Una breve introduzione al Calcolo delle probabilità

(Giorgio Letta, agosto 2010)

Il Calcolo delle probabilità, inteso come disciplina matematica, ha, rispetto ad altre branche della Matematica, come la Geometria, origini molto più recenti: la sua fondazione si fa comunemente risalire ad uno scambio epistolare che ebbe luogo nel 1654 tra Pascal e Fermat a proposito di una questione legata ai giochi d'azzardo. Altri illustri matematici, come Jacob Bernoulli, Laplace e Gauss (per non citare che i più grandi), si occuparono in seguito di Calcolo delle probabilità. Ma agli inizi del ventesimo secolo, questa disciplina, nonostante i grandi progressi compiuti, mancava ancora di un preciso quadro assiomatico. Non a caso, David Hilbert, in occasione di uno storico congresso internazionale dei matematici, svoltosi a Parigi nell'anno 1900, aveva inserito il problema di una fondazione assiomatica del Calcolo delle probabilità in una lista di grandi problemi matematici insoluti, che il secolo appena concluso lasciava idealmente in eredità al nuovo secolo. Finalmente, nel 1933, il matematico russo A. N. Kolmogorov espose, in un agile volumetto, la sua soluzione al problema posto da Hilbert: soluzione che consisteva nell'inquadrare il Calcolo delle probabilità nell'ambito della Teoria della misura. Dopo di allora, il Calcolo delle probabilità ha conosciuto uno sviluppo enorme, fino a diventare uno dei settori più importanti della Matematica, anche in ragione delle sue applicazioni ai più disparati problemi, in ogni campo del sapere.

Anziché tentare di passare in rassegna, in modo necessariamente superficiale e incompleto, i principali risultati e le principali applicazioni del moderno Calcolo delle probabilità, mi limiterò, in questa breve esposizione, a presentare l'assiomatizzazione di Kolmogorov, accompagnandola con alcuni brevi commenti e con alcuni esempi elementari. Questi commenti e questi esempi hanno soprattutto lo scopo di aiutare a capire quanto sia importante, per chi si occupi di applicazioni del Calcolo delle probabilità, tenere sempre rigorosamente distinte le tre fasi in cui ogni corretta applicazione deve articolarsi: la scelta del modello matematico, l'elaborazione matematica all'interno del modello scelto, e infine l'interpretazione dei risultati matematici ottenuti.

1. Esperimento aleatorio

La base intuitiva del Calcolo delle probabilità è costituita dal concetto di *esperimento aleatorio*. Un esperimento si dice aleatorio, per un certo individuo, in un certo istante, se l'individuo non è ancora in grado di indicarne con sicurezza il risultato (indipendentemente dal fatto che l'esperimento sia stato già eseguito o debba ancora essere eseguito).

Se l'individuo che si trova in una tale situazione d'incertezza è interessato al risultato dell'esperimento (per esempio, in vista di qualche scommessa), è naturale

che egli si preoccupi innanzitutto di fissare un "ventaglio completo di eventualità, a due a due incompatibili", ossia un insieme Ω, i cui elementi rappresentino ipotetici risultati dell'esperimento, con la certezza che, comunque vadano le cose, il risultato effettivo dell'esperimento "cadrà in Ω" (nel senso che sarà rappresentato da uno e da un sol elemento di Ω).

(1.1) Esempio. Si supponga che l'esperimento consista nel lanciare un dado. Se per "risultato" s'intende il numero della faccia che uscirà, si potrà prendere come Ω l'insieme $\{1, 2, 3, 4, 5, 6\}$.

(1.2) Esempio. Si supponga che l'esperimento sia costituito dalle estrazioni del lotto che saranno eseguite, sabato prossimo, sulla ruota di Napoli. Se per "risultato" s'intende l'insieme dei cinque numeri estratti (prescindendo dall'ordine di estrazione), si potrà prendere come Ω l'insieme di tutte le cinquine, intendendo per *cinquina* un insieme di cinque numeri interi compresi tra 1 e 90.

2. Eventi legati al risultato di un esperimento aleatorio

Tornando al caso generale di un arbitrario esperimento aleatorio, al quale sia stato associato (nel senso sopra chiarito) un certo insieme Ω di eventualità, consideriamo ora una qualsiasi parte A di Ω. Si può interpretare A come rappresentante un *evento legato al risultato dell'esperimento*: l'evento che si realizzerà se e solo se codesto risultato "cadrà in A". Si può anzi *identificare* questo evento con l'insieme A stesso.

