2 Modalità di convergenza

Gli strumenti fondamentali per studiare procedure e metodi statistici da un punto di vista matematico sono forniti dai teoremi di convergenza asintotica. Prima di poterli enunciare, è molto importante capire quali siano le principali modalità di convergenza per sequenze di varianili aleatorie, e studiare le relazioni che intercorrono tra loro.

Spiegazione Semplice

Cosa significa "Convergenza Asintotica"?
"Asintotico" è un modo complicato per dire: "cosa succede quando n (la dimensione del nostro campione, o il numero di esperimenti) diventa incredibilmente grande, tendendo all'infinito?". Stiamo studiando come una successione di variabili aleatorie (ad esempio, la media di n lanci di un dado) si "stabilizza" e "converge" verso un valore o una distribuzione finale. Capire questo ci permette di fidarci dei nostri metodi statistici quando abbiamo molti dati.

Definition 0.1 (((1): Convergenza in Legge, o Convergenza in Distribuzione))**.** *Sia $\{X_n\}_{n\in\mathbb{N}}$ una successione di variabili aleatorie; diremo che la sequenza converge in legge alla variabile aleatoria X (scritto $X_n \to_d X$) se e solo se si ha*

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x) \quad \textit{per ogni punto di continuità di } F_X(.).$$

Spiegazione Semplice

Cosa dice in parole povere? Questa è la forma più "debole" di convergenza.

- **Cosa NON dice:** Non dice che i *valori* di X_n e X sono vicini tra loro.
- **Cosa DICE:** Dice che la *distribuzione di probabilità* di X_n (pensa al suo "istogramma" o alla sua "curva a campana") assomiglia sempre di più alla distribuzione di X.
- $F_X(x)$ **(Funzione di Ripartizione):** È la funzione che risponde alla domanda: "Qual è la probabilità che la mia variabile X sia minore o uguale a x?". La convergenza richiede che il limite delle funzioni di ripartizione $F_{X_n}(x)$ sia uguale a $F_X(x)$.
- **Il punto chiave:** La probabilità $\mathbb{P}(X_n \leq x)$ diventa (quasi) la stessa di $\mathbb{P}(X \leq x)$, per n molto grande.

Remark 0.2. *Nella definizione precedente non abbiamo specificato né lo spazio di probabilità su cui è definita la sequenza $\{X_n\}_{n\in\mathbb{N}}$ né quello su cui è definita X. Non è stata una dimenticanza: non è infatti necessario che tali spazi siano gli stessi, anzi lo spazio di probabilità può persino essere diverso al variare di n, cioè potremmo considerare v.a. definite su sequenze $\{\Omega_n, \mathfrak{S}_n, \mathbb{P}_n\}$.*

Remark 0.3. *Si potrebbe pensare di introdurre una definizione diversa di convergenza in probabilità, lasciando cadere il requisito che la convergenza avvenga solo nei punti di continuità della funzione di distribuzione $F_X(.)$. Si vede facilmnente che questo darebbe adito però ad alcuni gravi paradossi. Si consideri ad esempio la sequenza deterministica $X_n = \frac{1}{n}$ considerate come variabili aleatorie, le X_n hanno funzione di ripartizione $F_{X_n}(x) = \mathbb{I}_{[\frac{1}{n},\infty)}(x)$. Prendiamo adesso la variabile aleatoria identicamente uguale a zero $X = 0$, la cui funzione di distribuzione è $F_X(x) = \mathbb{I}_{[0,\infty)}(x)$; abbiamo che*

$$\lim_{n\to\infty} F_{X_n}(0) = \lim_{n\to\infty} 0 = 0 \neq F_X(0) = 1.$$

Pertanto se richiedessimo la convergenza delle funzioni di continuità in ogni punto dovremmo concludere che la sequenza $\frac{1}{n}$ non converge in distribuzione a 0, quando $n \to \infty$.

Spiegazione Semplice

Perché la clausola "punti di continuità"? Questo esempio è la risposta.

- La successione $X_n = 1/n$ (che è $1, 1/2, 1/3, \ldots$) *ovviamente* converge a $X = 0$.
- La funzione di ripartizione $F_X(x)$ della variabile $X = 0$ ha un "salto" (una discontinuità) proprio in $x = 0$.
- Se guardiamo $F_{X_n}(0)$ (la probabilità che $1/n$ sia ≤ 0), questa è **sempre 0**.
- Ma $F_X(0)$ (la probabilità che 0 sia ≤ 0) è **1**.
- Quindi, nel punto $x = 0$, $0 \neq 1$, e la convergenza fallirebbe.

Per evitare questo paradosso, la definizione "ignora" questi specifici punti di salto.

