Complementi sulla convergenza di variabili aleatorie¹

Lucia Caramellino

Nel Calcolo delle Probabilità esistono molte nozioni di convergenza di successioni di variabili aleatorie (v.a.). In queste brevi note, ci concentreremo su tre di queste: la *convergenza quasi certa* (q.c.), la *convergenza in probabilità* e la *convergenza in legge* (anche detta *convergenza debole*).

Nel seguito, assumeremo di aver fissato uno spazio di probabilità $(\Omega, \mathcal{A}, \mathbb{P})$ sul quale sono definite le v.a. di interesse, che in generale potranno essere a valori in $\mathbb{R}^d$. Per $x \in \mathbb{R}^d$, denoteremo con $|x|$ la norma euclidea standard: $x = (x_1, \dots, x_d)$ e $|x|^2 = \sum_{k=1}^d |x_k|^2$.

Definizione 1 Una successione di v.a. $X_1, X_2, \dots$ converge quasi certamente alla v.a. X se

$$\{\omega : \lim_{n \rightarrow \infty} X_n(\omega) = X(\omega)\}$$

ha probabilità 1. Equivalentemente, se

$$\{\omega : \lim_{n \rightarrow \infty} X_n(\omega) = X(\omega)\}^c \quad (1)$$

ha probabilità 0. In simboli, scriveremo $X_n \xrightarrow{\text{q.c.}} X$ o anche “ $X_n \rightarrow X$ q.c.”.

Dunque, la convergenza q.c. è una semplice convergenza puntuale a meno di un insieme di probabilità nulla: $X_n \xrightarrow{\text{q.c.}} X$ se e solo se esiste N tale che $\mathbb{P}(N) = 0$ e per ogni $\omega \notin N$ si ha $X_n(\omega) \rightarrow X(\omega)$ (basta infatti prendere N come nella (1)).

Definizione 2 Una successione di v.a. $X_1, X_2, \dots$ converge in probabilità alla v.a. X se per ogni $\delta > 0$ si ha

$$\lim_{n \rightarrow \infty} \mathbb{P}(|X_n - X| > \delta) = 0. \quad (2)$$

In simboli, scriveremo $X_n \xrightarrow{P} X$ o anche “ $X_n \rightarrow X$ in probabilità”.

Osservazione 3 La richiesta “ $> \delta$ ” nella (2) può essere sostituita da “ $\geq \delta$ ”, cioè $X_n \xrightarrow{P} X$ se e solo se

$$\text{per ogni } \delta > 0 \text{ si ha } \lim_{n \rightarrow \infty} \mathbb{P}(|X_n - X| \geq \delta) = 0. \quad (3)$$

Infatti, poiché $\{|X_n - X| > \delta\} \subset \{|X_n - X| \geq \delta\}$, si ha $\mathbb{P}(|X_n - X| > \delta) \leq \mathbb{P}(|X_n - X| \geq \delta)$, dunque (3) implica che $X_n \xrightarrow{P} X$. Viceversa, supponiamo $X_n \xrightarrow{P} X$. Per ogni $\delta > 0$ si ha $\{|X_n - X| \geq \delta\} \subset \{|X_n - X| > \delta/2\}$. Quindi, essendo $\delta/2 > 0$,

$$\mathbb{P}(|X_n - X| \geq \delta) \leq \mathbb{P}(|X_n - X| > \delta/2) \xrightarrow{n \rightarrow \infty} 0$$

e dunque (3) è vera.

¹Note di complemento al corso *Probabilità e Statistica Matematica*, CdLT in Matematica a Tor Vergata, aa 2013/2014

Facciamo subito vedere che il limite q.c. o in probabilità è “unico”, nel senso seguente:

Proposizione 4 Sia $\overset{*}{\rightarrow}$ una notazione che indica $\overset{\text{q.c.}}{\rightarrow}$ oppure $\overset{P}{\rightarrow}$. Se $X_n \overset{*}{\rightarrow} X$ e $X_n \overset{*}{\rightarrow} \bar{X}$ allora $\mathbb{P}(X = \bar{X}) = 1$.

Dunque, quello che si può dimostrare che il limite è “unico quasi certamente”, nel senso che due v.a. limite della stessa successione sono uguali con probabilità 1, o equivalentemente sono diverse con probabilità 0. Questo è quanto tipicamente si ha in probabilità (o, più in generale, nella teoria della misura).

DIMOSTRAZIONE. Supponiamo che $\overset{*}{\rightarrow}$ indichi $\overset{\text{q.c.}}{\rightarrow}$. Se $X_n \overset{\text{q.c.}}{\rightarrow} X$ esiste N tale che $\mathbb{P}(N) = 0$ e se $\omega \notin N$ allora $X_n(\omega) \rightarrow X(\omega)$. Ma si ha anche $X_n \overset{\text{q.c.}}{\rightarrow} \bar{X}$, dunque esiste $\bar{N}$ tale che $\mathbb{P}(\bar{N}) = 0$ e se $\omega \notin \bar{N}$ allora $X_n(\omega) \rightarrow \bar{X}(\omega)$. Quindi, se $\omega \notin N \cup \bar{N}$, si ha contemporaneamente $X_n(\omega) \rightarrow X(\omega)$ e $X_n(\omega) \rightarrow \bar{X}(\omega)$ e, per l’unicità del limite, si ottiene $X(\omega) = \bar{X}(\omega)$. Abbiamo dimostrato che

$$(N \cup \bar{N})^c \subset \{X = \bar{X}\}$$

e quindi, $\mathbb{P}(X = \bar{X}) \geq \mathbb{P}((N \cup \bar{N})^c) = 1 - \mathbb{P}(N \cup \bar{N}) = 1 - 0 = 1$, da cui la tesi.

