Un complemento al Paradosso di Borel

Notazioni

$\mathbb{N}$ = insieme degli interi positivi.

$\mathbb{N}^*$ = insieme degli interi **strettamente** positivi.

Premessa terminologica

Converrà forse, prima di cominciare, spendere due parole per fissar bene il significato da attribuire al termine *paradosso* (nell'ambito di una teoria matematica) e per marcare la distanza di questo termine dal termine *sofisma*.

Un sofisma è una proposizione falsa, che qualcuno presenta come vera esibendone una pseudo-dimostrazione fondata su un uso scorretto delle regole logiche che governano la teoria.

Questa manomissione delle regole può essere involontaria, cioè dovuta a una cattiva conoscenza di quelle regole, oppure intenzionale.

In quest'ultimo caso, colui che presenta il sofisma conosce perfettamente le regole, ma cerca di approfittare con scaltrezza della loro complessità e della superficiale conoscenza che ne ha il destinatario del sofisma.

In definitiva si può dunque dire che un sofisma è un inganno (oppure un autoinganno). Con una buona conoscenza della teoria, e una sufficiente dose di pazienza, è sempre possibile smontarlo, individuando il punto, o i punti, in cui le regole sono state violate.

Ben diversa è la natura del paradosso. Per paradosso s'intende infatti una proposizione vera della teoria, ma che può apparire contraria al senso comune qualora si pretenda di attribuire agli oggetti matematici che intervengono nel suo enunciato una particolare interpretazione intuitiva.

Insomma un paradosso è un *teorema* (beninteso, nel senso primigenio di questo termine, e non in quello, ormai prevalente in Italia, di "cervellotico impianto accusatorio messo in piedi da giovane procuratore in cerca di facile notorietà"). Ma un paradosso non è semplicemente un teorema, bensì un teorema che sia accompagnato da una particolare interpretazione intuitiva degli oggetti matematici coinvolti nel suo enunciato, alla luce della quale la tesi del teorema appaia in contrasto col senso comune.

Il paradosso di Borel

Veniamo ora ad enunciare il Paradosso di Borel in modo formale, cioè come semplice teorema nell'ambito della Teoria dei Processi stocastici.

Siano E un insieme numerabile e μ una misura normalizzata su $\mathcal{P}(E)$ (tribù di tutte le parti di E), la quale si annulli solo sull'insieme vuoto.

Su uno spazio probabilizzato $(\Omega, \mathcal{A}, P)$ sia $(X_n)_{n\geq 1}$ un processo, avente $\big(E, \mathcal{P}(E)\big)$ come spazio degli stati e $\mathbb{N}^*$ come insieme dei tempi, tale che le variabili aleatorie X_n siano indipendenti e che ciascuna di esse abbia μ come legge.

Fissato un intero k strettamente positivo, si chiamerà *stringa* di lunghezza k una successione finita di elementi di E, della forma $(x_j)_{1\leq j\leq k}$.

Assegnata la stringa $(x_j)_{1\leq j\leq k}$, e posto

$$A_m = \bigcap_{j=1}^{k} \{X_{m+j} = x_j\} \quad \text{per } m \in \mathbb{N},$$

l'evento A_m sarà chiamato "apparizione dell'assegnata stringa nell'intervallo temporale $[m+1, m+k]$".

L'evento

$$\bigcup_{m\in\mathbb{N}} A_m$$

sarà invece detto "apparizione dell'assegnata stringa".

Il cosiddetto *Paradosso di Borel* afferma che quest'ultimo evento è quasi certo. Si tratta di un risultato del tutto elementare, che si può far discendere da un risultato più forte (e altrettanto elementare): è quasi certo addirittura l'evento

$$\bigcup_{n\in\mathbb{N}} A_{nk},$$

costituito dall'unione dei soli eventi della successione $(A_m)_{m\in\mathbb{N}}$ il cui indice sia un multiplo di k.

Per provare quest'ultimo risultato, basta osservare che gli eventi della successione $(A_{nk})_{n\in\mathbb{N}}$ sono indipendenti e che ciascuno di essi ha probabilità eguale al numero p così definito:

$$p = \prod_{j=1}^{k} q_j \quad \text{con} \quad q_j = \mu\{x_j\}.$$

Dunque il processo $(Z_n)_{n\geq 1}$ definito da

$$Z_n = I_{A_{nk}}$$

è un processo di Bernoulli, di parametro p, sicché è quasi certo, non solo l'evento

$$\bigcup_{n\geq 1}\{Z_n = 1\},$$

ma addirittura l'evento

$$\limsup_n \{Z_n = 1\}.$$

Naturalmente è ben nota, anche fuori dalla cerchia dei matematici, l'interpretazione grazie alla quale il risultato sopra dimostrato si è guadagnato il nome di *paradosso*.