Example 0.4. *Sia X_n una sequenza di variabilie binomiali con funzione di probabilità discreta*

$$Pr\{X_n = k\} = \binom{n}{k} p_n^k (1-p_n)^{n-k}, \quad k = 0, 1, 2, ..., n \text{ (0 altrimenti)},$$

dove $p_n := \frac{\lambda}{n}$ $\lambda > 0$. Abbiamo che $X_n \to_d X$, $X \sim Pois(\lambda)$. Infatti si verifica facilmente che

$$\begin{aligned}
\lim_{n\to\infty} Pr\{X_n = k\} &= \lim_{n\to\infty} \binom{n}{k} p_n^k (1-p_n)^{n-k} \\
&= \lim_{n\to\infty} \frac{n!}{k!(n-k)!} \frac{\lambda^k}{n^k} (1 - \frac{\lambda}{n})^n (1 - \frac{\lambda}{n})^{-k} \\
&= \frac{\lambda^k}{k!} \left(\lim_{n\to\infty} \frac{n!}{(n-k)!} \frac{1}{n^k} \right) \left(\lim_{n\to\infty} \left(1 - \frac{\lambda}{n}\right)^n \right) \left(\lim_{n\to\infty} \left(1 - \frac{\lambda}{n}\right)^{-k} \right) \\
&= \frac{\lambda^k}{k!} e^{-\lambda},
\end{aligned}$$

usando limiti notevoli conosciuti dai primi corsi di Analisi.

Spiegazione Semplice

Esempio: Binomiale $\to$ Poisson Questo è un esempio classico:

- La **Binomiale** conta i successi su n tentativi (es. quanti "Testa" su n lanci).
- La **Poisson** ($Pois(\lambda)$) conta "eventi rari" in un intervallo (es. quante telefonate arrivano a un call center in un'ora).

Se hai un numero n *altissimo* di tentativi (es. $n = 1.000.000$), ma una probabilità p_n *bassissima* (es. $p_n = \lambda/n = 5/1.000.000$), la Binomiale (che è difficile da calcolare) si comporta in pratica come una Poisson con media λ. La convergenza in legge ci dà la base teorica per fare questa approssimazione.

Definition 0.5 (((2): Convergenza in Probabilità o Convergenza Debole))**.** *Sia $\{X_n\}_{n\in\mathbb{N}}$una successione di variabili aleatorie definita sullo spazio di probabilità $\{\Omega, \mathfrak{S}, \mathbb{P}\}$; sia inoltre X, $X : \Omega \to \mathbb{R}$ Runa variabile aleatoria definita sullo stesso spazio di probabilità. Diremo che la sequenza converge in probabilità alla variabile aleatoria X (scritto $X_n \to_p X$) se e solo se si ha*

$$\lim_{n\to\infty} \mathbb{P}\{|X_n - X| > \epsilon\} = 0 \quad \text{per ogni } \epsilon > 0.$$

Spiegazione Semplice

Cosa dice in parole povere? Questa è una convergenza più "forte". Qui stiamo dicendo che i *valori* di X_n e X si avvicinano.
La formula $\mathbb{P}\{|X_n - X| > \epsilon\} \to 0$ si legge così: "La probabilità ($\mathbb{P}$) che la *distanza* (|...|) tra X_n e X sia più grande di un piccolo errore ϵ (non importa quanto piccolo, es. $\epsilon = 0.001$)... tende a zero man mano che n cresce".
In pratica: per n grande, è quasi impossibile che X_n e X siano "lontani" tra loro.

Lemma 0.6 (((2) ⇒ (1)))**.** *La convergenza in probabilità implica la convergenza in distribuzione, cioè* $\{X_n \to_p X\} \Rightarrow \{X_n \to_d X\}$.

Spiegazione Semplice

Se i valori di X_n e X tendono a essere "vicini" (Convergenza in Probabilità, la 2), è logico che anche le loro distribuzioni (i loro "istogrammi") debbano assomigliarsi (Convergenza in Legge, la 1).
La (2) è un requisito più stringente della (1). Se vale la (2), la (1) vale automaticamente.

Dimostrazione. Abbiamo che

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

Similmente

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

e dunque

$$F_X(x-\epsilon) - \mathbb{P}\{|X_n - X| > \epsilon\} \le F_{X_n}\{X \le x\}.$$

Di conseguenza

$$\begin{aligned} F_X(x-\epsilon) &= \lim \inf_{n\to\infty} [F_X(x-\epsilon) - \mathbb{P}\{|X_n - X| > \epsilon\}] \\ &\le \lim_{n\to\infty} F_{X_n}\{X \le x\} \\ &\le \lim \sup_{n\to\infty} F_{X_n}\{X \le x\} \\ &\le \lim \sup_{n\to\infty} [\mathbb{P}\{|X_n - X| > \epsilon\} + F_X\{X \le x + \epsilon\}] \\ &= F_X(x+\epsilon). \end{aligned}$$

Dato che ϵ è arbitrario, la convergenza segue immediatamente per tutti i punti di continuità di $F_X(.)$. □