Se $\overset{*}{\rightarrow}$ indica $\overset{P}{\rightarrow}$, abbiamo $\mathbb{P}(|X_n - X| > \delta) \xrightarrow{n \rightarrow \infty} 0$ e $\mathbb{P}(|X_n - \bar{X}| > \delta) \xrightarrow{n \rightarrow \infty} 0$, per ogni $\delta > 0$. Ora, per ogni n possiamo scrivere

$$|X - \bar{X}| = |X \pm X_n - \bar{X}| \leq |X - X_n| + |X_n - \bar{X}|.$$

Fissato $\delta > 0$, se $|X - X_n| \leq \delta/2$ e $|X_n - \bar{X}_n| \leq \delta/2$ allora $|X - \bar{X}| \leq \delta$. In notazioni insiemistiche, possiamo scrivere

$$\left\{|X - X_n| \leq \frac{\delta}{2}\right\} \cap \left\{|X_n - \bar{X}_n| \leq \frac{\delta}{2}\right\} \subset \{|X - \bar{X}| \leq \delta\}$$

e, passando al complementare,

$$\left\{|X - X_n| > \frac{\delta}{2}\right\} \cup \left\{|X_n - \bar{X}_n| > \frac{\delta}{2}\right\} \supset \{|X - \bar{X}| > \delta\}.$$

Allora,

$$\mathbb{P}(|X - \bar{X}| > \delta) \leq \mathbb{P}\left(|X - X_n| > \frac{\delta}{2}\right) + \mathbb{P}\left(|X_n - \bar{X}_n| > \frac{\delta}{2}\right) \xrightarrow{n \rightarrow \infty} 0,$$

ottenendo che $\mathbb{P}(|X - \bar{X}| > \delta) = 0$ per ogni $\delta > 0$, o anche $\mathbb{P}(|X - \bar{X}| \leq \delta) = 1$ per ogni $\delta > 0$. Ma allora, dev’essere necessariamente $\mathbb{P}(|X - \bar{X}| = 0) = 1$, da cui la tesi. $\square$

La convergenza q.c. è più forte della convergenza in probabilità. Infatti, si ha

Proposizione 5 Se $X_n \overset{\text{q.c.}}{\rightarrow} X$ allora $X_n \overset{P}{\rightarrow} X$.

DIMOSTRAZIONE. Ricordiamo che $x_n \rightarrow x$ se e solo se per ogni $\varepsilon > 0$ esiste $k \geq 1$ tale che per ogni $n \geq k$ si ha $|x_n - x| < \varepsilon$. Dunque, possiamo scrivere

$$\{\omega : \lim_{n \rightarrow \infty} X_n(\omega) = X(\omega)\} = \bigcap_{\varepsilon > 0} \bigcup_{k \geq 1} \bigcap_{n \geq k} \{|X_n - X| < \varepsilon\}.$$

Passando al complementare, che indicheremo con N , otteniamo

$$N = \bigcup_{\varepsilon > 0} \bigcap_{k \geq 1} \bigcup_{n \geq k} \{|X_n - X| \geq \varepsilon\}.$$

Per ipotesi, abbiamo $\mathbb{P}(N) = 0$. Ora, preso $\delta > 0$, si ha

$$\bigcap_{k \geq 1} \bigcup_{n \geq k} \{|X_n - X| \geq \delta\} \subset \bigcup_{\varepsilon > 0} \bigcap_{k \geq 1} \bigcup_{n \geq k} \{|X_n - X| \geq \varepsilon\} = N$$

e quindi, essendo N di probabilità nulla,

$$\mathbb{P}\left(\bigcap_{k > 0} \bigcup_{n \geq k} \{|X_n - X| \geq \delta\}\right) = 0.$$

Al variare di k , gli insiemi $A_k := \bigcup_{n \geq k} \{|X_n - X| \geq \delta\}$ formano una successione monotona decrescente (non crescente) di insiemi: $A_1 \supset A_2 \supset \dots$. Dunque, $A_k \downarrow A = \bigcap_{k \geq 1} A_k$ e per le proprietà di monotonia della probabilità,

$$\lim_{k \rightarrow \infty} \mathbb{P}(A_k) = \mathbb{P}(A) = 0.$$

Ora, osserviamo che per ogni $k \geq 1$,

$$\{|X_k - X| \geq \delta\} \subset \bigcup_{n \geq k} \{|X_n - X| \geq \delta\} = A_k$$

da cui segue che

$$\mathbb{P}(|X_k - X| \geq \delta) \leq \mathbb{P}(A_k) \xrightarrow{n \rightarrow \infty} 0,$$

e la tesi segue dall'Osservazione 3. $\square$

Le convergenze appena viste si “mantengono” sotto operazioni standard. Ad esempio, se $X_n \xrightarrow{\text{q.c.}} X$ (risp. $X_n \xrightarrow{P} X$) e $Y_n \xrightarrow{\text{q.c.}} X$ (risp. $Y_n \xrightarrow{P} Y$) allora $X_n + Y_n \xrightarrow{\text{q.c.}} X + Y$ (risp. $X_n + Y_n \xrightarrow{P} X + Y$). Si tratta però di proprietà che non discendono direttamente dalle definizioni, ed andrebbero dimostrate. La tecnica non è difficile, vediamo come si può fare in alcuni casi.

Proposizione 6 Sia $\xrightarrow{*}$ una notazione che indica $\xrightarrow{\text{q.c.}}$ oppure $\xrightarrow{P}$. Si ha:

- (i) se $X_n \xrightarrow{*} X$ e $Y_n \xrightarrow{*} Y$ allora $X_n + Y_n \xrightarrow{*} X + Y$;
- (ii) se $X_n \xrightarrow{*} c$, c costante, e f è una funzione continua allora $f(X_n) \xrightarrow{*} f(c)$;
- (iii) se $X_n \xrightarrow{*} X$ e $\{c_n\}_n$ è una successione di $\mathbb{R}$ tale che $c_n \rightarrow c$ allora $c_n X_n \xrightarrow{*} cX$;
- (iv) se $X_n \xrightarrow{*} X$ e $Y_n \xrightarrow{*} Y$ allora $Z_n = (X_n, Y_n) \xrightarrow{*} Z = (X, Y)$.