(2.1) Esempio. Nel caso del lancio di un dado, la parte $\{2, 4, 6\}$ dell'insieme Ω si può identificare con l'evento indicato dalle parole: "uscita di una faccia pari".

(2.2) Esempio. Nel caso del lotto, l'evento "uscita del 34" si può identificare con l'insieme di tutte le cinquine ammettenti 34 come elemento.

Nel caso generale, una volta scelto un certo insieme Ω come insieme delle eventualità, anche se è sempre possibile interpretare *ogni* parte di Ω come un evento legato al risultato dell'esperimento, può darsi che certe parti di Ω corrispondano a eventi non interessanti (ai fini di un determinato problema) oppure troppo complicati per poter essere studiati. In ciascun caso, dunque, e per ciascun problema da studiare, converrà scegliere una determinata classe $\mathcal{A}$ di parti di Ω, e riservare il nome di *eventi* agli elementi di questa classe. Sarà opportuno scegliere la classe $\mathcal{A}$ in modo tale che essa possieda buone proprietà di *stabilità* (rispetto alle più comuni operazioni insiemistiche). Precisamente, sarà conveniente esigere che l'intero insieme Ω sia un elemento di $\mathcal{A}$, che il complementare (rispetto a Ω) di un qualsiasi elemento di $\mathcal{A}$ sia ancora un elemento di $\mathcal{A}$ e che l'unione di una qualsiasi successione di elementi di $\mathcal{A}$ sia ancora un elemento di $\mathcal{A}$. (Come conseguenza, anche l'insieme vuoto e l'intersezione di una qualsiasi successione di elementi di $\mathcal{A}$ saranno ancora elementi di $\mathcal{A}$.) Una classe $\mathcal{A}$ con queste proprietà si chiama, in termini tecnici, una *tribù* (o, meno bene, una *σ–algebra*) su Ω. Inoltre, se $\mathcal{A}$ è una tribù su Ω, la coppia $(\Omega, \mathcal{A})$ si chiama

uno *spazio probabilizzabile*. Nell'ambito di un fissato spazio probabilizzabile $(\Omega, \mathcal{A})$, l'insieme Ω si chiama *l'insieme delle eventualità*, mentre la tribù $\mathcal{A}$ si chiama *la tribù degli eventi*. Usando questo linguaggio, possiamo così riassumere le considerazioni precedenti:

Quando si voglia studiare un esperimento aleatorio, il primo passo da compiere consiste nell'associargli uno spazio probabilizzabile $(\Omega, \mathcal{A})$.

3. Un po' di terminologia

Nell'ambito di un fissato spazio probabilizzabile $(\Omega, \mathcal{A})$, si adopera una terminologia particolarmente suggestiva: si usa dire che l'eventualità ω *realizza* l'evento A per dire che ω appartiene ad A. Inoltre:

- L'insieme vuoto, ossia l'evento che non è realizzato da alcuna eventualità, si denota con $\emptyset$ e si chiama anche l'*evento impossibile.*
- Se A è un evento, si denota con A^c, e si chiama il *complementare* di A (o anche l'evento *contrario* di A), l'evento che è realizzato da tutte e sole le eventualità che non realizzino A.
- Se A, B sono due eventi, si denota con $A \cup B$ la loro *unione*, ossia l'evento che è realizzato da tutte e sole le eventualità che realizzino *uno almeno* dei due eventi A, B.
- Se A, B sono due eventi, si denota con $A \cap B$ la loro *intersezione*, ossia l'evento che è realizzato da tutte e sole le eventualità che realizzino *entrambi* gli eventi A, B.
- Due eventi A, B si dicono tra loro *disgiunti* (o *incompatibili*) se non esiste alcuna eventualità che li realizzi entrambi, cioè se vale l'eguaglianza $A \cap B = \emptyset$.
- Si dice che l'evento A è *contenuto* nell'evento B (e si scrive $A \subset B$) se ogni eventualità che realizzi A realizza anche B.

Nell'ambito di un fissato spazio probabilizzabile $(\Omega, \mathcal{A})$ si può anche introdurre il concetto di *variabile aleatoria*. Noi ci limiteremo a definirne un caso particolare. Dato un insieme non vuoto E, che sia finito oppure infinito numerabile (come l'insieme degli interi), una funzione X definita su Ω e a valori in E si chiama una *variabile aleatoria discreta* su $(\Omega, \mathcal{A})$ se, per ciascun elemento x di E, l'insieme costituito dalle eventualità ω tali che si abbia $X(\omega) = x$ (insieme che si denota concisamente con $\{X = x\}$) è un evento.