Remark 0.7. *E' immediato verificare che l'implicazione inversa non vale in generale; strettamente parlando, non avrebbe nemmeno senso porre la domanda, visto che (2) richiede che le variabili siano definite tutte sullo stesso spazio di probabilità, mentre (1) no. Comunque a fini esplicativi si possono prendere banali controesempi: sia* $\{X_n\}_{n\in\mathbb{N}}$ *una successione di variabili aleatorie Gaussiane standard indipendenti, e sia* X *una altra Gaussiana standard indipendente da questa sequenza: è ovvio che* $X_n \triangleq X$ *pertanto* $X_n \to_d X$*; mentre invece*

$$Pr\{|X_n - X| > \epsilon\} = Pr\{|X| > \frac{\epsilon}{\sqrt{2}}\} \not\to 0,$$

dove abbiamo usato $X_n - X \triangleq N(0, 2)$*. Una parziale implicazione inversa esiste però nel caso in cui sia abbia convergenza in distribuzione ad una costante.*

Spiegazione Semplice

Contro-esempio: (1) NON implica (2) Questo è un contro-esempio cruciale. Immagina X_n come il risultato di una Gaussiana standard (curva a campana) e X come il risultato di *un'altra* Gaussiana standard, indipendente dalla prima.

- **Convergono in Legge (1)?** Sì! Hanno la stessa identica distribuzione. L'istogramma di X_n è già uguale all'istogramma di X.
- **Convergono in Probabilità (2)?** No! La probabilità che la loro differenza sia grande (es. $|X_n - X| > \epsilon$) non va a zero. Poiché sono indipendenti, X_n e X possono essere molto diversi a ogni estrazione. La loro differenza $X_n - X$ è a sua volta una variabile aleatoria (una $N(0, 2)$) e la probabilità che sia "lontana da zero" non svanisce.

Il fatto che le distribuzioni siano le stesse non significa che i valori estratti in un dato momento siano vicini.

Lemma 0.8. *Sia X_n una successione di variabili aleatorie tali per cui $X_n \to_d c$ dove $c \in \mathbb{R}$ una qualsiasi costante. Allora esiste uno spazio di probabilità $\{\Omega, \mathfrak{S}, \mathbb{P}\}$ ed una successione di variabili aleatorie $\{X'_n\}$ definite su questo spazio tali per cui $X'_n \underline{d} X_n$ per ogni n e $X'_n \to_p c$.*

Spiegazione Semplice

L'eccezione: la costante L'implicazione (1) $\Rightarrow$ (2) è falsa in generale, *tranne* in un caso speciale: se il limite è una costante c (un numero, non una variabile aleatoria).
Se la distribuzione di X_n "collassa" su un singolo numero c (es. la varianza della media campionaria va a zero), allora la convergenza in legge e la convergenza in probabilità diventano la stessa cosa.

Dimostrazione. Notiamo innanzitutto che $F_c(x) = \mathbb{I}_{[c,\infty)}(x)$, da cui

$$\begin{aligned} Pr\{|X'_n - c| > \epsilon\} &= Pr\{X'_n > c + \epsilon\} + Pr\{X'_n < c - \epsilon\} \\ &= 1 - F_{X_n}(c + \epsilon) + F_{X_n}(c - \epsilon) \\ &\to 1 - \mathbb{I}_{[c,\infty)}(c + \epsilon) + \mathbb{I}_{[c,\infty)}(c - \epsilon) = 0. \end{aligned}$$

□

Remark 0.9. *Intuitivamente, la (1) sta richiedendo che X_n e X "tendano a dare gli stessi numeri con la stessa probablità", mentre la (2) sta richiedendo che "in una singola estrazione, il valore di X_n e X tende ad essere vicino". E' quindi naturale che la (2) sia una richiesta più forte della (1).*

Definition 0.10 (((3): Convergenza in Media r-esima))**.** *Sia $\{X_n\}_{n\in\mathbb{N}}$ una successione di variabili aleatorie definita sullo spazio di probabilità $\{\Omega, \mathfrak{S}, \mathbb{P}\}$; sia inoltre X, $X : \Omega \to \mathbb{R}$ Runa variabile aleatoria definita sullo stesso spazio di probabilità. Diremo che la sequenza converge in media r-esima alla variabile aleatoria X (scritto $X_n \to_r X$) se e solo se si ha*

$$\lim_{n\to\infty} \mathbb{E}|X_n - X|^r = 0.$$

Spiegazione Semplice

Cosa dice in parole povere? Questa è una convergenza ancora più forte. Non chiede solo *che la probabilità* dell'errore vada a zero, ma che il *valore medio* (atteso) dell'errore (elevato alla r) vada a zero.