DIMOSTRAZIONE. (i) Supponiamo che $\xrightarrow{*}$ indichi $\xrightarrow{q.c.}$. Se $X_n \xrightarrow{q.c.} X$ esiste N_1 tale che $\mathbb{P}(N_1) = 0$ e se $\omega \notin N_1$ allora $X_n(\omega) \rightarrow X(\omega)$. E se $Y_n \xrightarrow{q.c.} Y$ esiste N_2 tale che $\mathbb{P}(N_2) = 0$ e se $\omega \notin N_2$ allora $Y_n(\omega) \rightarrow Y(\omega)$. Preso $N = N_1 \cup N_2$, allora $\mathbb{P}(N) = 0$ e se $\omega \notin N$ allora valgono contemporaneamente $X_n(\omega) \rightarrow X(\omega)$ e $Y_n(\omega) \rightarrow Y(\omega)$, dunque $X_n(\omega) + Y_n(\omega) \rightarrow X(\omega) + Y(\omega)$, da cui la tesi.

Supponiamo che $\xrightarrow{*}$ indichi $\xrightarrow{P}$. Si ha

$$|X_n + Y_n - (X + Y)| = |X_n - X + Y_n - Y| \leq |X_n - X| + |Y_n - Y|.$$

Quindi, preso $\delta > 0$, se $|X_n - X| \leq \frac{\delta}{2}$ e $|Y_n - Y| \leq \frac{\delta}{2}$ allora $|X_n + Y_n - (X + Y)| < \delta$. Riscritto in notazioni insiemistiche, otteniamo

$$\left\{ |X_n - X| \leq \frac{\delta}{2} \right\} \cap \left\{ |Y_n - Y| \leq \frac{\delta}{2} \right\} \subset \left\{ |X_n + Y_n - (X + Y)| \leq \delta \right\}$$

e, passando al complementare,

$$\left\{ |X_n + Y_n - (X + Y)| > \delta \right\} \subset \left\{ |X_n - X| > \frac{\delta}{2} \right\} \cup \left\{ |Y_n - Y| > \frac{\delta}{2} \right\}.$$

Allora,

$$\begin{aligned} \mathbb{P}(|X_n + Y_n - (X + Y)| > \delta) &\leq \mathbb{P}\left(\left\{ |X_n - X| > \frac{\delta}{2} \right\} \cup \left\{ |Y_n - Y| > \frac{\delta}{2} \right\}\right) \\ &\leq \mathbb{P}\left(|X_n - X| > \frac{\delta}{2}\right) + \mathbb{P}\left(|Y_n - Y| > \frac{\delta}{2}\right) \xrightarrow{n \rightarrow \infty} 0, \end{aligned}$$

e dunque $\lim_{n \rightarrow \infty} \mathbb{P}(|X_n + Y_n - (X + Y)| > \delta) = 0$, da cui $X_n + Y_n \xrightarrow{P} X + Y$.

(ii) Se $\xrightarrow{*}$ indica $\xrightarrow{q.c.}$, la tesi è immediata: se $X_n \xrightarrow{q.c.} c$ esiste N tale che $\mathbb{P}(N) = 0$ e se $\omega \notin N$ allora $X_n(\omega) \rightarrow c$ e, poiché f è continua, $f(X_n(\omega)) \rightarrow f(c)$. Dunque, $f(X_n) \xrightarrow{q.c.} f(c)$.

Supponiamo che $\xrightarrow{*}$ indichi $\xrightarrow{P}$. Essendo f continua, per ogni $\varepsilon > 0$ esiste $\delta > 0$ tale che se $|x - c| < \delta$ allora $|f(x) - f(c)| < \varepsilon$. Ciò significa che per ogni $\varepsilon > 0$ esiste $\delta > 0$ tale che

$$\{|X_n - c| < \delta\} \subset \{|f(X_n) - f(c)| < \varepsilon\}$$

o equivalentemente (solito passaggio al complementare)

$$\{|f(X_n) - f(c)| \geq \varepsilon\} \subset \{|X_n - c| \geq \delta\}.$$

Allora,

$$\mathbb{P}(|f(X_n) - f(c)| \geq \varepsilon) \leq \mathbb{P}(|X_n - c| \geq \delta) \xrightarrow{n \rightarrow \infty} 0$$

da cui otteniamo $\lim_{n \rightarrow \infty} \mathbb{P}(|f(X_n) - f(c)| \geq \varepsilon) = 0$. L'arbitrarietà di $\varepsilon > 0$ e l'Osservazione 3 danno $f(X_n) \xrightarrow{P} f(c)$.

(iii) Se $\xrightarrow{*}$ indica $\xrightarrow{q.c.}$, la tesi è immediata: se $X_n \xrightarrow{q.c.} c$ esiste N tale che $\mathbb{P}(N) = 0$ e per ogni $\omega \notin N$, $X_n(\omega) \rightarrow c$. Quindi, per $\omega \notin N$, si ha $c_n X_n(\omega) \rightarrow cX(\omega)$, da cui $c_n X_n \xrightarrow{q.c.} cX$.

