B)$ è $\mathcal{E}$-misurabile.

Se poi per ogni $x \in E$, $N(x, \cdot)$ è una probabilità su $(F, \mathcal{F})$, allora N si dice nucleo di transizione **markoviano**.

Esempio 2.28 Supponiamo di avere le variabili aleatorie X e Y, a valori rispettivamente su $(E, \mathcal{E})$ ed $(F, \mathcal{F})$, e una versione $(\nu_x)_{x \in D}$ della legge condizionale di Y rispetto a X. Allora possiamo definire

$$N(x, B) := \begin{cases} \nu_x(B) & \text{se } x \in D, \\ \lambda(B) & \text{se } x \notin D, \end{cases}$$

dove λ è una qualunque misura di probabilità su $(F, \mathcal{F})$. Allora N è un nucleo di transizione markoviano. Infatti la condizione di misurabilità segue dal fatto che la funzione $x \to \int_F h(x, y)\, d\nu_x(y)$ deve essere $\mathcal{E}$-misurabile per ogni $h \in L^+(E \times F)$, e quindi in particolare per ogni $h(x, y) = \mathbf{1}_B(y)$, con $B \in \mathcal{F}$.

Se N è un nucleo markoviano e $f \in L^+(F, \mathcal{F})$, possiamo definire $Nf \in L^+(E, \mathcal{E})$ tramite

$$(Nf)(x) := \int_F f(y) N(x, dy) \qquad \forall x \in E$$

Se $(E, \mathcal{E}) = (F, \mathcal{F})$, allora possiamo inoltre definire ricorsivamente $N^0 f := f$, $N^k f := N(N^{k-1} f)$.

Nota 2.29 Se $f = \mathbf{1}_A$, con $A \in \mathcal{F}$, allora $(Nf)(x) = N(x, A)$.

Nota 2.30 Se E ed F sono discreti e $\mathcal{F} = \mathcal{P}(F)$, allora N è caratterizzato dalle quantità $p_{xy} := N(x, \{y\})$. Allora $P := (p_{xy})_{xy}$ si dice **matrice di transizione markoviana**, ed ha un numero finito o infinito di componenti a seconda che E ed F siano finiti o meno; inoltre si ha

$$(Nf)(x) = \sum_{y \in F} f(y) p_{xy} = (Pf)(x)$$

dove l'ultima uguaglianza può essere interpretata nel senso del prodotto tra matrici. Inoltre per ogni $k \geq 0$ si ha che

$$(N^k f)(x) = \sum_{y \in F} f(y) p^k_{xy} = (P^k f)(x)$$

dove $(p^k_{xy})_{xy}$ sono i coefficienti della k-esima potenza P^k della matrice P.

Supponiamo ora di avere un processo stocastico $X = (X_n)_n$ a valori in $(E, \mathcal{E})$ definito sullo spazio probabilizzato $(\Omega, \mathcal{A}, \mathbb{P})$, e di chiamare $(\mathcal{F}_n)_n$ la sua filtrazione naturale. Siano inoltre μ una misura di probabilità su $(E, \mathcal{E})$ e N un nucleo markoviano di $(E, \mathcal{E})$ in sè. Allora X si dice **catena di Markov** di legge iniziale μ e nucleo di transizione N se vale una delle seguenti proprietà equivalenti:

1. a) $\mathbb{P}_{X_0} = \mu$;
   b) per ogni $h \in L^+(E, \mathcal{E})$, $E[h(X_{n+1})|\mathcal{F}_n] = Nh(X_n) \quad \forall n \in \mathbb{N}$;

2. a) $\mathbb{P}_{X_0} = \mu$;
   b) per ogni $B \in \mathcal{E}$, $\mathbb{P}\{X_{n+1} \in B|\mathcal{F}_n\} = N(X_n, B)$ (dove $\mathbb{P}\{X_{n+1} \in B|\mathcal{F}_n\} := E[\mathbf{1}_B(X_{n+1})|\mathcal{F}_n]$);

3. per ogni $B_0, \ldots, B_n \in \mathcal{E}$ si ha che

$$\mathbb{P}\{X_0 \in B_0, \ldots, X_n \in B_n\} = \int_{B_0} \ldots \int_{B_{n-1}} N(x_{n-1}, B_n)\, N(x_{n-2}, dx_{n-1})\, \ldots\, N(x_0, dx_1)\, \mu(dx_0)$$

Vediamo ora che le tre definizioni sono equivalenti.