4. Scelta dello spazio probabilizzabile

Giova osservare che, nella scelta dello spazio probabilizzabile $(\Omega, \mathcal{A})$ da associare a un determinato esperimento aleatorio, c'è sempre una certa dose di *arbitrarietà*, e ciò per due ragioni principali:

(a) Innanzitutto, gli elementi di Ω *rappresentano* ipotetici risultati dell'esperimento, secondo un opportuno *codice*: è chiaro che la scelta di questo codice è, in

larga misura, arbitraria. Ad esempio, se l'esperimento consiste nel lancio di una moneta, e se ci s'interessa solo alla faccia che apparirà (testa o croce), si potrà prendere $\Omega = \{0, 1\}$, con la convenzione che 0 significhi croce, e 1 testa. Ma egualmente legittima sarebbe la convenzione inversa (0=testa; 1=croce), oppure la scelta, in luogo dell'insieme $\{0, 1\}$, di un qualsiasi altro insieme costituito da due elementi.

(b) In secondo luogo, una volta fissato l'insieme Ω delle eventualità, la scelta della tribù $\mathcal{A}$ dipende dal criterio in base al quale si decide di considerare certi eventi come interessanti e altri eventi come non interessanti.

5. Misurazione del grado di fiducia

In uno studio probabilistico di un complesso di eventi legati al risultato di un esperimento aleatorio, la scelta dello spazio probabilizzabile $(\Omega, \mathcal{A})$ costituisce soltanto il primo passo. Un secondo passo consiste nella scelta di una "misura di probabilità". Cercheremo di spiegare che cosa ciò significhi.

Cominciamo con l'osservare che, per definizione stessa di esperimento aleatorio, l'individuo che considera un tal esperimento non è in grado di stabilire con certezza, per ogni singolo evento A della tribù $\mathcal{A}$, se esso si realizzarà o no. Ciò tuttavia non gl'impedisce di sentire – su un piano meramente psicologico – un diverso *grado di fiducia* nei confronti dei diversi eventi della tribù $\mathcal{A}$.

Sarà allora naturale, per l'individuo, cercar di *misurare* questo grado di fiducia, associando a ciascun evento A della tribù $\mathcal{A}$ un numero $P(A)$, ossia definendo una funzione P sulla tribù $\mathcal{A}$. Per convenzione, si può prendere questa funzione a valori nell'intervallo $[0, 1]$, e assumente il valore 1 sull'evento Ω. Sarà anche naturale pretendere che essa sia *additiva*, nel senso che verifichi la relazione

$$P(A \cup B) = P(A) + P(B) \tag{5.1}$$

per ogni coppia A, B di eventi disgiunti. In realtà, ragioni di comodità matematica consigliano di pretendere che P verifichi una condizione di additività più forte (*additività numerabile*), ottenuta considerando, anziché una coppia di eventi disgiunti, un'arbitraria famiglia *numerabile* di eventi a due a due disgiunti. Una funzione P con queste proprietà è detta una *misura normalizzata*, o anche una *misura di probabilità*, sulla tribù $\mathcal{A}$.

(5.2) Osservazione. Se P è una misura di probabilità su $\mathcal{A}$, allora P è *isotona*, nel senso che, per ogni coppia A, B di eventi con $A \subset B$, si ha $P(A) \leq P(B)$. Infatti, essendo B eguale all'unione dei due eventi disgiunti A e $B \cap A^c$, l'additività di P implica la relazione

$$P(B) = P(A) + P\left(B \cap A^c\right). \tag{5.3}$$

Da questa stessa relazione, prendendo $B = \Omega$, si deduce anche $1 = P(A) + P(A^c)$, ossia

$$P(A^c) = 1 - P(A). \tag{5.4}$$

(5.5) Definizione. Se P è una misura di probabilità sulla tribù $\mathcal{A}$, allora la terna $(\Omega, \mathcal{A}, P)$ è detta uno *spazio probabilizzato*, e, per ogni elemento A di $\mathcal{A}$, il numero $P(A)$ è detto la *probabilità* dell'evento A (secondo la misura P).

