3 Stochastic Orders of Magnitude

In molti casi può essere utile avere alcuni strumenti per studiare il comportamento di variabili aleatorie che non convergono in nessuno dei sensi discusso nel capitolo precedente, ma nonstante questo rimagono limitate, dopo essere state opportunamente normalizzate. A questo fine, introduciamo la seguente

Definition 44 (Stochastic Orders of Magnitude: $O_p(.) e o_p(.)$)**.** *Sia $\{X_n\}_{n\in\mathbb{N}}$ una successione di variabili aleatorie e sia $\{f_n\}_{n\in\mathbb{N}}$ una successione deterministica, $f_n > 0$ per ogni n. Diremo che $X_n = O_p(f_n)$ se e solo se*

$$\forall\epsilon > 0 \quad \exists K \in \mathbb{R}, \exists n_0 \in \mathbb{N} : \forall n > n_0 \quad Pr\{\frac{|X_n|}{f_n} \leq K\} \geq 1 - \epsilon.$$

Diremo che $X_n = o_p(f_n)$ se e solo se

$$\forall\epsilon > 0, \forall\delta > 0, \exists n_0 \in \mathbb{N} : \forall n > n_0 \quad Pr\{\frac{|X_n|}{f_n} \leq \delta\} \geq 1 - \epsilon.$$

Spiegazione Semplice

Cosa significano "O grande P" e "o piccolo P"?
Queste notazioni sono un modo per descrivere quanto "grande" è una variabile aleatoria X_n man mano che n cresce. È l'equivalente in probabilità della notazione "O grande" e "o piccolo" che si usa in Analisi Matematica (es. per i limiti notevoli o in informatica per la complessità degli algoritmi).

- $X_n = O_p(f_n)$ (**O grande P**): X_n cresce **al massimo** come f_n. Dividendo X_n per f_n, il risultato $\frac{|X_n|}{f_n}$ non "esplode", ma rimane "limitato in probabilità" (cioè, con probabilità molto alta, sta sotto un qualche numero K).
- $X_n = o_p(f_n)$ (**o piccolo P**): X_n cresce **più lentamente** di f_n. Dividendo X_n per f_n, il risultato $\frac{X_n}{f_n}$ "collassa" a zero (converge in probabilità a 0).

Esempi comuni:

- $X_n = O_p(1)$: Si usa $f_n = 1$. Significa che X_n è **stocasticamente limitata**. Non esplode all'infinito. È "controllata".
- $X_n = o_p(1)$: Si usa $f_n = 1$. Significa che X_n **converge in probabilità a 0**. È un modo più compatto per scrivere $X_n \to_p 0$.

Remark 45. *Dalla definizione, segue immediatamente che $\{X_n = o_p(f_n)\} \Leftrightarrow \{\frac{X_n}{f_n} \to_p 0\}$.*

Remark 46. *Nella definizione di $X_n = O_p(f_n)$ avremmo potuto rimpiazzare la condizione $\exists n_0 \in \mathbb{N} : \forall n > n_0$ con la condizione (apparentemente più generale) $\forall n \in \mathbb{N}$. Infatti se la sequenza $\frac{|X_n|}{f_n}$ è limitata per n sufficientemente grande è sempre possibile scegliere una costante K' (magari più grande di K) per la quale l'affermazione $Pr\{\frac{|X_n|}{f_n} \leq K'\} \geq 1 - \epsilon$ sia soddisfatta con $n = 1, ..., n_0$.*

Remark 47. *E' evidente che $\{X_n = o_p(f_n)\} \Rightarrow \{X_n = O_p(f_n)\}$. Sul piano discorsivo, $O_p(.)$ implica che la sequenza normalizzata X_n/f_n appartiene ad un compatto; se la sequenza X_n/f_n converge è zero, appartiene ad un compatto arbitrariamente piccolo intorno all'origine, per n grande abbastanza.*