- **Caso $r = 1$ (Media):** $\mathbb{E}[|X_n - X|] \to 0$. La media dell'errore assoluto va a zero.
- **Caso $r = 2$ (Media Quadratica):** $\mathbb{E}[(X_n - X)^2] \to 0$. Questo è lo "Scarto Quadratico Medio" (MSE) usato ovunque in statistica. È un requisito molto forte perché "punisce" molto gli errori grandi (elevandoli al quadrato).

Remark 0.11. *Affinchè la definizione abbia senso stiamo implicitamente richiedendo che tutte le variabili aleatorie coinvolte abbiano momenti r-esimi finiti.*

Remark 0.12. *I casi più importanti sono quelli che corrispondono ar* $r = 2$ *(Convergenza in Media Quadratica) er* $r = 1$.

Lemma 0.13. *La convergenza in media r-esima implica la convergenza in probabilità, cioè* $\{X_n \to_r X\} \Rightarrow \{X_n \to_p X\}$.

> **Spiegazione Semplice**
>
> Questo passaggio è logico e si basa sulla Disuguaglianza di Markov (vista nel Capitolo 1). Se la *media* del tuo errore (es. $\mathbb{E}[|X_n - X|^r]$) sta andando a zero, allora la *probabilità* che quell'errore sia più grande di un numero ϵ deve per forza andare a zero anche lei.
> La (3) è più forte della (2).

Dimostrazione. La dimostrazione è una immediata conseguenza della disuguaglianza di Markov:

$$\lim_{n\to\infty} Pr\{|X_n - X| > \epsilon\} \leq \lim_{n\to\infty} \frac{\mathbb{E}[|X_n - X|^r]}{\epsilon^r}$$

□

Remark 0.14. *L'implicazione opposta è falsa in generale; in effetti, la convergenza in probabilità non implica nemmeno l'esistenza di momenti di un ordine qualsiasi* $r > 0$. *Un controesempio più interessante è fornito dalla seguente sequenza:*

$$X_n = \begin{cases} 0 & \textit{con probabilità } 1 - \frac{1}{n} \\ n & \textit{con probabilità } \frac{1}{n} \end{cases}$$

Questa sequenza ha valor medio finito e costante $\mathbb{E}[X_n] = 1$), *ma converge in probabilità alla variabile aleatoria identicamente pari a zero, che ha ovviamente valor medio nullo.*

> **Spiegazione Semplice**
>
> **Contro-esempio: (2) NON implica (3)** Questo è un contro-esempio fantastico. X_n vale 0 quasi sempre, ma con una probabilità piccolissima $1/n$ "esplode" e vale n.
>
> - **Converge in Probabilità (2) a 0?** Sì. La probabilità che X_n sia "lontano" da 0 (cioè che sia n) è $\mathbb{P}(|X_n - 0| > \epsilon) = 1/n$. Man mano che $n \to \infty$, questa probabilità va a 0.
> - **Converge in Media r=1 (3) a 0?** No. Calcoliamo la media: $\mathbb{E}[|X_n - 0|] = \left(0 \times \left(1 - \frac{1}{n}\right)\right) + \left(n \times \frac{1}{n}\right) = 0 + 1 = 1$. La media dell'errore è *costante* e non va a 0.
>
> La (2) è "ingannata" dal fatto che l'evento $X_n = n$ è sempre più raro. Ma la (3) "vede" che quell'evento, per quanto raro, è così "disastroso" (un errore di n) da mantenere la media dell'errore alta.

Remark 0.15. *C'è una parziale freccia inversa tra la convergenza in probabilità e quella in mediar-esima. Intutivamente la convergenza in probabilità non implica la convergenza in media r-esima perché può capitare che con probabilità sempre più piccola vengano assunti valori sempre più grandi, come nel contro-esempio precedente; se però imponiamo che le variabili siano limitate, la possibilità di questi contro-esempi cade. In aprticolare, abbiamo il Lemma che segue.*

Lemma 0.16. *Sia* $\{X_n\}$ *una successione di variabili aleatorie limitate* $|X_n| \leq M \in \mathbb{R}$ *definite sullo spazio di probabilità* $\{\Omega, \mathfrak{S}, \mathbb{P}\}$ *e tale per cui* $X_n \to_p X$ *(X è evidentemente definita sullo stesso spazio di probabilità; si noti che* $|X| \leq M$ *con probabilità 1). Allora* $X_n \to_r X$ *per ogni* $r \in \mathbb{R}$.