Per quanto ovvio, è forse utile sottolineare che, secondo la definizione precedente, non ha senso parlare di "probabilità di un evento", se non nell'ambito di un ben definito spazio probabilizzato. In particolare, se si è costruito soltanto uno spazio probabilizzabile $(\Omega, \mathcal{A})$, non ha ancora senso chiedersi quale sia la probabilità di un assegnato evento A (elemento della tribù $\mathcal{A}$): infatti esistono, in generale, molte misure di probabilità sulla tribù $\mathcal{A}$, e non è detto che queste assumano tutte lo stesso valore sull'evento A.

Usando il linguaggio sopra introdotto, possiamo ora così riassumere le considerazioni precedenti:

Compito preliminare, per un individuo che intenda studiare dal punto di vista probabilistico un esperimento aleatorio, è quello di associargli un opportuno spazio probabilizzato $(\Omega, \mathcal{A}, P)$.

6. Scelta della misura di probabilità

Abbiamo già richiamato l'attenzione sul carattere inevitabilmente arbitrario della scelta dello spazio probabilizzabile $(\Omega, \mathcal{A})$ da associare a un determinato esperimento aleatorio. Ancor meno scontata e automatica è la scelta di P. Quale sarà il modo migliore di compierla? Osserviamo, a questo proposito, che la misura di probabilità P da scegliere è un oggetto matematico, col quale si vuole "fotografare" la distribuzione della propria fiducia tra i diversi eventi. Converrà dunque scegliere P in modo che questa fotografia risulti "il più fedele possibile".

Naturalmente sarebbe vano cercar di precisare che cosa ciò significhi dal punto di vista matematico. Tuttavia nella scelta di P può essere di grande aiuto la Teoria della misura. Questa può in molti casi trasformare la scelta di P addirittura in una "scelta obbligata", qualora a P s'impongano certe condizioni aggiuntive, più o meno "naturali" (suggerite, per esempio, da ragioni di simmetria). Nelle prossime due sezioni, vedremo appunto due esempi di "scelta obbligata" (il primo banale, il secondo tutt'altro che banale).

Resta ad ogni modo il fatto che lo spazio probabilizzato $(\Omega, \mathcal{A}, P)$ da associare a un determinato esperimento aleatorio è un "modello matematico" e che la sua scelta è, in ogni caso, un'operazione *pre–matematica*. Chiedersi se una certa scelta sia "giusta o sbagliata" non ha dunque senso: o, perlomeno, non ha lo stesso senso che chiedersi se siano giusti o sbagliati determinati conti eseguiti nell'ambito di un particolare modello scelto.

7. Un caso di scelta obbligata: estrazioni del lotto

Consideriamo un caso veramente banale: quello in cui l'insieme Ω sia finito e la tribù $\mathcal{A}$ degli eventi coincida con la tribù di tutte le parti di Ω. In questo caso, per costruire una misura normalizzata sulla tribù di tutte le parti di Ω, basta assegnarne

i valori sulle parti costituite da un sol elemento (cioè sui *singoletti*): e questi valori si possono scegliere in modo arbitrario, con l'unico vincolo che siano compresi tra 0 e 1 e ammettano 1 come somma. In altre parole: per definire la più generale misura di probabilità sulla tribù di tutte le parti di Ω, basta associare a ciascun elemento di Ω un numero compreso tra 0 e 1, che sarà comodo chiamare un *peso*, in modo tale che la somma di tutti questi numeri sia eguale a 1. Dopo di ciò, si definirà la misura P nel modo seguente: per ogni parte A di Ω, il valore $P(A)$ che P assume su A è la somma dei *pesi* associati ai singoli elementi di A.

In particolare, quando i pesi associati ai singoli elementi di Ω siano eguali tra loro, ossia tutti eguali a $1/n$, dove n denota il numero degli elementi di Ω, la corrispondente misura di probabilità P si chiama la *ripartizione uniforme* su Ω. Il suo valore $P(A)$ sulla generica parte A di Ω è la somma di tanti pesi, tutte eguali a $1/n$, quanti sono gli elementi di A, ossia è il rapporto tra il numero m delle eventualità che realizzano A e il numero totale delle eventualità:

$$P(A) = m \cdot (1/n) = \frac{m}{n} \cdot \tag{7.1}$$

Sarà questa la misura da scegliere qualora si giudichi sensato, per ragioni di simmetria suggerite dalla particolare natura dell'esperimento studiato, trattare in modo imparziale i diversi singoletti.