Proposition 48 (Algebra degli stochastic orders)**.** *Siano $\{f_n, g_n\}$ due sequenze deterministiche strettamente positive. 1) Abbiamo che*

$$\{X_n = O_p(f_n), Y_n = O_p(g_n)\} \Rightarrow \{X_n + Y_n = O_p(\max(f_n, g_n))\}$$

$$\{X_n = O_p(f_n), Y_n = O_p(g_n)\} \Rightarrow \{X_n \times Y_n = O_p(f_n \times g_n)\}$$

2) Inoltre

$$\{X_n = o_p(f_n), Y_n = o_p(g_n)\} \Rightarrow \{X_n + Y_n = o_p(\max(f_n, g_n))\}$$

$$\{X_n = o_p(f_n), Y_n = o_p(g_n)\} \Rightarrow \{X_n \times Y_n = o_p(f_n \times g_n)\}$$

3) Infine

$$\{X_n = O_p(f_n), Y_n = o_p(g_n)\} \Rightarrow \{X_n \times Y_n = o_p(f_n \times g_n)\}$$

Spiegazione Semplice

Come si comportano "O grande P" e "o piccolo P" in algebra?
Questa proposizione ci dà le regole per sommare e moltiplicare questi ordini. Sono le stesse dell'Analisi Matematica:

- **Somma (Regola 1 e 2):** Quando sommi due termini, quello che "vince" (cioè quello che cresce più velocemente) detta l'ordine di grandezza. Esempio: $O_p(n^2) + O_p(n) = O_p(n^2)$.
- **Prodotto (Regola 1 e 2):** Quando moltiplichi, gli ordini si moltiplicano. Esempio: $O_p(n^2) \times O_p(n) = O_p(n^3)$.
- **Regola mista (Regola 3):** Questa è la più importante. Se moltiplichi qualcosa di *limitato* (un O_p) per qualcosa che *va a zero* (un o_p), il risultato *va a zero* (o_p).

Dimostrazione. Per la dimostrazione del punto 1), fissato $\epsilon > 0$ prendiamo due costanti C_1 C_2 sufficientemente grandi tali per cui

$$Pr\{\frac{|X_n|}{f_n} > C_1\} < \frac{\epsilon}{2}$$

e

$$Pr\{\frac{|Y_n|}{g_n} > C_2\} < \frac{\epsilon}{2}$$

l'esistenza di tali costanti è garantita dall'ipotesi $X_n = O_p(f_n)$, $Y_n = O_p(g_n)$. Abbiamo evidentemente

$$Pr\{|X_n + Y_n| > (C_1 \vee C_2)(f_n \vee g_n)\}$$

$$\leq Pr\{|X_n| > C_1 f_n\} + Pr\{|Y_n| > C_2 g_n\} = \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon,$$

dove la prima disuguaglianza segue dal fatto che l'evento $\{|X_n + Y_n| > (C_1 \vee C_2)(f_n \vee g_n)\}$ implica necessariamente il verificarsi di almeno uno dei due eventi $\{|X_n| > C_1 f_n\}$ $\{|Y_n| > C_2 g_n\}$. In maniera analoga possiamo notare che

$$Pr\{|X_n \times Y_n| > (C_1 \times C_2)(f_n \times g_n)\}$$

$$\leq Pr\{|X_n| > C_1 f_n\} + Pr\{|Y_n| > C_2 g_n\} = \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon,$$

di nuovo perché l'evento $\{|X_n \times Y_n| > (C_1 \times C_2)(f_n \times g_n)\}$ implica necessariamente il verificarsi di almeno uno dei due eventi $\{|X_n| > C_1 f_n\} o \{|Y_n| > C_2 g_n\}$. La dimostrazione del punto 2) è del tutto analoga. Per il punto 3), per ipotesi abbiamo $X_n = O_p(f_n)$ $Y_n = o_p(g_n)$; notiamo che si ha, per qualsiasi costante $C_1 > 0$

$$Pr\{|X_n \times Y_n| > C_1(f_n \times g_n)\}$$

$$= Pr\{|X_n \times Y_n| > (\frac{C_1}{C_2} \times C_2)(f_n \times g_n)\}$$

$$\leq Pr\{|X_n| > C_2 f_n\} + Pr\{|Y_n| > \frac{C_1}{C_2} g_n\}.$$

Per ipotesi, possiamo scegliere C_2 tale per cui $Pr\{|X_n| > C_2 f_n\} \leq \frac{\epsilon}{2}$; sempre per ipotesi, per $\frac{C_1}{C_2}$ fissato esiste senz'altro n_0 tale che