Dimostrazione. Abbiamo che

$$\begin{aligned} E|X_n - X|^r &= E|X_n - X|^r I_{\{|X_n - X| > \epsilon\}} + E|X_n - X|^r I_{\{|X_n - X| \leq \epsilon\}} \\ &\leq (2M)^r E[I_{\{|X_n - X| > \epsilon\}}] + \epsilon^r = (2M)^r Pr\{|X_n - X| > \epsilon\} + \epsilon^r, \end{aligned}$$

e la dimostrazione è conclusa per l'arbitrarietà di ϵ e l'ipotesi sulla convergenza in probabilità. □

Example 0.17 (Legge debole dei grandi numeri)**.** *Sia* $\{X_n\}_{n\in\mathbb{N}}$ *una successione di variabili aleatorie indipendenti ed identicamente distribuite definita sullo spazio di probabilità* $\{\Omega, \mathfrak{S}, \mathbb{P}\}$*, con* $\mathbb{E}[X] = \mu$ *e* $\mathbb{E}[(X-\mu)^2] = \sigma^2 < \infty$.*. Allora*

$$\overline{X}_n := \frac{1}{n}\sum_{i=1}^{n} X_i \to_2 \mu.$$

Infatti

$$\begin{aligned}\mathbb{E}[(\overline{X}_n - \mu)^2] &= Var[\frac{1}{n}\sum_{i=1}^{n} X_i] \\ &= \frac{1}{n^2}\sum_{i=1}^{n} Var[X_i] \\ &= \frac{\sigma^2}{n}\end{aligned}$$

Spiegazione Semplice

La Legge Debole dei Grandi Numeri (che nel Capitolo 1 abbiamo visto come conseguenza di Chebyshev) è un esempio di convergenza in media quadratica (e quindi anche in probabilità).
Lo scarto quadratico medio della media campionaria $\overline{X}_n$ dalla media vera μ è $\mathbb{E}[(\overline{X}_n - \mu)^2] = \frac{\sigma^2}{n}$.
Dato che $\lim_{n\to\infty} \frac{\sigma^2}{n} = 0$, abbiamo la convergenza in media quadratica ($r = 2$).

Remark 0.18. *Le ipotesi della legge debole dei grandi numeri possono essere grandemente generalizzate. Ad esempio, si può sostituire l'indipendenza con l'incorrelazione, e l'identica distribuzione con l'ipotesi che il valor medio sia costante e la varianza uniformemente limitata.*

Remark 0.19. *E' importante ricordare la relazione che esiste tra la convergenza in media di ordini diversi.*

Lemma 0.20. *Sia* $\{X_n\}_{n\in\mathbb{N}}$ *una successione di variabili aleatorie definita sullo spazio di probabilità* $\{\Omega, \mathfrak{S}, \mathbb{P}\}$*; sia inoltre X,* $X : \Omega \to \mathbb{R}$ *una variabile aleatoria definita sullo stesso spazio di probabilità. Allora* $\{X_n \to_{r_2} X\} \Rightarrow \{X_n \to_{r_1} X\}$ *per ogni* $0 < r_1 < r_2$.

Dimostrazione. Per la dimostrazione, basta considerare la funzione convessa $f(x) := |x|^{r_2/r_1}$ e notare che, per la disuguaglianza di Jensennn

$$f(\mathbb{E}|X_n - X|^{r_1}) = (\mathbb{E}|X_n - X|^{r_1})^{r_2/r_1} \le \mathbb{E}[f(|X_n - X|^{r_1})] = (\mathbb{E}|X_n - X|^{r_2})$$

e quindi

$$(\mathbb{E}|X_n - X|^{r_1})^{1/r_1} \le (\mathbb{E}|X_n - X|^{r_2})^{1/r_2},$$

$$\{\lim_{n\to\infty} \mathbb{E}|X_n - X|^{r_2} = 0\} \Rightarrow \{\lim_{n\to\infty} \mathbb{E}|X_n - X|^{r_1} = 0\}.$$

□

Spiegazione Semplice

La convergenza in media quadratica ($r = 2$) è più forte della convergenza in media ($r = 1$). Se vale la prima, vale anche la seconda. In generale, più r è alto, più la convergenza è "forte".

Definition 0.21 (((4). Convergenza quasi certa))**.** *Sia* $\{X_n\}_{n\in\mathbb{N}}$ *una successione di variabili aleatorie definita sullo spazio di probabilità* $\{\Omega, \mathfrak{S}, \mathbb{P}\}$*; sia inoltre X,* $X : \Omega \to \mathbb{R}$ *Runa variabile aleatoria definita sullo stesso spazio di probabilità. Diciamo che* X_n *converge quasi certamente a X, scritto* $X_n \to_{q.c.} X$*, se e solo se*

$$\mathbb{P}\{\omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\} = \mathbb{P}\{\lim_{n\to\infty} X_n = X\} = 1.$$

Spiegazione Semplice

Cosa dice in parole povere? Questa è la forma più "forte" di convergenza. È la più intuitiva ma la più difficile da dimostrare.

- ω **(omega):** Pensa a ω come a un "universo" o un "esperimento completo". È un singolo, infinito lancio di monete, per esempio.
- $X_n(\omega)$**:** È il valore della variabile X_n in quello specifico universo (es. la media dei primi n lanci *di quella specifica sequenza infinita*).
- **La formula:** $\mathbb{P}\{\omega : \lim X_n(\omega) = X(\omega)\} = 1$ si legge: "La probabilità di (ri)trovarsi in un universo ω dove la successione di numeri $X_n(\omega)$ converge *esattamente* al numero $X(\omega)$ (nel senso classico dell'Analisi 1) è 1 (cioè è il 100%)".