Supponiamo che $\xrightarrow{*}$ indichi $\xrightarrow{P}$. Poiché $c_n \rightarrow c$, per ogni intero $K > 1$ esiste n_K tale che se $n \geq n_K$ allora $|c_n - c| < 1/K$. Inoltre, essendo convergente, $\{c_n\}_n$ è una successione limitata: esiste $C > 0$ tale che $|c_n| \leq C$ per ogni n . Dunque, fissato $K > 1$, per $n \geq n_K$ possiamo scrivere

$$|c_n X_n - cX| = |c_n X_n \pm c_n X - cX| \leq |c_n| \cdot |X_n - X| + |c_n - c| \cdot |X| \leq C |X_n - X| + \frac{1}{K} \cdot |X|.$$

Procedendo in modo analogo al punto (i), otteniamo facilmente che, per ogni $\delta > 0$,

$$\{|c_n X_n - cX| > \delta\} \subset \left\{ C |X_n - X| > \frac{\delta}{2} \right\} \cup \left\{ \frac{1}{K} \cdot |X| > \frac{\delta}{2} \right\}.$$

Dunque, per $n \geq n_K$,

$$\mathbb{P}(|c_n X_n - cX| > \delta) \leq \mathbb{P}\left(|X_n - X| > \frac{\delta}{2C}\right) + \mathbb{P}\left(|X| > \frac{K\delta}{2}\right).$$

Passando al limite per $n \rightarrow \infty$ (lim sup perché non sappiamo ancora se il limite esiste), si ha

$$\limsup_{n \rightarrow \infty} \mathbb{P}(|c_n X_n - cX| > \delta) \leq \limsup_{n \rightarrow \infty} \mathbb{P}\left(|X_n - X| > \frac{\delta}{2C}\right) + \mathbb{P}\left(|X| > \frac{K\delta}{2}\right) = \mathbb{P}\left(|X| > \frac{K\delta}{2}\right),$$

in cui abbiamo usato che $X_n \xrightarrow{P} X$. Ricapitolando, abbiamo dimostrato che per ogni $\delta > 0$ e $K > 1$ si ha

$$\limsup_{n \rightarrow \infty} \mathbb{P}(|c_n X_n - cX| > \delta) \leq \mathbb{P}\left(|X| > \frac{K\delta}{2}\right).$$

e quindi, passando al limite per $K \rightarrow \infty$, per ogni $\delta > 0$ si ha

$$\limsup_{n \rightarrow \infty} \mathbb{P}(|c_n X_n - cX| > \delta) \leq \limsup_{K \rightarrow \infty} \mathbb{P}\left(|X| > \frac{K\delta}{2}\right).$$

Studiamo il limite a destra e mostriamo che è nullo. Indicando con F la f.r. di X , si ha

$$\mathbb{P}\left(|X| > \frac{K\delta}{2}\right) = \mathbb{P}\left(X > \frac{K\delta}{2}\right) + \mathbb{P}\left(X < -\frac{K\delta}{2}\right) \leq 1 - F\left(\frac{K\delta}{2}\right) + F\left(-\frac{K\delta}{2}\right)$$

e per $K \rightarrow \infty$,

$$\mathbb{P}\left(|X| > \frac{K\delta}{2}\right) \rightarrow 1 - 1 + 0 = 0.$$

Possiamo quindi scrivere, per ogni $\delta > 0$,

$$\limsup_{n \rightarrow \infty} \mathbb{P}(|c_n X_n - cX| > \delta) = 0,$$

il che dà $\lim_{n \rightarrow \infty} \mathbb{P}(|c_n X_n - cX| > \delta) = 0$, cioè $c_n X_n \xrightarrow{P} cX$.

(iv) Ricordiamo che se $z = (x, y)$ allora si ha sempre $|z| \leq |x| + |y|$. Dunque,

$$|Z_n - Z| \leq |X_n - X| + |Y_n - Y|.$$

Tale disuguaglianza dà immediatamente la tesi per la convergenza q.c. Per la convergenza in probabilità, con la solita tecnica dimostriamo facilmente che, per ogni $\delta > 0$,

$$\{|Z_n - Z| > \delta\} \subset \left\{|X_n - X| > \frac{\delta}{2}\right\} \cup \left\{|Y_n - Y| > \frac{\delta}{2}\right\}$$

e di conseguenza,

$$\mathbb{P}(|Z_n - Z| > \delta) \leq \mathbb{P}\left(|X_n - X| > \frac{\delta}{2}\right) + \mathbb{P}\left(|Y_n - Y| > \frac{\delta}{2}\right).$$

La tesi segue passando al limite per $n \rightarrow \infty$. $\square$

Osservazione 7 La dimostrazione del punto (ii) in relazione alla convergenza q.c. vale anche se la v.a. limite non è costante, dunque possiamo affermare che se $X_n \xrightarrow{q.c.} X$ e f è continua allora $f(X_n) \xrightarrow{q.c.} f(X)$. In realtà è possibile dimostrare lo stesso risultato anche per la convergenza in probabilità: se $X_n \xrightarrow{P} X$ e f è continua allora $f(X_n) \xrightarrow{P} f(X)$. La dimostrazione richiede però alcune nozioni di teoria della misura che esulano dagli obiettivi del II anno di LT in Matematica (gli studenti interessati potranno vedere questo ed altri argomenti nei corsi successivi di probabilità). Ad esempio, sfruttando questa proprietà e la (iv) della Proposizione 6, è facile verificare [esercizio!] il seguente risultato:

se $X_n \xrightarrow{P} X$ e $Y_n \xrightarrow{P} Y$, con $\mathbb{P}(Y_n = 0) = 0$ per ogni n grande e $\mathbb{P}(Y = 0) = 0$, allora

$$\frac{X_n}{Y_n} \xrightarrow{P} \frac{X}{Y}.$$

La legge dei grandi numeri può essere riscritta in termini della convergenza in probabilità:

Teorema 8 Siano $X_1, X_2, \dots$ v.a. i.i.d. con momento secondo finito, e siano $\mu = \mathbb{E}(X_1)$, $\sigma^2 = \text{Var}(X_1)$. Posto $\bar{X}_n$, la media empirica delle prime n v.a., cioè

$$\bar{X}_n = \frac{1}{n} \sum_{k=1}^n X_k,$$

per ogni $\delta > 0$ si ha

$$\mathbb{P}(|\bar{X}_n - \mu| > \delta) \leq \frac{\sigma^2}{n\delta^2}.$$

Come conseguenza, $\bar{X}_n \xrightarrow{P} \mu$.