Ad esempio, nel caso del lotto (si veda l'Esempio (1.2)), la ripartizione uniforme sull'insieme Ω di tutte le cinquine appare come la più naturale, tra tutte le misure di probabilità sulla tribù delle parti di Ω, in quanto l'unica che tratti in modo imparziale le varie cinquine possibili. Gli accorgimenti usati nell'eseguire le estrazioni (i sofisticati e trasparenti apparecchi di estrazione che sostituiscono oggi i bambini bendati di una volta, e la vigile presenza di un funzionario dell'Intendenza di Finanza) hanno appunto lo scopo di togliere al giocatore ogni motivo *razionale* per rifiutare un tale atteggiamento d'imparzialità.

8. Un altro caso di scelta obbligata: lanci di una moneta

Consideriamo l'esperimento aleatorio che consiste nell'eseguire una successione di lanci di una moneta arrestandosi al primo "successo" (dove, con questo termine, intendiamo convenzionalmente "apparizione di testa"). Poniamoci la questione seguente.

Questione. *Per ciascun intero n (con $n \geq 1$), qual è la probabilità che l'esperimento termini col lancio n–esimo?*

Dal punto di vista empirico, l'esperimento sopra descritto appare ben definito, facilissimo da eseguire e, per giunta, di durata molto breve. Ma, se tentiamo di formalizzarlo in modo matematicamente rigoroso, c'imbattiamo in una difficoltà: comunque grande si fissi un intero n, non possiamo escludere, da un punto di vista puramente logico, che il primo successo appaia *dopo* il lancio n–esimo. Anzi, non possiamo neppure escludere che il primo successo *non ci sia*, nel senso che tutti i lanci diano come risultato "croce", sicché l'esperimento *non abbia termine*.

Siamo dunque costretti a inventarci un esperimento aleatorio *ideale* consistente in una successione *infinita* di lanci.

Compiamo con ciò un salto gigantesco, che ci fa perdere ogni aderenza con la realtà empirica. È chiaro infatti che nessun mortale potrà mai eseguire infiniti lanci di una moneta. Ma un tal distacco dalla realtà non è un guaio, né una novità. Ciò che noi vogliamo costruire è un *modello matematico* per un esperimento ideale: ed è ben noto che i modelli matematici corrispondono molto spesso a idealizzazioni estreme, e persino grottesche, di situazioni empiriche. Lanciamoci dunque, senza ulteriori indugi, nell'impresa di costruire un adeguato modello matematico da associare a quell'ideale esperimento aleatorio. Più precisamente, proponiamoci di costruire uno spazio probabilizzato $(\Omega, \mathcal{A}, P)$ e, su di esso, una successione infinita $(X_1, X_2, \ldots)$ di variabili aleatorie, a valori in $\{0, 1\}$, la quale rappresenti la successione dei risultati dei singoli lanci.

Cominciamo con l'introdurre alcune locuzioni che ci aiuteranno a rendere più spedito il linguaggio. Chiameremo *cifre binarie* i due numeri 0 e 1. Inoltre, dato un qualsiasi insieme I, chiameremo *stringa binaria* avente I come insieme degli indici (o, in modo più conciso, *I–stringa*) ogni famiglia di cifre binarie, avente I come insieme degli indici. In altre parole, una stringa binaria avente I come insieme degli indici è una funzione che ad ogni elemento i di I associa una cifra binaria x_i. Diremo che la stringa è *nulla* se le cifre x_i sono tutte eguali a 0.

Ciò posto, se adottiamo il codice che identifica "testa" con 1 e "croce" con 0, possiamo scegliere come insieme delle eventualità da associare al nostro ideale esperimento aleatorio l'insieme Ω costituito da tutte le stringhe binarie aventi come insieme degli indici l'insieme $\mathbb{N}^*$ degli interi maggiori di zero. Naturalmente, sarà da intendere che la generica $(x_1, x_2, \ldots)$ di queste stringhe rappresenti il risultato che si ottiene quando per ciascun indice i, il risultato del lancio di indice i sia rappresentato dalla cifra binaria x_i (cioè sia "testa" oppure "croce", secondo che x_i sia 1 oppure 0).

Introduciamo, per ogni indice i, la funzione X_i così definita su Ω: per ogni elemento ω di Ω, il valore $X_i(\omega)$ che X_i assume in ω è la cifra binaria di indice i nella stringa ω. L'insieme $\{X_i = 1\}$ costituito da tutti gli elementi di Ω sui quali la funzione X_i assume il valore 1 sarà dunque l'insieme costituito da tutte le stringhe binarie che presentano la cifra 1 in corrispondenza dell'indice i. Analogo significato avrà, naturalmente, l'insieme $\{X_i = 0\}$.