Dimostrazione.

3) $\Rightarrow$ 2) Il punto 2.*a*) segue subito. Il punto 2.*b*) è equivalente a

$$E[\mathbf{1}_{\{X_{n+1} \in B\}} \mathbf{1}_A] = E[N(X_n, B)\mathbf{1}_A] \qquad \forall A \in \mathcal{F}_n, B \in \mathcal{E}$$

Per dimostrare l'uguaglianza di sopra, basta dimostrarla per ogni A in una opportuna base di generatori di $\mathcal{F}_n$. Consideriamo allora la base costituita dagli eventi della forma $\{X_0 \in B_0, \ldots, X_n \in B_n\}$, con $B_0, \ldots, B_n \in \mathcal{E}$. Per il punto 3) si ha che

$$\begin{aligned} E[\mathbf{1}_{\{X_{n+1} \in B\}} \mathbf{1}_{\{X_0 \in B_0, \ldots, X_n \in B_n\}}] &= \mathbb{P}\{X_0 \in B_0, \ldots, X_n \in B_n, X_{n+1} \in B\} = \\ &= \int_{B_0} \ldots \int_{B_n} N(x_n, B)\, N(x_{n-1}, dx_n)\, \ldots\, N(x_0, dx_1)\, \mu(x_0) = \\ &= E[N(X_n, B)\mathbf{1}_{\{X_0 \in B_0, \ldots, X_n \in B_n\}}] \end{aligned}$$

2) $\Rightarrow$ 1) Il punto 2.*b*) implica che il punto 1.*b*) è vero per ogni h positiva e semplice: difatti sia la speranza condizionale al primo membro che l'operatore integrale N sono lineari rispetto ad h. A questo punto, applicando il teorema di convergenza monotona ad entrambi i membri, si ottiene la tesi.

1) $\Rightarrow$ 3) Per induzione su n: se $n = 0$, allora $\mathbb{P}\{X_0 \in B_0\} = \mu(B_0)$. Se la tesi vale per ogni $B_0, \ldots, B_n \in \mathcal{E}$, allora

$$\begin{aligned} &\mathbb{P}\{X_0 \in B_0, \ldots, X_n \in B_n, X_{n+1} \in B_{n+1}\} = E[\mathbf{1}_{\{X_{n+1} \in B_{n+1}\}} \mathbf{1}_{\{X_0 \in B_0, \ldots, X_n \in B_n\}}] = \\ &= E[N(X_n, B_{n+1})\mathbf{1}_{\{X_0 \in B_0, \ldots, X_n \in B_n\}}] = \int_{B_0 \times \ldots \times B_n} N(x_n, B_{n+1})\, d\mathbb{P}_{X_0, \ldots, X_n}(x_0, \ldots, x_n) = \\ &= \int_{B_0} \ldots \int_{B_n} N(x_n, B_{n+1})\, N(x_{n-1}, dx_n)\, \ldots\, N(x_0, dx_1)\, \mu(x_0) \end{aligned}$$

□

Alcuni autori riservano il nome "catena di Markov" al caso in cui E è un insieme discreto, mentre nel caso in cui E non è discreto il processo $(X_n)_n$ viene chiamato **processo**

di Markov a tempi discreti. Noi ci atterremo alla prima definizione data, chiamando catena di Markov una qualunque successione di variabili aleatorie che goda delle proprietà equivalenti 1)–3), senza distinguere se E sia un insieme discreto o meno. Notiamo però che se E è discreto la proprietà 3) è equivalente alla seguente:

3'. per ogni $x_0, \dots, x_n \in E$ si ha che

$$\mathbb{P}\{X_0 = x_0, \dots, X_n = x_n\} = \mu(x_0) p_{x_0,x_1} \dots p_{x_{n-1},x_n}$$

dove $(p_{xy})_{xy}$ è la matrice di transizione di X.