Inutile dire che si tratta di un'interpretazione molto audace. Per dirla più brutalmente: se non di pura follia, si tratta di pura celia. Infatti l'esperimento aleatorio consistente nel pigiare a caso, a ciascun istante, un tasto di una tastiera (senza mai arrestarsi) è un esperimento che nessun essere vivente (uomo o scimmia che sia) potrà mai eseguire.

Siamo insomma di fronte al solito vezzo, consistente nell'impiegare, per un problema di probabilità nel quale intervenga uno schema infinito, lo stesso linguaggio che è abituale, e naturale, nel caso di uno schema finito.

Ma, lasciando da parte ogni possibile interpretazione, è interessante occuparsi di altri problemi matematici che siano formulabili nello stesso quadro d'ipotesi del Paradosso di Borel e che siano un po' meno banali del paradosso stesso.

Uno di questi problemi consiste nello studio della variabile aleatoria S (a valori in $\mathbb{N} \cup \{\infty\}$), così definita:

$$S(\omega) = \inf \{m \in \mathbb{N} : \omega \in A_m\} .$$

Essa rappresenta il tempo che precede la prima apparizione dell'assegnata stringa. È interessante calcolarne esplicitamente la speranza.

Questo problema è proposto al lettore, come esercizio, in ciascuno dei due seguenti libri di D. Williams:

Probability with Martingales (1991),

Weighing the Odds (2001).

Come unico suggerimento, l'autore accenna, in modo molto conciso e informale, a un metodo di risoluzione, fondato sull'impiego delle martingale, del quale attribuisce la paternità al matematico cinese S.Y.R. Li, senza fornire ulteriori precisazioni.

Converrà seguire una via leggermente diversa (non facente uso delle martingale) che mi permisi di proporre a Williams e che lo stesso Williams ebbe la gentilezza di giudicare "nicer" rispetto a quella da lui suggerita.

Innanzitutto si può osservare che, grazie al Paradosso di Borel, la variabile aleatoria S è quasi certamente finita. Dunque, pur di modificare il processo $(X_n)_{n\geq 1}$ su un opportuno evento trascurabile, si può supporre che S sia dappertutto finita. Si può allora dire che, per ogni intero j compreso tra 1 e k, la variabile aleatoria X_{S+j} coincide con la costante x_j:

$$X_{S+j} = x_j \quad \text{per} \quad 1 \leq j \leq k. \tag{1}$$

Per ogni intero positivo n, si denoti con $\mathcal{F}_n$ la minima tribù che renda misurabili le X_j con $j \leq n$. Si ponga inoltre

$$T = S + k. \tag{2}$$

Allora T è, rispetto alla filtrazione $(\mathcal{F}_n)_{n\geq 0}$, un tempo d'arresto. Infatti, per ogni intero positivo n, si ha

$$\{T \leq n\} = \bigcup_{m\in\mathbb{N},\, m+k\leq n} \bigcap_{1\leq j\leq k} \{X_{m+j} = x_j\} \in \mathcal{F}_n.$$

Il tempo d'arresto T rappresenta l'istante in cui si compie la prima apparizione dell'assegnata stringa.

Poiché T differisce da S solo per l'aggiunta della costante k, basterà calcolare esplicitamente la speranza di T.

A questo scopo, per ogni intero j compreso tra 1 e k, si denoti con f_j la funzione così definita su E:

$$f_j = q_j^{-1} I_{\{x_j\}} \quad \text{con} \quad q_j = \mu\{x_j\}. \tag{3}$$

Sia n un intero positivo. Per ogni intero j compreso tra 1 e k, la variabile aleatoria $f_j \circ X_{n+j}$ ha speranza unitaria, come mostra la relazione

$$\begin{aligned} \int f_j \circ X_{n+j} \,\mathrm{d}P &= q_j^{-1} \int I_{\{x_j\}} \circ X_{n+j} \,\mathrm{d}P \\ &= q_j^{-1} P\{X_{n+j} = x_j\} = 1. \end{aligned}$$

Ne segue che ha speranza unitaria anche la variabile aleatoria U_n così definita:

$$U_n = \prod_{j=1}^{k} f_j \circ X_{n+j}. \tag{4}$$

L'evento $\{T > n\}$, essendo un elemento della tribù $\mathcal{F}_n$, è indipendente dal blocco $[X_{n+j}]_{j\geq 1}$, e quindi da U_n. Si ha dunque

$$P\{T > n\} = \int_{\{T>n\}} U_n \,\mathrm{d}P. \tag{5}$$

Questa relazione, poiché la variabile aleatoria U_n è nulla sull'evento $\{S > n\}$, identico a $\{T > n + k\}$, può essere scritta nella forma

$$
\begin{aligned}
P\{T > n\} &= \textstyle\int_{\{n<T\leq n+k\}} U_n \, \mathrm{d}P \\
&= \textstyle\sum_{h=1}^{k} \int_{\{T=n+h\}} U_n \, \mathrm{d}P.
\end{aligned} \tag{6}
$$