Gli insiemi della forma $\{X_i = 1\}$ sono speciali parti di Ω. Tra tutte le tribù su Ω che contengano la classe di quegli insiemi, ne esiste una *minima* (cioè contenuta in tutte le altre). Sarà naturale scegliere proprio questa come la tribù $\mathcal{A}$ degli eventi. Ad essa appartengono, naturalmente, anche tutti gli insiemi della forma $\{X_i = 0\}$. Più in generale, fissato un intero n (con $n \geq 1$), se $(x_1, \ldots, x_n)$ è una stringa binaria avente $\{1, \ldots, n\}$ come insieme degli indici, allora l'insieme

$$\{X_1 = x_1\} \cap \cdots \cap \{X_n = x_n\} \tag{8.1}$$

è un elemento di $\mathcal{A}$, ossia un evento. Si osservi che, per ciascun indice i, grazie alla scelta da noi adottata per la tribù $\mathcal{A}$, la funzione X_i è una variabile aleatoria. La sua interpretazione è evidente: X_i rappresenta il risultato del lancio di indice i.

Fissato n, gli eventi della forma (8.1) (al variare in tutti i modi possibili della stringa binaria $(x_1, \ldots, x_n)$) formano una *partizione* di Ω, ossia sono a due a due disgiunti e hanno come unione l'intero insieme Ω. Essi sono tanti quante le stringhe binarie aventi $\{1, \ldots, n\}$ come insieme degli indici, cioè sono in numero di 2^n.

Rimane infine da scegliere una misura di probabilità P sulla tribù $\mathcal{A}$. Se si giudica opportuno trattare in modo imparziale le due facce della moneta, sarà naturale esigere che la misura P sia tale che, *per ogni* n, essa attribuisca la stessa probabilità a ciascuno dei 2^n eventi della forma (8.1), cioè dia a ciascuno di essi, come probabilità, 2^{-n}:

$$P\big(\{X_1 = x_1\} \cap \cdots \cap \{X_n = x_n\}\big) = 2^{-n}. \tag{8.2}$$

Esisterà effettivamente una misura di probabilità su $\mathcal{A}$ che verifichi questa condizione? Fortunatamente, la risposta è affermativa. Ma la dimostrazione è tutt'altro che immediata. Essa si basa su risultati non banali di Teoria della misura.

Resta così completata la costruzione di uno spazio probabilizzato $(\Omega, \mathcal{A}, P)$ da considerare come un adeguato modello matematico da associare all'ideale esperimento aleatorio consistente nell'eseguire una successione infinita di lanci di una moneta.

9. L'istante del primo successo

Rimanendo nel quadro dell'esperimento aleatorio ideale considerato nella sezione precedente, mantenendo le notazioni ivi introdotte e immaginando inoltre (al solo scopo di rendere più spedito e suggestivo il linguaggio) che i lanci avvengano agli istanti $1, 2, \ldots$, vogliamo ora definire una variabile aleatoria T che rappresenti "il tempo da attendere per veder apparire il primo successo".

Cominciamo col ricordare che il simbolo $\mathbb{N}^*$ designa l'insieme degli interi maggiori di zero. Questo insieme ammette 1 come minimo elemento, ma è sprovvisto di massimo elemento. Talvolta è utile considerare l'insieme ottenuto aggiungendo a $\mathbb{N}^*$ un ulteriore elemento, denotato con ∞, e prolungare all'insieme così ottenuto la relazione d'ordine di $\mathbb{N}^*$ convenendo che si abbia $n < \infty$ per ogni intero n. Il nuovo insieme, che indicheremo con $\mathbb{N}^* \cup \{\infty\}$, ammette dunque ∞ come massimo elemento.

Consideriamo ora la funzione T così definita su Ω: per ogni elemento ω di Ω che non sia la stringa nulla, definiamo $T(\omega)$ come il minimo intero n tale che si abbia $X_n(\omega) = 1$. Se poi ω è la stringa nulla, poniamo $T(\omega) = \infty$.

Allora gli insiemi della forma $\{T = n\}$ (con n intero maggiore di zero) sono eventi. Si ha infatti $\{T = 1\} = \{X_1 = 1\}$ e, per ogni intero n maggiore di 1,

$$\{T = n\} = \{X_1 = 0\} \cap \cdots \cap \{X_{n-1} = 0\} \cap \{X_n = 1\}. \tag{9.1}$$

Anche l'insieme $\{T = \infty\}$ (costituito dalla sola stringa nulla) è un evento: esso è infatti il complementare dell'unione di tutti gli eventi della forma $\{T = n\}$, con n intero maggiore di zero.