Ignora un set di "universi sfortunati" (es. una sequenza di infinite "Teste") che hanno probabilità 0 di accadere. Per tutti gli altri (probabilità 1), la successione X_n converge a X.
La Legge *Forte* dei Grandi Numeri usa questa convergenza.

Remark 0.22. *Notiamo che l'evento $\{\omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\}$ può essere scritto come*

$$\{\omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\} = \cap_{k=1}^{\infty} \cup_{n=1}^{\infty} \cap_{m=n}^{\infty} \{\omega : |X_m(\omega) - X(\omega)| < \frac{1}{k}\}$$

cioè l'insieme di quegli ω tali per cui, per scelto $\frac{1}{k}$ piccolo quanto si vuole, $|X_m(\omega) - X(\omega)| < \frac{1}{k}$ definitivamente. Possiamo quindi definire la convergenza quasi certa imponendo che il complementare di questo evento abbia probabilità nulla, cioè

$$\begin{aligned}0 &= \mathbb{P}\left(\left[\cap_{k=1}^{\infty} \cup_{n=1}^{\infty} \cap_{m=n}^{\infty} \{\omega : |X_m(\omega) - X(\omega)| < \frac{1}{k}\}\right]^c\right) \\ &= \mathbb{P}\left(\cup_{k=1}^{\infty} \cap_{n=1}^{\infty} \cup_{m=n}^{\infty} \{\omega : |X_m(\omega) - X(\omega)| \geq \frac{1}{k}\}\right) \\ &= \lim_{n\to\infty} \mathbb{P}\left(\cup_{m=n}^{\infty} \{\omega : |X_m(\omega) - X(\omega)| \geq \frac{1}{k}\}\right) \quad \textit{per ogni } k = 1, 2, ...\end{aligned}$$

Da quest'ultima riga è immediato vedere che la convergenza quasi certa implica la convergenza debole, (4) $\Rightarrow$ (2).

Spiegazione Semplice

Se la successione X_n converge a X "quasi certamente" (cioè, tranne che in casi sfortunati con probabilità 0), allora la probabilità che X_n e X siano lontani deve per forza andare a zero.
La (4) è più forte della (2).

Remark 0.23. *La convergenza debole al contrario non implica la convergenza quasi certa. Un controesempio può essere costruito come segue: sia U una variabile aleatoria uniforme in [0,1], definita su uno spazio di probabilità adeguato. Consideriamo la sequenza $\{X_n\}_{n\in\mathbb{N}}$ definita come*

$$\begin{aligned}X_1 &= \mathbb{I}_{[0,\frac{1}{2})}(U), \quad X_2 = \mathbb{I}_{[\frac{1}{2},1]}(U) \\ X_3 &= \mathbb{I}_{[0,\frac{1}{4})}(U), \quad X_4 = \mathbb{I}_{[\frac{1}{4},\frac{2}{4})}(U), \quad X_5 = \mathbb{I}_{[\frac{2}{4},\frac{3}{4})}(U), \quad X_6 = \mathbb{I}_{[\frac{3}{4},1]}(U) \dots \\ X_7 &= \mathbb{I}_{[0,\frac{1}{8})}(U), \quad X_8 = \mathbb{I}_{[\frac{1}{8},\frac{2}{8})}(U), \dots, X_{14} = \mathbb{I}_{[\frac{7}{8},1]}(U), \\ X_{15} &= \mathbb{I}_{[0,\frac{1}{16})}(U), \quad X_{16} = \mathbb{I}_{[\frac{1}{16},\frac{2}{16})}(U), \dots, X_{30} = \mathbb{I}_{[\frac{15}{16},1]}(U), \\ X_{2^q-1} &= \mathbb{I}_{[0,\frac{1}{2^q})}(U), \quad X_{2^q} = \mathbb{I}_{[\frac{1}{2^q},\frac{2}{2^q})}(U), \dots\end{aligned}$$

Si ha che

$$\lim_{n\to\infty} Pr\{|X_n| > 0\} = 0, \quad \lim_{n\to\infty} Pr\{\cup_{m=n}^{\infty} |X_m| > 0\} = 1,$$

quindi la sequenza converge in probabilità a zero, anche se assume infinitamente spesso il valore 1 e pertanto non può convergervi quasi certamente.

Spiegazione Semplice

Contro-esempio: (2) NON implica (4) Questo è il contro-esempio più famoso. Immagina una "luce" che "spazza" l'intervallo [0,1].