DIMOSTRAZIONE. Ricordiamo la disuguaglianza di Chebyshev: se Z è una v.a. di media m e varianza s^2 allora per ogni $\delta > 0$ si ha

$$\mathbb{P}(|Z - m| > \delta) \leq \frac{s^2}{\delta^2}.$$

Applichiamo tale disuguaglianza a $Z = \bar{X}_n$. Usando l'indipendenza delle v.a. (varianza della somma = somma delle varianze), si ha

$$\mathbb{E}(\bar{X}_n) = \frac{1}{n} \sum_{k=1}^n \underbrace{\mathbb{E}(X_k)}_{=\mu} = \mu \quad \text{e} \quad \text{Var}(\bar{X}_n) = \frac{1}{n^2} \sum_{k=1}^n \underbrace{\text{Var}(X_k)}_{=\sigma^2} = \frac{\sigma^2}{n}.$$

Dunque, per $\delta > 0$, si ha

$$\mathbb{P}(|\bar{X}_n - \mu| > \delta) \leq \frac{\sigma^2}{n\delta^2}.$$

Se $n \rightarrow \infty$, otteniamo $\lim_{n \rightarrow \infty} \mathbb{P}(|\bar{X}_n - \mu| > \delta) = 0$, e quindi $\bar{X}_n \xrightarrow{P} \mu$. $\square$

In realtà, nelle stesse ipotesi è possibile dimostrare un risultato più forte: la legge dei grandi numeri vale quasi certamente, ovvero $\bar{X}_n \xrightarrow{q.c.} \mu$ (anche questo si vedrà nei corsi più avanzati). In ogni caso, la legge dei grandi numeri afferma che, per n grande, la media empirica $\bar{X}_n$ dà un'approssimazione della media teorica μ . Questo è particolarmente utile in molte situazioni pratiche, quando μ non è nota e la si vuole stimare in qualche modo. Dunque, grazie alla LGN, possiamo stimare μ con la media empirica su n osservazioni, con n grande - quanto grande è un problema che affronteremo dopo il Teorema del Limite Centrale.

Un'applicazione classica della legge dei grandi numeri, o più precisamente delle tecniche che abbiamo usato per dimostrare la legge dei grandi numeri, è data dal teorema di Weierstrass sull'approssimazione uniforme delle funzioni continue su un compatto tramite polinomi. Il teorema che presentiamo, e che rappresenta uno dei tanti esempi di risultati non propriamente in probabilità che si possono dimostrare con tecniche puramente probabilistiche, dà un'espressione esplicita dei polinomi approssimanti in termini dei *polinomi di Bernstein*. Assumiamo che il compatto in questione sia l'intervallo $[0, 1]$ ed indichiamo con $C([0, 1])$ la classe delle funzioni continue su $[0, 1]$.

Teorema 9 Sia $f \in C([0, 1])$. Per $n \geq 1$, sia

$$P_n(x) = \sum_{k=0}^n f\left(\frac{k}{n}\right) \binom{n}{k} x^k (1-x)^{n-k}, \quad x \in [0, 1].$$

Allora,

$$\lim_{n \rightarrow \infty} \sup_{x \in [0, 1]} |f(x) - P_n(x)| = 0.$$

DIMOSTRAZIONE. Vediamo subito qual è il nesso con la probabilità. Fissato $x \in [0, 1]$, siano $X_1, X_2, \dots$ v.a. i.i.d. bernuolliane di parametro x . Ricordiamo che la media e la varianza comuni sono date da $\mu = x$ e $\sigma^2 = x(1-x)$ rispettivamente. Allora, la LGN dà $\bar{X}_n \xrightarrow{P} x$, e per la Proposizione 6 si ha

$$f(\bar{X}_n) \xrightarrow{P} f(x).$$

Quindi, per n grande, $f(x)$ si può approssimare con $f(\bar{X}_n)$. Inoltre, sappiamo che $S_n = X_1 + \dots + X_n$ ha legge $\text{Bi}(n, x)$, quindi

$$\mathbb{E}(f(\bar{X}_n)) = \mathbb{E}\left(f\left(\frac{S_n}{n}\right)\right) = \sum_{k=0}^n f\left(\frac{k}{n}\right) \mathbb{P}(S_n = k) = \sum_{k=0}^n f\left(\frac{k}{n}\right) \binom{n}{k} x^k (1-x)^{n-k} = P_n(x).$$

Dunque,

$$f(x) - P_n(x) = f(x) - \mathbb{E}(f(\bar{X}_n)) = \mathbb{E}(f(x) - f(\bar{X}_n)).$$

Dimostriamo allora che $\sup_{x \in [0, 1]} |\mathbb{E}(f(x) - f(\bar{X}_n))| \xrightarrow{n \rightarrow \infty} 0$. Per ogni $\delta > 0$, possiamo scrivere

$$\begin{aligned} |\mathbb{E}(f(x) - f(\bar{X}_n))| &\leq \mathbb{E}(|f(x) - f(\bar{X}_n)|) \\ &= \mathbb{E}(|f(x) - f(\bar{X}_n)| \mathbf{1}_{|\bar{X}_n - x| < \delta}) + \mathbb{E}(|f(x) - f(\bar{X}_n)| \mathbf{1}_{|\bar{X}_n - x| \geq \delta}) \end{aligned}$$

Poiché f è continua su un compatto, f è limitata: esiste $M > 0$ tale che $|f(y)| \leq M$ per ogni $y \in [0, 1]$. Dunque,

$$|f(x) - f(\bar{X}_n)| \leq |f(x)| + |f(\bar{X}_n)| \leq 2M$$

perché $x, \bar{X}_n \in [0, 1]$. Sostituendo, otteniamo

$$\begin{aligned} |\mathbb{E}(f(x) - f(\bar{X}_n))| &\leq \mathbb{E}(|f(x) - f(\bar{X}_n)| \mathbf{1}_{|\bar{X}_n - x| < \delta}) + 2M\mathbb{E}(\mathbf{1}_{|\bar{X}_n - x| \geq \delta}) \\ &= \mathbb{E}(|f(x) - f(\bar{X}_n)| \mathbf{1}_{|\bar{X}_n - x| < \delta}) + 2M\mathbb{P}(|\bar{X}_n - x| \geq \delta). \end{aligned}$$