Dunque la funzione T è una variabile aleatoria discreta su $(\Omega, \mathcal{A})$, a valori in $\mathbb{N}^* \cup \{\infty\}$. Dalle relazioni precedenti risulta inoltre che, per ogni intero n maggiore di zero, si ha, come caso particolare di (8.2),

$$P\{T = n\} = 2^{-n}. \tag{9.2}$$

Infine, dall'ovvia eguaglianza

$$\{T > n\} = \{X_1 = 0\} \cap \cdots \cap \{X_n = 0\}, \tag{9.3}$$

si deduce (sempre come caso particolare di (8.2))

$$P\{T > n\} = 2^{-n}. \tag{9.4}$$

L'evento $\{T = \infty\}$ è contenuto in ciascuno degli eventi della forma $\{T > n\}$. La sua probabilità, per l'isotonia della misura P, è dunque maggiorata da 2^{-n} per ogni n, ossia è nulla:

$$P\{T = \infty\} = 0. \tag{9.5}$$

L'evento $\{T = \infty\}$ si può interpretare come "assenza di successi".

La variabile aleatoria T si può interpretare come "il tempo da attendere per veder apparire il primo successo". In particolare, sull'evento $\{T < \infty\}$, essa si può interpretare come "l'istante del primo successo".

I risultati matematici (9.5) e (9.2) che abbiamo ottenuto all'interno del modello matematico $(\Omega, \mathcal{A}, P)$ possono essere così interpretati in termini dell'esperimento aleatorio (ideale) dal quale eravamo partiti e al quale abbiamo associato quel modello:

In una successione infinita di lanci di una moneta, la probabilità che non esca mai "testa" è nulla, mentre, per ogni intero n maggiore di zero, la probabilità che si ottenga "testa" per la prima volta nel lancio n–esimo è eguale a 2^{-n}.

10. Il paradosso di Borel

Consideriamo ancora l'esperimento aleatorio ideale consistente in una successione infinita di lanci di una moneta, e conserviamo le notazioni e il linguaggio delle Sezioni 8 e 9. Inoltre, fissato un intero h maggiore di zero, decomponiamo l'insieme $\mathbb{N}^*$ degli "istanti di lancio" in una successione infinita $(I_k)_{k \geq 1}$ d'intervalli temporali di lunghezza h, nel modo seguente: I_1 è l'intervallo $\{1, \ldots, h\}$ costituito dai primi h interi, I_2 è l'intervallo $\{h + 1, \ldots, 2h\}$ costituito dai successivi h interi, e così via.

Ci domandiamo quale sia la probabilità che, in uno almeno degli intervalli temporali I_k così definiti, si abbia totale assenza di successi (ossia apparizione di sole croci). Per poter formalizzare in modo preciso questa domanda, denotiamo, per ogni k, con S_k la somma delle variabili aleatorie X_i con i appartenente a I_k. Questa

somma è una variabile aleatoria che rappresenta il numero dei successi ottenuti nell'intervallo temporale I_k. L'evento di cui si chiede di calcolare la probabilità è dunque l'evento A così definito:

$$A = \{S_1 = 0\} \cup \{S_2 = 0\} \cup \cdots \tag{10.1}$$

Sarà più comodo occuparsi del suo complementare:

$$A^c = \{S_1 \neq 0\} \cap \{S_2 \neq 0\} \cap \cdots \tag{10.2}$$

L'evento A^c è contenuto, per ogni intero $n \geq 1$, nell'evento B_n così definito:

$$B_n = \{S_1 \neq 0\} \cap \cdots \cap \{S_n \neq 0\}. \tag{10.3}$$

Dunque, per l'isotonia della misura P, si ha

$$P(A^c) \leq P(B_n). \tag{10.4}$$

Calcoliamo la probabilità di B_n. A questo scopo, consideriamo l'intervallo temporale J_n definito da