- Prima X_1 illumina [0, 1/2], X_2 illumina [1/2, 1].
- Poi X_3 illumina [0, 1/4], X_4 illumina [1/4, 2/4], X_5 illumina [2/4, 3/4], ...
- Poi con intervalli di 1/8, 1/16, e così via.
- X_n vale 1 se il punto U è "illuminato", 0 altrimenti.
- **Converge in Probabilità (2) a 0?** Sì. Per ogni n grande, la "luce" è strettissima (es. larga 1/1000). La probabilità di essere "illuminati" ($\mathbb{P}(|X_n| > 0)$) tende a 0.
- **Converge Quasi Certo (4) a 0?** No. Qualsiasi punto U tu scelga (es. $U = 0.5$), la "luce" continuerà a spazzare e a "colpirlo" *infinite volte*. La successione $X_n(0.5)$ sarà 0, 0, 1, 0, 0, 1, 0, 0, 1... e **non converge** a 0.

Definition 0.24 (((5). Convergenza completa))**.** *Sia* $\{X_n\}_{n\in\mathbb{N}}$ *una successione di variabili aleatorie definita sullo spazio di probabilità* $\{\Omega, \mathfrak{S}, \mathbb{P}\}$*; sia inoltre* X*,* $X : \Omega \to \mathbb{R}$ *Runa variabile aleatoria definita sullo stesso spazio di probabilità. Diciamo che* X_n *converge completamente a* X*, scritto* $X_n \to_{c.c.} X$*, se e solo se*

$$\sum_{n=1}^{\infty} \mathbb{P}\{|X_n - X| > \epsilon\} < \infty, \quad \textit{per ogni } \epsilon > 0.$$

Spiegazione Semplice

Questa è una forma di convergenza super-forte, usata per lo più nelle dimostrazioni. Non chiede solo che $\mathbb{P}\{|X_n - X| > \epsilon\}$ tenda a 0, ma che tenda a 0 *così velocemente* che la sua serie (la somma infinita di tutti i termini) *converga* a un numero finito (come fa $\sum 1/n^2$).

Remark 0.25. *La convergenza completa implica quella quasi certa* ((5) $\Rightarrow$ (4))*; infatti, per la subadditività della misura di probabilità abbiamo che*

$$\lim_{n\to\infty} \mathbb{P}(\cup_{m=n}^{\infty}\{\omega : |X_m(\omega) - X(\omega)| \geq \frac{1}{k}\}) \leq \lim_{n\to\infty} \sum_{m=n}^{\infty} \mathbb{P}(\{\omega : |X_m(\omega) - X(\omega)| \geq \frac{1}{k}\}) = 0,$$

perchè il resto n-esimo di una serie convergente va a zero. Il vice versa non è vero: consideriamo ad esempio una variabile uniforme in [0,1] U, ed introduciamo la sequenza

$$X_n := \mathbb{I}_{[0,\frac{1}{n}]}(U).$$

Chiaramente

$$\sum_{n=1}^{N} Pr\{X_n > 0\} = \sum_{n=1}^{N} \frac{1}{n} \to \infty \quad \textit{per } N \to \infty$$

d'altra parte però per ogni ω *tale che* $U(\omega) \neq 0$ *abbiamo* $\lim_{n\to\infty} \mathbb{I}_{[0,\frac{1}{n}]}(U) = 0$.

Remark 0.26. *Come abbiamo visto, in generale la convergenza in probabilità è molto più debole della convergenza quasi certa, e quindi a maggior ragione di quela completa. E' comunque possibile trovare una parziale controimplicazione, come nel Lemma che segue.*

Lemma 0.27. *Sia* $\{X_n\}_{n\in\mathbb{N}}$ *una successione di variabili aleatorie definita sullo spazio di probabilità* $\{\Omega, \mathfrak{S}, \mathbb{P}\}$*; sia inoltre* X*,* $X : \Omega \to \mathbb{R}$ *Runa variabile aleatoria definita sullo stesso spazio di probabilità, e si abbia la convergenza in probabilità* $X_n \to_p X$*. Allora esiste una sottosuccessione* X_{n_k} *tale che* $X_{n_k} \to_{c.c.} X$*.*

Dimostrazione. Per la dimostrazione, è sufficiente scegliere la sottosuccessione X_{n_k} tale per cui

$$Pr\{|X_{n_k} - X| > \frac{1}{k}\} \leq \frac{1}{2^k}$$

in modo che si abbia, per ogni $\epsilon > 0$

$$\sum_{n_k=1}^{\infty} Pr\{|X_{n_k} - X| > \epsilon\} \leq \sum_{n_k=1}^{\lceil \frac{1}{\epsilon} \rceil} Pr\{|X_{n_k} - X| > \epsilon\} + \sum_{n_k=\lceil \frac{1}{\epsilon} \rceil+1}^{\infty} Pr\{|X_{n_k} - X| > \frac{1}{k}\}$$

$$\leq \lceil \frac{1}{\epsilon} \rceil + \sum_{n_k=1}^{\infty} \frac{1}{2^k} < \infty.$$

□

Remark 0.28. *Si può verificare che non esiste implicazione, né in un senso né nell'altro, tra la convergenza completa e la convergenza in media r-esima. Si consideri infatti la sequenza:*

$$X_n = \begin{cases} 0 & \text{con probabilità } 1 - \frac{1}{n^2} \\ n & \text{con probabilità } \frac{1}{n^2} \end{cases}$$

Le X_n convergono completamente (e quindi quasi certamente) alla variabile aleatoria che vale identicamente zero, ma non convergono nemmeno in media prima; infatti $\mathbb{E}[X_n] = 1$ per ogni n.