Inoltre, le 3 definizioni sono equivalenti ad una quarta:

4. a) $\mathbb{P}_{X_0} = \mu$;
   b) per ogni $x_0, \dots, x_{n+1} \in E$, $\mathbb{P}\{X_{n+1} = x_{n+1} | X_0 = x_0, \dots X_n = x_n\} = p_{x_n,x_{n+1}}$

Corollario 2.31 *Se X è una catena di Markov di nucleo N, allora*

1. *$E[h(X_{n+k}) \mid \mathcal{F}_n] = (N^k h)(X_n)$ per ogni $n, k \geq 0$, $h \in L^+(E, \mathcal{E})$.*
2. *Se E è discreto, allora*

$$\mathbb{P}\{X_{n+1} = x_{n+1}, \dots, X_{n+k} = x_{n+k} \mid X_0 = x_0, \dots X_n = x_n\} = p_{x_n,x_{n+1}} \cdot \dots \cdot p_{x_{n+k-1},x_{n+k}}$$

Nota 2.32 Dal corollario appena visto si ricava che, se E è discreto, si ha che $p_{xy} = \mathbb{P}\{X_{n+1} = y \mid X_n = x\}$; inoltre si può interpretare

$$p^k_{xy} = (N^k \mathbf{1}_{\{y\}})(x) = E[\mathbf{1}_{\{y\}}(X_{n+k}) \mid X_n = x] = \mathbb{P}\{X_{n+k} = y \mid X_n = x\}$$

come la probabilità di passare dallo stato x allo stato y in esattamente k passi.

Esempio 2.33 (passeggiata aleatoria su $\mathbb{Z}$) Prendiamo $(X_n)_n$ i.i.d. di legge data da

$$\mathbb{P}\{X_n = 1\} = \mathbb{P}\{X_n = -1\} = \frac{1}{2}$$

e definiamo $S_n := \sum_{i=1}^n X_i$. Intuitivamente S rappresenta il moto di una particella che può occupare solo posizioni su un reticolo intero, e ad ogni istante si muove avanti o indietro con probabilità 1/2. Allora S è una catena di Markov, con spazio degli stati $E = \mathbb{Z}$, legge iniziale $\mu = \delta_0$ e nucleo di transizione $N(x, \cdot) = \frac{1}{2}\delta_1 + \frac{1}{2}\delta_{-1}$. In questo caso la matrice di transizione $P = (p_{xy})_{xy}$ ha infinite componenti, date da

$$p_{xy} = \begin{cases} \frac{1}{2} & \text{se } |x - y| = 1, \\ 0 & \text{altrimenti} \end{cases}$$

Esempio 2.34 (*rating* di obbligazioni) Una obbligazione è un titolo finanziario che obbliga una parte (l'emittente) a pagare una certa somma dopo una scadenza, a condizione che l'emittente non fallisca nel frattempo. Per osservare questo evento prima che accada, ci sono le cosiddette agenzie di *rating*, che assegnano ad ogni obbligazione un *rating*, con l'interpretazione che più è alto il *rating*, più è bassa la probabilità di fallimento dell'obbligazione. In un modello molto semplice, una obbligazione può avere *rating* A, B, C o D (dove D corrisponde al fallimento) e passare da un rating all'altro secondo la matrice di transizione

$$P = \begin{pmatrix} 0.9 & 0.09 & 0.009 & 0.001 \\ 0.05 & 0.9 & 0.04 & 0.01 \\ 0.01 & 0.04 & 0.9 & 0.05 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Notiamo che la probabilità di passare da uno stato A, B o C ad un qualunque altro stato è maggiore di zero, mentre una volta che una obbligazione è nello stato D non può più tornare negli altri tre.

2.9 Classificazione degli stati

In questa sezione supponiamo che lo spazio degli stati E sia discreto. La necessitá di questa ipotesi sará evidente fin dalla prima definizione.

Definizione 2.35 *Diciamo che lo stato $i \in E$* **comunica** *con lo stato $j \in E$ (e si scrive $i \to j$) se esiste $n > 0$ tale che*

$$p_{ij}^{(n)} = \mathbb{P}\{X_n = j \mid X_0 = i\} > 0$$

Definiamo poi per ogni $j \in E$ il **primo tempo di visita di** *j*

$$\tau_j := \inf\{n \mid X_n = j\}$$

e per ogni $i, j \in E$,

$$\rho_{ij} := \mathbb{P}\{\tau_j < +\infty \mid X_0 = i\}$$

Ovviamente si ha che $\rho_{ij} > 0$ se e solo se $i \to j$.