$$Pr\{|Y_n| > \frac{C_1}{C_2} g_n\} \leq \frac{\epsilon}{2}$$

per ogni $n > n_0$, ed il risultato è dimostrato. □

E' naturale domandarsi come si inseriscano gli stochastic orders of magnitude nella gerarchia di modalità di convergenza che abbiamo discusso nella sezione precedente. Informalmente, possiamo considerare una sequenza $X_n = O_p(1)$ come se fosse nella categoria "0", nel senso che tale proprietà è implicata da tutte le modalità di convergenza precedentemente introdotte:

$$(5) \Rightarrow (4) \Rightarrow (2) \Rightarrow (1) \Rightarrow O_p(1),$$

$$(3) \Rightarrow (2) \Rightarrow (1) \Rightarrow O_p(1);$$

più rigorosamente, possiamo dimostrare il risultato qui di seguito.

Lemma 49. *Sia X_n una sequenza di variabili aleatorie che converge in distribuzione ad una variabile limite X. Allora $X_n = O_p(1)$*

> **Spiegazione Semplice**
>
> **Convergenza $\implies$ Limitatezza Stocastica**
> Questo lemma collega la convergenza del Capitolo 2 a questa nuova notazione.
> Dice che se una successione di variabili aleatorie X_n converge in distribuzione (la forma più debole di convergenza, $X_n \to_d X$), allora deve essere "stocasticamente limitata" ($X_n = O_p(1)$).
> In parole povere: se una successione di istogrammi X_n sta convergendo a un istogramma "finale" X, questi istogrammi non possono "scappare" all'infinito. Devono rimanere "controllati" e "limitati" (con alta probabilità) per potersi stabilizzare.

Dimostrazione. Notiamo prima che per ogni $\epsilon > 0$ esiste sicuramente un compatto $K = K_\epsilon \subset \mathbb{R}$ tale per cui $Pr\{X \in K_\epsilon\} > 1 - \epsilon$. E' sufficiente infatti prendere $K_\epsilon = [-M_\epsilon, M_\epsilon]$, dove $M_\epsilon > 0 \in \mathbb{R}$ tale per cui $F_X(-M_\epsilon) < \frac{\epsilon}{2}$ e $F_X(M_\epsilon) > 1 - \frac{\epsilon}{2}$; questo M_ϵ esiste senz'altro perché tutte le funzioni di distribuzione devono soddisfare $\lim_{x\to-\infty} F_X(x) = 0$, $\lim_{x\to\infty} F_X(x) = 1$ (questa proprietà è talvolta chiamata la regolarità delle misure di probabilità). Preso ora $\bar{\epsilon} > 0$, scegliamo $M_{\epsilon/2}$ punto di continuità di F_X tale per cui

$$Pr\{X \in [-M_{\epsilon/2}, M_{\epsilon/2}]\} > 1 - \frac{\epsilon}{2}$$

si scelga ora n_0 tale per cui $|F_{X_n}(-M_{\epsilon/2}) - F_X(-M_{\epsilon/2})| \leq \frac{\epsilon}{4}$ e $|F_{X_n}(M_{\epsilon/2}) - F_X(M_{\epsilon/2})| \leq \frac{\epsilon}{4}$ per tutti gli $n > n_0$; allora per un tale n abbiamo che

$$\begin{gathered} Pr\{|X_n| > M_{\epsilon/2}\} \\ = F_{X_n}(-M_{\epsilon/2}) + 1 - F_{X_n}(M_{\epsilon/2}) \\ \leq |F_{X_n}(-M_{\epsilon/2}) - F_X(-M_{\epsilon/2})| + F_X(-M_{\epsilon/2}) \\ +1 - F_X(M_{\epsilon/2}) + |F_X(M_{\epsilon/2}) - F_{X_n}(M_{\epsilon/2})| \\ \leq \frac{\epsilon}{4} + (F_X(-M_{\epsilon/2}) + 1 - F_X(M_{\epsilon/2})) + \frac{\epsilon}{4} \leq \frac{\epsilon}{4} + \frac{\epsilon}{2} + \frac{\epsilon}{4} = \epsilon, \end{gathered}$$