Spiegazione Semplice

Questo Remark mostra che non c'è legame tra la convergenza forte (completa) e la convergenza in media (r-esima). Dà un esempio: $X_n = n$ con probabilità $1/n^2$, e 0 altrimenti.

- **Converge completamente (5)?** Sì. La somma delle probabilità $\sum \mathbb{P}(|X_n| > \epsilon) = \sum 1/n^2$ è una serie convergente (fa $\pi^2/6$).
- **Converge in media r=1 (3)?** No. Il testo afferma che $\mathbb{E}[X_n] = 1$ per ogni n (quindi non tende a 0). (Nota: c'è un'apparente discrepanza tra l'esempio $X_n = n$ con prob $1/n^2$ che darebbe $\mathbb{E}[X_n] = 1/n$, e l'affermazione $\mathbb{E}[X_n] = 1$. L'esempio del Remark 30, $X_n = n$ con prob $1/n$, ha $\mathbb{E}[X_n] = 1$. La conclusione del Remark 42 si basa sul fatto che l'aspettativa non converge a 0).

Abbiamo quindi stabilito le implicazioni

$$(5) \Rightarrow (4) \Rightarrow (2) \Rightarrow (1),$$
$$(3) \Rightarrow (2) \Rightarrow (1);$$

concludiamo mostrando una implicazione tra (1) e (4).

Spiegazione Semplice

Riassunto delle Implicazioni
Ecco la gerarchia delle convergenze:

- **Completa (5)** $\Longrightarrow$ **Quasi Certa (4)** $\Longrightarrow$ **In Probabilità (2)** $\Longrightarrow$ **In Legge (1)**
- **In Media r-esima (3)** $\Longrightarrow$ **In Probabilità (2)**

Nessun'altra freccia è vera in generale! Ad esempio, la "quasi certa" (4) non implica la "media r-esima" (3), e viceversa.

Proposition 0.29 (Skorohod)**.** *Sia $\{X_n\}_{n\in\mathbb{N}}$ una successione di variabili aleatorie definita sullo spazio di probabilità tale per cui $X_n \to_d X$. Allora esiste uno spazio di probabilità $\{\Omega, \mathfrak{S}, \mathbb{P}\}$ su cui sono definite variabili aleatorie X'_n, X' tali per cui $X_n \underline{d} X'_n$, $X \underline{d} X'$, e*

$$\mathbb{P}\{\lim_{n\to\infty} X'_n = X'\} = 1.$$

Spiegazione Semplice

Questo è un teorema molto tecnico ma affascinante. Dice che anche se la convergenza in legge (1) è debolissima, possiamo *sempre* "ricostruire" le variabili X_n e X su un nuovo spazio di probabilità "magico" (chiamiamole X'_n e X') in modo tale che:

1. Le nuove variabili X'_n e X' abbiano le stesse identiche distribuzioni di quelle vecchie (X_n e X).
2. Su questo nuovo spazio, le variabili X'_n convergano a X' nel senso più forte possibile, quello "quasi certo" (4).

È uno strumento potentissimo che permette ai matematici di "trasformare" una convergenza debole in una forte, per facilitare le dimostrazioni.

Dimostrazione. Per semplicità consideriamo il caso in cui le funzioni di distribuzione $F_{X_n}(.)$ $F_X(.)$ siano crescenti e continue. Prendiamo $\{\Omega, \mathfrak{A}, \mathbb{P}\} = \{[0,1], \mathbb{B}[0,1], Leb\}$ con la variabile aleatoria identità $U(\omega) = \omega$, cioè l'uniforme in $[0, 1]$. Definiamo

$$X'_n = F_{X_n}^{-1}(U), \quad X' = F_X^{-1}(U);$$

queste inverse sono ben poste per le ipotesi sulla funzione di ripartizione ed il risultato segue immediatamente perchè

$$Pr(X'_n \leq x) = Pr(F_{X_n}^{-1}(U) \leq x) = Pr(U \leq F_{X_n}(x)) = F_{X_n}(x),$$

e similmente per X'. La convergenza quasi certa è una conseguenza della convergenza (quasi) ovunque delle funzioni di distribuzione e delle loro inverse. $\square$