Nota 2.36 *Se $i \to j$ e $j \to k$, allora anche $i \to k$. Difatti $i \to j$ significa che esiste n tale che $p_{ij}^{(n)} > 0$, e $j \to k$ significa che esiste m tale che $p_{jk}^{(m)} > 0$; allora si ha che*

$$p_{ik}^{(n+m)} = \sum_{l \in E} p_{il}^{(n)} p_{lk}^{(m)} \geq p_{ij}^{(n)} p_{jk}^{(m)} > 0$$

Questo significa che per vedere quali stati comunicano non serve calcolarsi tutte le matrici di transizione P^n ma conta solo sapere in che posizione sono gli zeri nella matrice di transizione.

Esempio 2.37 Riprendiamo l'Esempio 2.34. Si vede immediatamente che gli stati A, B e C comunicano con tutti gli altri, mentre invece lo stato D comunica solo con se stesso.

In questo esempio notiamo subito che sullo stato D si puó dire qualcosa di piú: difatti si vede subito che, se si parte dallo stato D, non c'é modo di passare ad un altro stato. Questo induce alla seguente classificazione degli stati.

Definizione 2.38 *Diciamo che lo stato $i \in E$ é:*

- **ricorrente** *se $\rho_{ii} = 1$. Difatti in questo caso lo stato i viene visitato q.c. almeno una volta dopo ogni visita: questo significa che viene visitato un numero infinito di volte.*
- **transitorio** *se $\rho_{ii} < 1$. Difatti in questo caso c'é una probabilitá positiva che lo stato i non venga piú visitato dopo una visita: si puó dimostrare che questo implica che lo stato viene visitato un numero finito di volte.*
- **assorbente** *se $i \to j$ solo per $j = i$. Difatti questo significa che l'unico stato che si puó raggiungere da i é i stesso.*

Proposizione 2.39 *Siano $i, j \in E$. Allora:*

- *Se $j \to i$ ma $i \nrightarrow j$, allora j é transitorio.*
- *Se i é ricorrente e $i \to j$, allora anche j é ricorrente.*

Dato che il fatto di essere stati ricorrenti o transitori in un certo senso si propagano tramite la comunicazione, possiamo poi raggruppare gli stati di E in sottoinsiemi.

Definizione 2.40 *Un sottoinsieme $C \subset E$ si dice* **classe chiusa** *se per ogni $i \in C$ si ha che $i \to j$ implica $j \in C$. Una classe chiusa C si dice poi* **irriducibile** *se per ogni $i, j \in C$ si ha che $i \to j$.*

In particolare, uno stato assorbente costituisce da solo una classe irriducibile. In generale, in una classe irriducibile tutti gli stati comunicano tra di loro, mentre in una classe chiusa gli stati non comunicano con stati esterni, ma potrebbero non comunicare tutti tra di loro (potrebbe ad esempio trattarsi dell'unione di piú classi irriducibili).

Si puó dimostrare che esiste un'unica decomposizione di E del tipo

$$E = T \cup C_1 \cup C_2 \cup \dots$$

dove T é l'insieme di tutti gli stati transitori e le C_i sono tutte classi irriducibili, che possono essere in numero finito o infinito. Si puó anche dimostrare che se E ha cardinalitá finita, allora T é un sottoinsieme proprio di E, cioé esistono sicuramente stati ricorrenti.

2.10 Misure invarianti

Anche in questa sezione supponiamo che lo spazio degli stati E sia discreto. A differenza della sezione precedente, peró, i concetti qui presenti si riescono ad estendere a situazioni molto piú generali.

Definizione 2.41 *La misura (in genere di probabilitá) μ su E si dice* **misura invariante** *per la catena di Markov con matrice di transizione P se $\mu = \mu P$.*

Questo implica che se $\mathbb{P}_{X_0} = \mu$, allora $\mathbb{P}_{X_1} = \mu P = \mu$, e quindi per induzione tutte le $(X_n)_n$ sono identicamente distribuite con legge marginale μ.