Manipoliamo ora il primo termine. La funzione f è uniformemente continua (perché continua su un compatto): per ogni $\varepsilon > 0$ esiste $\delta > 0$ tale che per ogni $\xi, \eta \in [0, 1]$ con $|\xi - \eta| < \delta$ si ha $|f(\xi) - f(\eta)| < \varepsilon$. Fissato quindi $\varepsilon > 0$, prendiamo δ come appena scritto ed osserviamo che se $|\bar{X}_n - x| < \delta$ allora $|f(\bar{X}_n) - f(x)| < \varepsilon$. Allora,

$$|f(x) - f(\bar{X}_n)| \mathbf{1}_{|\bar{X}_n - x| < \delta} < \varepsilon \mathbf{1}_{|\bar{X}_n - x| < \delta} \leq \varepsilon.$$

Passando alle medie, per tali ε e δ otteniamo

$$|\mathbb{E}(f(x) - f(\bar{X}_n))| \leq \varepsilon + 2M\mathbb{P}(|\bar{X}_n - x| \geq \delta).$$

Ora, poiché $\bar{X}_n \xrightarrow{P} x$, passando al limite per $n \rightarrow \infty$ ed usando l'arbitrarietà di ε , otteniamo $|\mathbb{E}(f(x) - f(\bar{X}_n))| \rightarrow 0$, cioè $P_n(x) \rightarrow f(x)$. Vogliamo però dimostrare che tale convergenza è uniforme, dunque torniamo alla disuguaglianza scritta sopra ed applichiamo la disuguaglianza di Chebyshev (ricordiamo che $\text{Var}(\bar{X}_n) = \frac{1}{n}\text{Var}(X_1) = \frac{1}{n}x(1-x)$): otteniamo

$$|\mathbb{E}(f(x) - f(\bar{X}_n))| \leq \varepsilon + 2M \frac{x(1-x)}{n\delta^2} \leq \varepsilon + \frac{M}{2n\delta^2},$$

dove si è usato $x(1-x) \leq \frac{1}{4}$ per ogni $x \in [0, 1]$. Abbiamo quindi ottenuto a destra una stima valida per ogni $x \in [0, 1]$. Possiamo scrivere

$$\sup_{x \in [0, 1]} |\mathbb{E}(f(x) - f(\bar{X}_n))| \leq \varepsilon + \frac{M}{2n\delta^2}$$

e passando al limite per $n \rightarrow \infty$,

$$\limsup_{n \rightarrow \infty} \sup_{x \in [0, 1]} |\mathbb{E}(f(x) - f(\bar{X}_n))| \leq \varepsilon.$$

Ricordando che $\varepsilon > 0$ è arbitrario, otteniamo

$$\exists \lim_{n \rightarrow \infty} \sup_{x \in [0, 1]} |\mathbb{E}(f(x) - f(\bar{X}_n))| = 0,$$

da cui la tesi. $\square$

Vediamo ora alcune conseguenze della legge dei grandi numeri in statistica.

In statistica, uno *stimatore* per un parametro (incognito) ϑ è una funzione $T(X_1, \dots, X_n)$ delle osservazioni (=v.a. i.i.d.) $X_1, \dots, X_n$ che assume valori nello spazio di appartenenza di ϑ . Dunque, $\bar{X}_n$ è uno stimatore della media μ . Questo stimatore è anche *non distorto*, perché

in media vale esattamente la quantità che si vuole stimare, cioè μ : $\mathbb{E}(\bar{X}_n) = \mu$. Ma grazie alla LGN, $\bar{X}_n$ è anche *consistente*², perché $\bar{X}_n \xrightarrow{P} \mu$

Dunque, $\bar{X}_n$ è uno *stimatore non distorto e consistente* della media. E per la varianza? La LGN consente di trovare uno stimatore consistente? E questo stimatore è non distorto? La risposta è sì. Vediamo come si fa.

Il primo passo è di costruire la “varianza empirica”, cioè la varianza che si può calcolare sulla base di n osservazioni. Poniamo dunque

$$\hat{\sigma}_n^2 = \frac{1}{n} \sum_{k=1}^n (X_k - \bar{X}_n)^2.$$

Poiché $\hat{\sigma}_n^2 \geq 0$, $\hat{\sigma}_n^2$ può essere preso come stimatore della varianza (che, in generale, assume valori ≥ 0). Per studiare la consistenza, osserviamo che si ha

$$\hat{\sigma}_n^2 = \frac{1}{n} \sum_{k=1}^n X_k^2 - \bar{X}_n^2.$$

Infatti,

$$\begin{aligned} \hat{\sigma}_n^2 &= \frac{1}{n} \sum_{k=1}^n (X_k^2 - 2X_k\bar{X}_n + \bar{X}_n^2) = \frac{1}{n} \sum_{k=1}^n X_k^2 - 2\frac{1}{n} \bar{X}_n \sum_{k=1}^n X_k + \bar{X}_n^2 \\ &= \frac{1}{n} \sum_{k=1}^n X_k^2 - 2\bar{X}_n^2 + \bar{X}_n^2 = \frac{1}{n} \sum_{k=1}^n X_k^2 - \bar{X}_n^2 \end{aligned}$$