D'altra parte, la definizione (4) di U_n mostra che, sull'evento $\{T = n + h\}$ (identico a $\{S + k - h = n\}$), la variabile aleatoria U_n coincide con la variabile aleatoria $\prod_{j=1}^{k} f_j \circ X_{S+k-h+j}$, la quale, a sua volta, coincide (in virtù di (1)), con la costante $\prod_{j=1}^{k} f_j(x_{k-h+j})$. Dunque, ponendo

$$
c_h = \textstyle\prod_{j=1}^{k} f_j(x_{k-h+j}), \tag{7}
$$

si potrà scrivere la relazione (6) nella forma

$$
P\{T > n\} = \textstyle\sum_{h=1}^{k} c_h \, P\{T = n + h\}. \tag{8}
$$

Il risultato così ottenuto basta per fornire la seguente espressione esplicita della speranza di T:

$$\begin{aligned}
\int T \, \mathrm{d}P &= \textstyle\sum_{n \geq 0} P\{T > n\} \\
&= \textstyle\sum_{n \geq 0} \sum_{h=1}^{k} c_h \, P\{T = n + h\} \\
&= \textstyle\sum_{h=1}^{k} c_h \sum_{n \geq 0} P\{T = n + h\} \\
&= \textstyle\sum_{h=1}^{k} c_h \, P\{T \geq h\} \\
&= \textstyle\sum_{h=1}^{k} c_h .
\end{aligned}$$

Osservazione. Sia h un intero compreso tra 1 e k. Allora, per ogni intero j compreso tra 1 e h, si ha, grazie alla definizione della funzione f_j,

$$f_j\left(x_{k-h+j}\right) = \begin{cases} q_j^{-1} & \text{se } x_j = x_{k-h+j} \\ \\ 0 & \text{altrimenti.} \end{cases}$$

Pertanto la costante c_h definita dalla relazione (7) può essere espressa nel modo seguente:

$$c_h = \begin{cases} \prod_{j=1}^{h} q_j^{-1} & \text{se } x_j = x_{k-h+j} \text{ per } 1 \leq j \leq h, \\ \\ 0 & \text{altrimenti.} \end{cases}$$

Si vede dunque che la costante c_h non è nulla se, e solo se, l'indice h è tale che siano tra loro identiche le due stringhe

$$\left(x_j\right)_{1 \leq j \leq h}, \qquad \left(x_{k-h+j}\right)_{1 \leq j \leq h}$$

(che possono esser chiamate, rispettivamente, il segmento iniziale e il segmento finale, di lunghezza h, dell'assegnata stringa). Sarà comodo convenire che, quando l'intero h sia dotato di questa proprietà, esso sia detto un *indice pesante* per l'assegnata stringa.

Si osservi che, in ogni caso, un particolare indice pesante per l'assegnata stringa $\left(x_j\right)_{1 \leq j \leq k}$ è l'indice k, ossia la lunghezza della stringa.

Esempio. Particolarizzando le ipotesi, si supponga ora che l'insieme E sia costituito dalle venticinque lettere maiuscole dell'alfabeto inglese e che μ sia la *ripartizione uniforme* su E, ossia la misura di probabilità su $\mathcal{P}(E)$ il cui valore su ciascun singoletto è 25^{-1}.

Si supponga inoltre, come nei citati libri di Williams, che l'assegnata stringa abbia lunghezza 11 e sia precisamente la stringa $(x_j)_{1\leq j\leq 11}$ così definita:

$$\begin{aligned}
x_1 &= A\\
x_2 &= B\\
x_3 &= R\\
x_4 &= A\\
x_5 &= C\\
x_6 &= A\\
x_7 &= D\\
x_8 &= A\\
x_9 &= B\\
x_{10} &= R\\
x_{11} &= A
\end{aligned}$$

Per questa stringa, i soli indici pesanti sono i numeri 1, 4, 11 e si ha $\quad c_1 = 25, \quad c_4 = 25^4, \quad c_{11} = 25^{11}.$

La speranza del tempo d'arresto T è data dunque da

$$\int T\,\mathrm{d}P = 25 + 25^4 + 25^{11}.$$

Calcolo esplicito della funzione generatrice

Lasciando da parte il semplice esempio di Williams, torniamo al quadro generale d'ipotesi del Paradosso di Borel.

Il risultato già visto, cioè il calcolo esplicito della speranza del tempo d'arresto T in funzione dei coefficienti c_h, è stato notevolmente ampliato da Rita Giuliano, la quale ne ha dedotto il calcolo esplicito (sempre in funzione dei coefficienti c_h) della funzione generatrice di T (e dunque la determinazione esplicita della legge di T).