Teorema 2.42 (Markov-Kakutani) *Se E é finito, allora esiste una misura invariante.*

Dimostrazione. (traccia) Indichiamo con $d := |E|$; allora ogni misura su E puó essere identificata con un vettore $\mu \in \mathbb{R}^d$ a coordinate non negative. Preso un $\mu \in \mathbb{R}^d$ con $\sum_{i=1}^d \mu_i = 1$ (cioé una misura di probabilitá), definiamo $\mu_n := \frac{1}{n}\sum_{k=0}^{n-1} \mu P^k$, che é ancora una probabilitá su E. Poiché l'insieme delle probabilitá su E, visto come sottoinsieme di $\mathbb{R}^d$, é un insieme chiuso e limitato, e quindi compatto, esiste un punto di accumulazione π, cioé tale che esiste una sottosuccessione convergente $\mu_{n_k} \to \pi$ nella topologia di $\mathbb{R}^d$. Allora si ha che

$$\pi P - \pi = \lim_{n\to\infty} \frac{1}{n}\left(\sum_{k=0}^{n-1} \mu P^{k+1} - \sum_{k=0}^{n-1} \mu P^k\right) = \lim_{n\to\infty} \frac{1}{n}\left(\mu P^n - \mu\right) = 0$$

poiché $|P| \leq 1$. In altre parole, π é una misura invariante. □

Notiamo che, anche se E é finito, non necessariamente la misura invariante é unica, nemmeno se ci restringiamo alle probabilitá. Inoltre, poiché la combinazione convessa di probabilitá invarianti é ancora una probabilitá invariante, l'insieme delle probabilitá invarianti é un insieme convesso.

Definizione 2.43 *Una misura π si dice* **reversibile** *se è soddisfatta la seguente equazione, detta* **del bilancio dettagliato***:*

$$\pi_i p_{ij} = \pi_j p_{ji} \qquad \forall i, j \in E$$

È immediato vedere che se π è reversibile allora è anche invariante. Difatti per ogni $j \in E$ si ha che

$$(\pi P)_j = \sum_{i\in E} \pi_i p_{ij} = \sum_{i\in E} \pi_j p_{ji} = \pi_j \sum_{i\in E} p_{ji} = \pi_j \cdot 1 = \pi_j$$

È utile osservare che è più facile soddisfare l'equazione del bilancio dettagliato che l'equazione di stazionarietà $\pi = \pi P$. Può però succedere che una misura sia stazionaria senza essere reversibile.

Presentiamo ora una condizione sotto cui la misura invariante non solo é unica, ma presenta anche delle interessanti proprietá di convergenza.

Definizione 2.44 *La matrice di transizione P si dice* **regolare** *se esiste $m > 0$ tale che $p_{ij}^{(m)} > 0$ per ogni $i, j \in E$.*

Se E é finito e la matrice P é regolare, allora significa che E é costituito da un'unica classe ricorrente irriducibile.

Teorema 2.45 (Markov) *Se P é regolare, allora esiste un'unica misura invariante π, con la proprietá che $\lim_{n\to\infty} p_{ij}^{(n)} = \pi_j$ per ogni $i \in E$.*

Corollario 2.46 *Se X é una catena di Markov con matrice di transizione P regolare e chiamiamo π l'unica misura invariante, allora $X_n \rightharpoonup \pi$ per ogni legge iniziale.*

Dimostrazione. Dimostrare la tesi equivale a dire che per ogni funzione f continua e limitata su E si ha che $E[f(X_n)] \to \int_E f\, d\pi$. Dato che E é numerabile, possiamo identificare f con $f_i := f(i)$, $i \in E$, e $\int_E f\, d\pi = \sum_{i\in E} f_i \pi_i$. Se chiamiamo μ la legge iniziale di X, allora

$$\lim_{n\to\infty} (\mu P^n)_j = \lim_{n\to\infty} \sum_{i\in E} \mu_i p_{ij}^{(n)} = \sum_{i\in E} \mu_i \pi_j = \pi_j \sum_{i\in E} \mu_i = \pi_j \cdot 1 = \pi_j$$

e quindi per ogni f continua e limitata su E si ha che

$$\lim_{n\to\infty} E[f(X_n)] = \lim_{n\to\infty} \int_E f\, d\mathbb{P}_{X_n} = \lim_{n\to\infty} \sum_{j\in E} f_j (\mu P^n)_j = \sum_{j\in E} f_j \pi_j = \int_E f\, d\pi$$

□

2.11 Arresto ottimale

Come applicazione della teoria vista finora, consideriamo il seguente problema: data una catena di Markov X, vogliamo massimizzare il valore assunto da una data funzione della variabile X_n quando la arrestiamo ad un tempo di arresto ottimale.