$$J_n = \{1, \ldots, nh\} = I_1 \cup \cdots \cup I_n \tag{10.5}$$

e, per ogni J_n–stringa $(x_1, \ldots, x_{nh})$, consideriamo l'evento

$$\{X_1 = x_1\} \cap \cdots \cap \{X_{nh} = x_{nh}\}. \tag{10.6}$$

Sappiamo che gli eventi di questa forma sono tanti quante le J_n–stringhe (cioè 2^{nh}). Inoltre essi sono a due a due disgiunti, e ciascuno di essi ha probabilità eguale a 2^{-nh}. D'altra parte, l'evento B_n è l'unione di tutti gli eventi esprimibili nella forma (10.6) mediante una J_n–stringa $(x_1, \ldots, x_{nh})$ che verifichi la condizione seguente: *nessuna delle sue restrizioni ad uno degli intervalli $I_1, \ldots, I_n$ sia nulla*. Se conveniamo di chiamare *speciali* le J_n–stringhe verificanti questa condizione, possiamo dunque dire che la probabilità di B_n si ottiene semplicemente moltiplicando la comune probabilità 2^{-nh} degli eventi della forma (10.6) per il numero, diciamolo s_n, delle J_n–stringhe speciali. Per calcolare s_n, osserviamo che scegliere una J_n–stringa speciale equivale a scegliere, per ogni intero k compreso tra 1 e n, una qualsiasi tra le I_k–stringhe non nulle (che sono in numero di $2^h - 1$): ciò si può fare in $(2^h - 1)^n$ modi diversi. Si ha dunque $s_n = (2^h - 1)^n$, e quindi

$$P(B_n) = s_n\, 2^{-nh} = \left(1 - 2^{-h}\right)^n. \tag{10.7}$$

La probabilità così calcolata, essendo la potenza n–esima del numero $1 - 2^{-h}$ (strettamente compreso tra 0 e 1), tende a zero al tendere di n all'infinito (nel senso che, comunque si fissi un numero reale ϵ maggiore di zero, essa risulta minore di ϵ non appena n sia abbastanza grande). Grazie alla diseguaglianza (10.4), ne segue $P(A^c) = 0$, ossia $P(A) = 1$ (si veda (5.4)).

Il risultato matematico che abbiamo appena ottenuto all'interno del modello matematico $(\Omega, \mathcal{A}, P)$ si può così interpretare in termini di lanci di una moneta:

Comunque grande sia l'intero h fissato, se si esegue una successione infinita di lanci di una moneta, la probabilità che, in uno almeno degli intervalli temporali della forma I_k, si abbia totale assenza di successi (cioè apparizione di sole croci) è eguale a 1.

La nostra intuizione fatica ad accettare come vero un simile risultato. Siamo dunque in presenza di un *paradosso*. Lo si chiama infatti, in letteratura, il *paradosso di Borel*, dal nome del matematico francese Émile Borel (1871-1956). Come si spiega la natura paradossale di quel risultato? La risposta è molto semplice: paradossale non è né lo schema infinito che abbiamo costruito né il risultato matematico che, nel suo ambito, abbiamo dimostrato. Paradossale è soltanto l'*interpretazione* che di quello schema e di quel risultato si dà in termini di lanci di una moneta, cioè in termini di un particolare esperimento aleatorio. Ma il fatto che questa interpretazione conduca a un paradosso non provoca il minimo stupore se si tiene presente che l'esperimento in questione è un esperimento ideale, assolutamente irrealizzabile.

D'altra parte, l'utilità del modello infinito sopra introdotto è innegabile per due ragioni: innanzitutto, come abbiamo visto, esso permette di dare un senso preciso al concetto di "istante del primo successo"; in secondo luogo, nel suo ambito, è possibile, per ogni fissato intero n, formalizzare in modo molto semplice l'esperimento aleatorio che consiste nell'eseguire una successione finita di n lanci: basta per questo considerare, anziché la successione infinita delle variabili aleatorie X_i, la successione finita formata dalle prime n di esse.

Una situazione come la precedente è, del resto, molto comune in matematica: l'introduzione di un opportuno schema infinito si rivela spesso utile, se non addirittura indispensabile, per trattare schemi di tipo finito. Illuminanti, in proposito, sono le seguenti parole del grande matematico Ennio De Giorgi:

"... in matematica, anche se ci si vuole limitare a procedimenti finitistici, si devono ammettere regole di tipo non finitistico. Ad esempio, un'addizione fra due interi è un'operazione che si fa con un numero finito di passi, ma per definire l'addizione si è costretti a parlare dell'insieme (infinito) degli interi naturali."

Giorgio Letta
Dipartimento di Matematica "Leonida Tonelli"
Largo Bruno Pontecorvo, 5
56127 Pisa
E-mail: letta@dm.unipi.it