da cui segue immediatamente che $X_n = O_p(1)$, come richiesto. □

Example 50. *E' interessante produrre alcuni esempi in cui la condizione $X_n = O_p(1)$ non vale; tre particolarmente semplici sono le sequenze*

$$X_n = n, \quad X_n \triangleq N(0, n), \quad X_n \triangleq U(-n, n).$$

Nel corso di Analisi 1 si è visto, come importante corollario del Teorema di Bolzano-Weierstrass, che ogni successione limitata ammette sempre una sottosuccessione (estratta) convergente. La nozione di sequenza $O_p(1)$ si può interpretare come una forma di equilimitatezza stocastica (detta tipicamente tightness) nella letteratura internazionale; è naturale domandarsi se possa esistere una proprietà simile per successioni di variabili aleatorie. La risposta (affermativa) è fornita dal seguente Teorema.

Theorem 51 (Prohorov 1956)**.** *Ogni successione $X_n = O_p(1)$ ammette una estratta che converge in distribuzione.*

> **Spiegazione Semplice**
>
> **Il "Bolzano-Weierstrass" della Probabilità**
> Questo teorema è l'esatto analogo stocastico (probabilistico) del famoso Teorema di Bolzano-Weierstrass.
>
> - **Teorema di Bolzano-Weierstrass:** Una successione *limitata* (es. $x_n = \sin(n)$) ha sempre una sotto-successione *convergente.*
> - **Teorema di Prohorov:** Una successione *stocasticamente limitata* (cioè $X_n = O_p(1)$) ha sempre una sotto-successione *convergente in distribuzione.*

Ci dice che se una sequenza di variabili aleatorie non "esplode", allora una sua "parte" (X_{n_k}) deve per forza "stabilizzarsi" verso una qualche distribuzione.

Remark 52. *Il teorema di Prohorov è in realtà molto più generale di questo, perché vale per qualsiasi successione tight di elementi aleatori a valori su uno spazio completo, metrico e separabile.*

Dimostrazione. Dal Teorema di Helly-Bray (si veda ad esempio....) sappiamo che ogni successione di funzioni di distribuzione ammette una estratta convergente ad una funzione di distribuzione possibilmente difettiva, cioè una funzione F^* che sia non decrescente e continua da destra, ma per la quale non valga necessariamente $\lim_{x\to\infty} F^*(-x) = 0$ e $\lim_{x\to\infty} F^*(x) = 1$. Sia F_{n_k} questa estratta, e dimostriamo che se la corrispondente sequenza di variabile aleatorie è tight necessariamente la funzione limite F^* deve indurre una misura di massa 1 (cioè deve valere $\lim_{x\to\infty}(F^*(x) - F^*(-x)) = 1$). Poiché F^* è non decrescente, il limite $\lim_{x\to\infty} F^*(x) = \eta$ sicuramente esiste; supponiamo per assurdo che sia strettamente minore di 1. Allora per ogni insieme compatto $K \subset \mathbb{R}$ si avrebbe necessariamente $Pr\{X_n \in K\} \leq Pr\{X_n \leq \max(K)\} \leq \eta < 1$ e la sequenza non potrebbe essere $O_p(1)$. Il teorema di Levy ci garantisce che una successione di funzioni di ripartizione individua una variabile aleatoria limite (in distribuzione) se e solo se le corrispondenti funzioni caratteristiche convergono ad una funzione limite continua all'origine. Questa condizione può essere interpretata come una condizione di tightness; ad esempio, per la sequenza di Gaussiane $N(0, n)$ le corrispondenti funzioni caratteristiche sono date da $\phi_n(t) = \exp(-n\frac{t^2}{2})$ che convergono ad una funzione limitex $\phi_\infty(t) = I_{\{t=0\}}(t)$, evidentemente discontinua all'origine. $\square$