Poniamo ora il problema in modo più matematico. Supponiamo di avere una catena di Markov $X = (X_n)_n$ a valori in $(E, \mathcal{E})$ (non necessariamente discreto) definita sullo spazio probabilizzato $(\Omega, \mathcal{A}, \mathbb{P})$, di legge iniziale μ e nucleo di transizione N. Consideriamo inoltre $\varphi \in L^+(E, \mathcal{E})$ e $\gamma > 0$. Vogliamo trovare un tempo di arresto $\hat{\tau}$ che massimizzi la quantità $E[\gamma^\tau \varphi(X_\tau)]$ tra tutti i tempi di arresto $\tau \le K$, dove $K \in \mathbb{N}$ è un dato orizzonte temporale. Questo significa trovare un tempo di arresto $\hat{\tau} \le K$ tale che

$$E[\gamma^{\hat{\tau}} \varphi(X_{\hat{\tau}})] \ge E[\gamma^\tau \varphi(X_\tau)] \qquad \forall \tau \le K \text{ tempo di arresto}$$

In tutta questa sezione intenderemo che i tempi di arresto lo sono rispetto alla filtrazione naturale $(\mathcal{F}_n^X)_n$ di X. Un tempo di arresto $\hat{\tau}$ che soddisfa la proprietà sopra sarà detto **tempo di arresto ottimale**, o **ottimo**.

Il problema considerato sopra, come parecchi problemi del Calcolo delle Probabilità che coinvolgono processi di Markov, ha una soluzione data da un algoritmo deterministico.

Teorema 2.47 *Se definiamo la successione di funzioni $V_n : E \to \mathbb{R}^+$, $n = 0, \dots, K$, come*

$$\begin{aligned} V_K(x) &:= \varphi(x), \\ V_n(x) &:= \max(\varphi(x), \gamma N V_{n+1}(x)), \qquad n = K-1, \dots, 0 \end{aligned}$$

Allora il tempo di arresto

$$\hat{\tau} := \inf\{n \mid V_n(X_n) = \varphi(X_n)\}$$

è un tempo di arresto ottimo, e $E[\gamma^{\hat{\tau}} \varphi(X_{\hat{\tau}})] = E[V_0(X_0)]$.

Nota 2.48 Sicuramente $K \in \{n \mid V_n(X_n) = \varphi(X_n)\}$, quindi $\hat{\tau}$ è ben definito.

Dimostrazione. Consideriamo un tempo di arresto $\tau \leq K$. Per ogni $n = 0, \dots, K-1$ si ha allora che

$$\begin{aligned} E[\gamma^{\tau\wedge(n+1)} V_{\tau\wedge(n+1)}(X_{\tau\wedge(n+1)})] &= E[\gamma^\tau V_\tau(X_\tau)\mathbf{1}_{\{\tau<n+1\}} + \gamma^{n+1} V_{n+1}(X_{n+1})\mathbf{1}_{\{\tau\geq n+1\}}] = \\ &= E[\gamma^\tau V_\tau(X_\tau)\mathbf{1}_{\{\tau\leq n\}} + \gamma^{n+1} N V_{n+1}(X_n)\mathbf{1}_{\{\tau\geq n+1\}}] \leq \\ &\leq E[\gamma^\tau V_\tau(X_\tau)\mathbf{1}_{\{\tau\leq n\}} + \gamma^n V_n(X_n)\mathbf{1}_{\{\tau>n\}}] = \\ &= E[\gamma^{\tau\wedge n} V_{\tau\wedge n}(X_{\tau\wedge n})] \end{aligned}$$

per definizione delle $(V_n)_n$. Se però $\tau = \hat{\tau}$, allora su $\{\tau > n\}$ si ha che $V_n(X_n) = \gamma N V_{n+1}(X_n)$, e quindi la disuguaglianza è un'uguaglianza.