Ora, se $X_1, X_2, \dots$ sono i.i.d. anche la successione dei quadrati $X_1^2, X_2^2, \dots$ è fatta di v.a. i.i.d. E se X_1 ha momento quarto allora X_1^2 ha momento secondo. Dunque, se $X_1, X_2, \dots$ sono v.a. i.i.d. con momento quarto, la LGN applicata alla successione dei quadrati dà

$$\frac{1}{n} \sum_{k=1}^n X_k^2 \xrightarrow{P} \mathbb{E}(X_1^2).$$

Applicando ora il punto (i) e (ii) della Proposizione 6, otteniamo

$$\hat{\sigma}_n^2 = \frac{1}{n} \sum_{k=1}^n X_k^2 - \bar{X}_n^2 \xrightarrow{P} \mathbb{E}(X_1^2) - \mu^2 = \sigma^2.$$

Dunque, $\hat{\sigma}_n^2$ è uno *stimatore consistente* per σ^2 . È distorto? Sì:

$$\begin{aligned} \mathbb{E}(\hat{\sigma}_n^2) &= \frac{1}{n} \sum_{k=1}^n \mathbb{E}(X_k^2) - \mathbb{E}(\bar{X}_n^2) = \frac{1}{n} \sum_{k=1}^n \underbrace{(\text{Var}(X_k) + \mathbb{E}(X_k)^2)}_{=\sigma^2 + \mu^2} - \underbrace{(\text{Var}(\bar{X}_n) + \mathbb{E}(\bar{X}_n)^2)}_{=\frac{\sigma^2}{n} + \mu^2} \\ &= \sigma^2 - \frac{1}{n} \sigma^2 = \frac{n-1}{n} \sigma^2. \end{aligned}$$

²Per la precisione, si dovrebbe dire *consistente in senso debole*, visto che si sta usando la convergenza in probabilità. Se lo stimatore converge q.c. al parametro, si dice *consistente in senso forte*. E visto che qui si sta usando la LGN, che vale anche in senso forte, nelle nostre ipotesi si ha consistenza forte.

Ma non è difficile creare uno stimatore non distorto: basta considerare

$$s_n^2 = \frac{n}{n-1} \hat{\sigma}_n^2 = \frac{1}{n-1} \sum_{k=1}^n (X_k - \bar{X}_n)^2 = \frac{1}{n-1} \sum_{k=1}^n X_k^2 - \frac{n}{n-1} \bar{X}_n^2.$$

Ovviamente $\mathbb{E}(s_n^2) = \sigma^2$. Inoltre, $s_n^2 = c_n \sigma_n^2$, con $c_n = \frac{n}{n-1} \rightarrow 1$ e $\sigma_n^2 \xrightarrow{P} \sigma^2$. Quindi, per (iii) della Proposizione 6, otteniamo $s_n^2 \xrightarrow{P} \sigma^2$. Abbiamo quindi costruito uno stimatore consistente e non distorto della varianza. Possiamo riassumere come segue:

Proposizione 10 *Siano $X_1, X_2, \dots$ v.a. i.i.d. con momento secondo, di media μ e varianza σ^2 . Allora,*

$$\bar{X}_n = \frac{1}{n} \sum_{k=1}^n X_k.$$

è uno stimatore non distorto e consistente della media μ :

$$\mathbb{E}(\bar{X}_n) = \mu \text{ per ogni } n \text{ e } \bar{X}_n \xrightarrow{P} \mu.$$

Se per di più le v.a. hanno momento quarto, allora

$$s_n^2 = \frac{1}{n-1} \sum_{k=1}^n (X_k - \bar{X}_n)^2 = \frac{1}{n-1} \sum_{k=1}^n X_k^2 - \frac{n}{n-1} \bar{X}_n^2.$$

è uno stimatore non distorto e consistente della varianza σ^2 :

$$\mathbb{E}(s_n^2) = \sigma^2 \text{ per ogni } n \text{ e } s_n^2 \xrightarrow{P} \sigma^2.$$

Vediamo ora l'ultimo tipo di convergenza che affronteremo in questo corso. Limiteremo la trattazione a v.a. a valori in $\mathbb{R}$.

Definizione 11 *Siano date una successione di v.a. $X_1, X_2, \dots$ ed una v.a. X , tutte a valori in $\mathbb{R}$. Sia F_n la f.r. di X_n , $n \geq 1$, e sia F la f.r. di X . La successione $X_1, X_2, \dots$ converge in legge alla v.a. X se*

$$\lim_{n \rightarrow \infty} F_n(x) = F(x) \text{ per ogni } x \in \mathbb{R} \text{ tale che } \Delta F(x) = 0,$$

dove $\Delta F(x) = F(x) - F(x^-)$ denota il salto di F in x (dunque, $\Delta F(x) = 0$ significa che x è un punto di continuità per F). In simboli, scriveremo $X_n \xrightarrow{L} X$ o anche " $X_n \rightarrow X$ in legge".

Osserviamo che la convergenza in legge esprime il comportamento asintotico della successione delle leggi, non delle variabili aleatorie. Ad esempio, se $X_1, X_2, \dots$ sono identicamente distribuite (tutte con la stessa legge), allora ovviamente c'è convergenza in legge (la successione delle f.r. è costante!), indipendentemente da cosa le v.a. $X_1, X_2, \dots$ rappresentano.

Per quanto riguarda l'“unicità” del limite, osserviamo che, dalla definizione, abbiamo l'unicità della funzione di ripartizione limite in ogni punto x di continuità. Ciò garantisce l'unicità della

funzione di ripartizione limite: questo segue dal fatto che le funzioni monotone e limitate possono avere una quantità al più numerabile di discontinuità (si vedrà nei corsi più avanzati). Dunque, la v.a. limite in legge è “unica” nel senso che ha una unica, prefissata, funzione di ripartizione. E questo è quanto necessita: la convergenza in legge è data dalla convergenza delle funzioni di ripartizione!