Al fine di esporre il procedimento impiegato da Rita per giungere al suo risultato, converrà premettere qualche notazione e qualche osservazione preliminare.

Se Z è una variabile aleatoria su $(\Omega, \mathcal{A}, P)$ a valori in $\mathbb{N}$, si denoterà con G_Z la funzione generatrice di Z, ossia la funzione così definita sull'intervallo $]0,1]$:

$G_Z(u) = \int u^Z \mathrm{d}P = \sum_{n\geq 0} u^n P\{Z = n\}$ per $0 < u \leq 1$.

Per ogni intero n strettamente positivo, se, partendo dalla banale eguaglianza

$$\{T = n\} = \{T > n-1\} \setminus \{T > n\},$$

si tien conto dell'espressione di $P\{T > n\}$ fornita da (8) e dell'analoga espressione di $P\{T > n-1\}$, si trova

$$P\{T = n\}$$
$$= \sum_{h=1}^{k} c_h \Big[P\{T = n - 1 + h\} - P\{T = n + h\} \Big].$$

Inoltre, essendo $P\{T = k\} = c_k^{-1}$ e $P\{T = h\} = 0$ per ogni intero h con $1 \leq h < k$, si ha

$$\sum_{h=1}^{k} c_h \, P\{T = h\} = c_k c_k^{-1} = 1. \tag{9}$$

Tenendo conto di questi fatti, è possibile scrivere il valore $G_T(u)$ della funzione generatrice G_T nel generico punto u di $]0, 1]$ in ciascuna delle forme seguenti:

$$\sum_{n \geq 0} u^n \sum_{h=1}^{k} c_h \Big[P\{T = n - 1 + h\} - P\{T = n + h\} \Big],$$
$$\sum_{h=1}^{k} c_h \sum_{n \geq 0} u^n \Big[P\{T + 1 - h = n\} - P\{T - h = n\} \Big],$$
$$\sum_{h=1}^{k} c_h \sum_{n \geq 1} u^n \Big[P\{T + 1 - h = n\} - P\{T - h = n\} \Big]$$
$$- \sum_{h=1}^{k} c_h \, P\{T = h\}.$$

Grazie alla relazione (9), si ha dunque l'eguaglianza

$$G_T(u) = \sum_{h=1}^{k} c_h \Big[u^{1-h} G_T(u) - u^{-h} G_T(u) \Big] - 1,$$

dalla quale si deduce

$$(10) \qquad G_T(u) = \Big[1 + \textstyle\sum_{h=1}^{k} c_h \left(u^{-h} - u^{1-h}\right)\Big]^{-1},$$

ossia la desiderata espressione esplicita di $G_T(u)$ in funzione dei coefficenti c_h.

Questo risultato di Rita Giuliano è stato poi arricchito dalla seguente osservazione di Luca Pratelli.

Osservazione. Grazie all'eguaglianza $T = S + k$, la funzione generatrice di S è legata a quella di T dalla relazione $G_T(u) = u^k G_S(u)$. Perciò dall'eguaglianza (10) discende che la funzione generatrice di S coincide, sull'intervallo $]0, 1]$, col reciproco del polinomio Q così definito:

$$(11) \qquad Q(u) = u^k + (1 - u) \textstyle\sum_{h=1}^{k} c_h \, u^{k-h}.$$

Questo polinomio può esser messo nella forma

$$Q(u) = \textstyle\sum_{j=0}^{k} b_j u^j,$$

con $b_0 = c_k,\ b_k = 1-c_1,\ b_j = c_{k-j+1}$ per $1 \leq j \leq k-1$. Di conseguenza, se, per ogni intero positivo n, si pone

$$a_n = P\{S = n\},$$

e si scrive $G_S(u)$ nella forma $\sum_{n\geq 0} a_n u^n$, si trova

$$\left(\sum_{n\geq 0} a_n u^n\right)\left(\sum_{j=0}^{k} b_j u^j\right) = 1.$$

Ne segue, grazie al principio d'identità delle serie intere,

$$(12) \quad a_0 b_0 = 1, \qquad \sum_{j=0}^{h\wedge k} a_{h-j} b_j = 0 \quad \text{per} \quad h \geq 1.$$

È dunque possibile (essendo noti i coefficienti b_j) ottenere ricorsivamente i coefficienti a_n, ossia la densità discreta della legge di S. Basta, per questo, scrivere le relazioni (12) nella forma

$$a_0 = b_0^{-1}, \quad a_h = -b_0^{-1} \sum_{j=1}^{h\wedge k} a_{h-j} b_j \quad \text{per } h \geq 1.$$