Quanto sopra visto significa che l'applicazione $n \to E[\gamma^{\tau\wedge n} V_{\tau\wedge n}(X_{\tau\wedge n})]$ è decrescente per ogni tempo di arresto $\tau \leq K$, e costante se $\tau = \hat{\tau}$. Questo significa che

$$E[V_0(X_0)] = E[\gamma^{\tau\wedge 0} V_{\tau\wedge 0}(X_{\tau\wedge 0})] \geq E[\gamma^{\tau\wedge K} V_{\tau\wedge K}(X_{\tau\wedge K})] = E[\gamma^\tau V_\tau(X_\tau)] \geq E[\gamma^\tau \varphi(X_\tau)]$$

dove l'ultima disuguaglianza segue dalla definizione delle $(V_n)_n$. Se però $\tau = \hat{\tau}$, allora $V_\tau(X_\tau) = \varphi(X_\tau)$ per definizione di $\hat{\tau}$, e quindi tutte le disuguaglianze di sopra sono uguaglianze. □

Nota 2.49 Se si vuole risolvere un problema di minimo invece che di massimo, l'algoritmo è esattamente lo stesso: basta sostituire il max che compare nella definizione di V_n con un min.

Esempio 2.50 (Problema del segretario) Questo esempio è un problema classico del Calcolo delle Probabilità, che consiste in questo: ci vengono presentate, una alla volta, K opportunità, e ad ogni istante possiamo solo scegliere se accettare quella che abbiamo davanti o rifiutarla e continuare, attendendone una migliore (che ovviamente potrebbe non arrivare). Una volta rifiutata, una opportunità è persa per sempre. Per questo, il problema è noto anche come "problema del matrimonio" o "problema dell'appartamento".

Vediamo il problema in una delle sue forme più semplici. Supponiamo di avere delle variabili aleatorie $X_0, \dots, X_K$ i.i.d. di legge $U(0,1)$, e di voler massimizzare $E[X_\tau]$ su tutti i tempi di arresto τ. Innanzitutto è molto facile verificare che $(X_n)_n$ è una catena di

Markov con spazio degli stati $E = (0,1)$, di legge iniziale $U(0,1)$ e nucleo di transizione $N(x,\cdot) = U(0,1)$ per ogni $x \in (0,1)$. Allora

$$(N\varphi)(x) = \int_E \varphi(y)\ N(x,dy) = \int_0^1 \varphi(y)\ dy \qquad \forall x \in (0,1)$$

Dobbiamo quindi massimizzare $E[\gamma^\tau \varphi(X_\tau)]$, con $\gamma = 1$, $\varphi(x) = x$. Applichiamo l'algoritmo dell'arresto ottimale:

$$\begin{aligned} V_K(x) &= x, \\ V_{K-1}(x) &= \max\left(x, \frac{1}{2}\right), \qquad \left(\text{infatti } \int_0^1 V_K(x)\ dx = \frac{1}{2}\right), \\ V_{K-2}(x) &= \max\left(x, \frac{5}{8}\right), \qquad \left(\int_0^1 V_{K-1}(x)\ dx = \frac{5}{8}\right), \end{aligned}$$

e per induzione si può dimostrare che

$$V_n(x) = \max(x, v_n), \qquad \text{con } v_n = \frac{1}{2}(1 + v_{n+1}^2)$$

La successione $(v_{K-n})_n$ tende a 1 molto lentamente: ecco una tabella per $K - n \leq 4$:

n	0	1	2	3	4
V_{K-n}	0	0.5	0.625	0.695	0.741

Un tempo di arresto ottimale è dato da

$$\hat{\tau} = \min\{n \leq K \mid V_n(X_n) = \varphi(X_n)\} = \min\{n \leq K \mid X_n \geq v_n\}$$

Bibliografia

[1] M. Capinski, E. Kopp, *Measure, integral and probability*, Springer 1999

[2] G. Folland, *Real analysis: modern techniques and their applications*, John Wiley & Sons, 1984

[3] A. Klenke, *Probability theory*, Springer, 2008

[4] G. Letta, *Probabilità elementare*, Zanichelli, 1994