La convergenza in legge è più debole delle precedenti. Infatti, si ha

Proposizione 12 *Se $X_n \xrightarrow{P} X$ allora $X_n \xrightarrow{L} X$.*

DIMOSTRAZIONE. Sia $x \in \mathbb{R}$ fissato. Preso $\delta > 0$, si ha

$$\begin{aligned} F_n(x) &= \mathbb{P}(X_n \leq x) = \mathbb{P}(X_n \leq x, X \leq x + \delta) + \mathbb{P}(X_n \leq x, X > x + \delta) \\ &\leq \mathbb{P}(X \leq x + \delta) + \mathbb{P}(X_n \leq x, X > x + \delta) \\ &= F(x + \delta) + \mathbb{P}(X_n \leq x, X > x + \delta). \end{aligned}$$

Poiché $\{X_n \leq x, X > x + \delta\} \subset \{|X_n - X| > \delta\}$, otteniamo

$$F_n(x) \leq F(x + \delta) + \mathbb{P}(|X_n - X| > \delta).$$

Passando al limite per $n \rightarrow \infty$ (lim sup perché non sappiamo ancora se esiste) ed usando l'ipotesi $X_n \xrightarrow{P} X$, otteniamo

$$\limsup_{n \rightarrow \infty} F_n(x) \leq F(x + \delta). \quad (4)$$

Vediamo ora una minorazione del lim inf. Si ha

$$\begin{aligned} 1 - F_n(x) &= \mathbb{P}(X_n > x) = \mathbb{P}(X_n > x, X > x - \delta) + \mathbb{P}(X_n > x, X \leq x - \delta) \\ &\leq \mathbb{P}(X > x - \delta) + \mathbb{P}(X_n > x, X \leq x - \delta) \\ &= 1 - F(x - \delta) + \mathbb{P}(X_n > x, X \leq x - \delta). \end{aligned}$$

Poiché $\{X_n > x, X \leq x - \delta\} \subset \{|X_n - X| > \delta\}$, otteniamo

$$1 - F_n(x) \leq 1 - F(x - \delta) + \mathbb{P}(|X_n - X| > \delta).$$

Passando al lim sup per $n \rightarrow \infty$ ed usando l'ipotesi $X_n \xrightarrow{P} X$, otteniamo

$$\limsup_{n \rightarrow \infty} (1 - F_n(x)) \leq 1 - F(x - \delta).$$

Ma $\limsup_{n \rightarrow \infty} (1 - F_n(x)) = 1 - \liminf_{n \rightarrow \infty} F_n(x)$, quindi

$$\liminf_{n \rightarrow \infty} F_n(x) \geq F(x - \delta).$$

Mettendo insieme questa disuguaglianza con la (4), per ogni $x \in \mathbb{R}$ e $\delta > 0$ abbiamo

$$F(x - \delta) \leq \liminf_{n \rightarrow \infty} F_n(x) \leq \limsup_{n \rightarrow \infty} F_n(x) \leq F(x + \delta).$$

Passando al limite per $\delta \rightarrow 0$, per ogni $x \in \mathbb{R}$ si ha

$$F(x^-) \leq \liminf_{n \rightarrow \infty} F_n(x) \leq \limsup_{n \rightarrow \infty} F_n(x) \leq F(x^+) = F(x).$$

Ora, se $\Delta F(x) = F(x) - F(x^-) = 0$, le quantità a destra e a sinistra sono uguali. Dunque, se $\Delta F(x) = 0$, esiste $\lim_{n \rightarrow \infty} F_n(x)$ e tale limite è proprio $F(x)$, quindi $X_n \xrightarrow{\mathcal{L}} X$. $\square$

In generale, se $X_n \xrightarrow{\mathcal{L}} X$ non è detto che $X_n \xrightarrow{P} X$. Ma tale implicazione è vera se la v.a. limite è costante:

Proposizione 13 *Se $X_n \xrightarrow{\mathcal{L}} c$ allora $X_n \xrightarrow{P} c$.*

DIMOSTRAZIONE. Indichiamo con F_n la f.r. di X_n e con F la f.r. di $X = c$. Ricordiamo che

$$F(x) = \begin{cases} 0 & x < c \\ 1 & x \geq c \end{cases}$$

Dunque $\Delta F(x) = 0$ se e solo se $x \neq c$ e quindi $X_n \xrightarrow{\mathcal{L}} X$ equivale a $F_n(x) \rightarrow F(x)$ per ogni $x \neq c$. Ora, preso $\delta > 0$, si ha

$$\begin{aligned} \mathbb{P}(|X_n - c| > \delta) &= \mathbb{P}(\{X_n < c - \delta\} \cup \{X_n > c + \delta\}) \\ &\leq \mathbb{P}(X_n \leq c - \delta) + \mathbb{P}(X_n > c + \delta) \\ &= F_n(c - \delta) + 1 - F_n(c + \delta). \end{aligned}$$

Poiché $\delta > 0$, $c - \delta < c < c + \delta$ e quindi $F_n(c \pm \delta) \rightarrow F(c \pm \delta)$. Dunque, otteniamo

$$\limsup_{n \rightarrow \infty} \mathbb{P}(|X_n - c| > \delta) \leq F(c - \delta) + 1 - F(c + \delta) = 0 + 1 - 1 = 0,$$

da cui $X_n \xrightarrow{P} X$. $\square$

Esiste un criterio molto utile per la convergenza in legge, legato alla convergenza (puntuale) delle funzioni caratteristiche (che, come sappiamo, caratterizzano le leggi). Si tratta di un risultato molto noto in letteratura (la cui dimostrazione sarà tralasciata in questo corso perché prevede nozioni piuttosto avanzate), dovuto a Paul Lévy:

Teorema 14 *Sia $X_1, X_2, \dots$ una successione di v.a., con funzione caratteristica rispettivamente $\varphi_1, \varphi_2, \dots$, e sia X una v.a. con funzione caratteristica φ . Allora, $X_n \xrightarrow{\mathcal{L}} X$ se e solo se $\lim_{n \rightarrow \infty} \varphi_n(\theta) = \varphi(\theta)$, per ogni θ .*