La candela

E così, nell'intervallo tra la precedente puntata e questa, il 53 è finalmente uscito sulla ruota di Venezia, dopo forse 180 estrazioni. Pare che il Lotto abbia pagato 600 milioni (di euro) e i giornali hanno titolato "sbancato il Lotto." Penso che qualsiasi impresa vorrebbe essere sbancata in quel modo: pagando 600 milioni dopo averne incassati almeno il quadruplo...

Ottimo pretesto per toccare un tema a carattere interdisciplinare: la probabilità. Tema a molte facce, come sapete: a parte il Lotto e i giochi in generale, è d'importanza fondamentale in fisica, ma anche in genetica come pure nella diagnostica medica e in non so quanti altri ambiti.

Ma cominciamo pure dall'evento che ho preso come occasione per il mio "incipit": i ritardi al Lotto, o più in generale nei giochi d'azzardo. Anche questo sottotema può essere affrontato da più lati, alcuni dei quali però (per es. la psicologia) non sono di mia competenza. A me interessa particolarmente un aspetto che esce un po' dalle trattazioni più tradizionali, pur restando strettamente nell'ambito scientifico. Ma l'unico modo che vedo per spiegare ciò che ho in mente, è di sviluppare il discorso.

Si può partire al modo tradizionale: come mai è quasi universale la convinzione che un numero in ritardo abbia maggiore probabilità di uscire? qual è il ragionamento (sbagliato) che quasi tutti fanno?

Rispondere è facile: ci si appella alla "legge dei grandi numeri," che si presume dica che alla lunga ogni evento si deve presentare un numero di volte proporzionale alla sua probabilità. Dato che la probabilità di un singolo estratto al Lotto è 1/18, ogni numero dovrebbe uscire in media una volta ogni 18 estrazioni; se è in ritardo, quel numero deve "colmare" il ritardo e quindi la sua uscita diventa più probabile.

Lo stesso ragionamento si fa nel caso anche più semplice della roulette: dato che neri e rossi sono equiprobabili, dopo una serie di rossi il nero sarà più probabile, e viceversa.

Cerchiamo di mostrare su questo esempio che cosa realmente dice la legge dei grandi numeri. Consideriamo un certo numero di lanci della roulette, per es. 10, e occupiamoci solo del colore, trascurando anche lo 0, che serve, com'è noto, a dare un vantaggio al banco (come se già non ne avesse abbastanza, grazie ai giocatori sconsiderati ...). In 10 lanci sono possibili una quantità di risultati, visto che a ogni lancio può uscire il rosso o il nero: tutte le sequenze da `NNNNNNNNNN` a `RRRRRRRRRR`, passando per `NRRRNRRNNN`, `RRRRNNNNRR`, ecc. In totale sono $2^{10} = 1024$.

Chiediamoci: quante sono le sequenze in cui si presentano 5 rossi e 5 neri, in qualunque ordine? Non mi aspetto che ricordiate come si fa il calcolo, per mezzo dei coefficienti binomiali, per cui vi do il risultato: 252, ossia quasi 1/4 del totale. Per l'esattezza, il 24.6%.

Immaginiamo ora di ripetere il calcolo con 20 lanci, per un totale di $2^{20} = 1\,048\,576$ sequenze possibili. Tra queste, quelle con esattamente 10 rossi e 10 neri sono 184 576, ossia il 17.6%.

* * *

Finora non abbiamo fatto che del calcolo combinatorio, e la probabilità non si è ancora vista. Ma ora cominciamo a farla entrare in gioco, con una serie di asserzioni:

a) supponiamo che l'esito di un lancio sia *casuale*, e che nero e rosso siano *equiprobabili*;

b) supponiamo che l'esito di ciascun lancio sia *indipendente* da quello degli altri lanci.

Queste asserzioni, e soprattutto il significato delle parole in corsivo, le dovremo discutere e giustificare. Ma per ora procediamo dandole per buone.

Dalle due ipotesi segue immediatamente che una qualsiasi sequenza ha la stessa probabilità di qualunque altra, e che se noi siamo interessati all'insieme di sequenze che hanno una certa caratteristica comune (per es. 5 rossi su 10 lanci) la probabilità di questo *evento composto* la otteniamo semplicemente calcolando il numero di sequenze "buone" e dividendolo per il numero totale.

Ma è quello che abbiamo già fatto: possiamo quindi riformulare i risultati visti sopra dicendo che su 10 lanci c'è una probabilità del 24.6% di avere 5 rossi, mentre su 20 lanci la probabilità di 10 rossi è del 17.6%, dunque *minore*.

È facile intuire che la stessa tendenza si manterrà anche aumentando il numero di lanci: la probabilità di avere *ugual numero* di rossi e di neri decresce sempre. Detto di sfuggita, si può anche dimostrare che decresce approssimativamente come l'inverso della radice quadrata del numero di lanci, ma questo non ci serve.

Sembrerebbe dunque di aver trovato tutt'altro da quello che ci aspettavamo: al crescere del numero di lanci, il pareggio tra rossi e neri *diventa sempre più improbabile*!

Ma proviamo a cambiare la proprietà "buona" della quale vogliamo calcolare la probabilità: chiediamoci quale sia la probabilità che il numero di rossi *non scarti più del* 10% da quello atteso, ossia dalla metà del totale dei lanci. Con 10 lanci, se escono 4 rossi oppure 6, siamo già fuori per il 20%; quindi solo 5 rossi vanno bene, e la probabilità richiesta è la stessa di prima: 24.6%.

Però con 20 lanci anche un numero di rossi pari a 9 o a 11 va bene, perché lo scarto relativo è giusto il 10%; dunque dobbiamo accettare come buoni un numero di rossi da 9 a 11 inclusi. Occorre rifare il calcolo, e di nuovo vi do il risultato in termini di probabilità: si trova 49.7%. Se prendiamo per es. 50 lanci, i casi buoni vanno da 23 a 27, e la probabilità risulta 52.0%.

La novità è che ora la probabilità *cresce* col numero n di lanci, e si può forse indovinare che cosa capita facendo crescere n indefinitamente: *la probabilità tende a* 1. Bene, proprio questo è il contenuto della *legge dei grandi numeri*:

> In n prove indipendenti di un evento che ha probabilità p in una prova, la probabilità che il numero di successi non scarti più di una data frazione ε dal valore atteso np tende a 1 per $n \to \infty$.

Nel nostro caso $p = 1/2$, e avevamo scelto $\varepsilon = 0.1$; il numero di successi era il numero di rossi negli n lanci. (Per inciso, quella che ho enunciata si chiama di solito "legge *debole* dei grandi numeri," perché ce n'è anche una versione più "forte," ma a noi basterà questa.)

Occorre osservare che l'enunciato della legge dei grandi numeri sembra un po' un gioco di parole, perché dice che *la probabilità* che il numero di successi per un evento di *data probabilità* ecc. tende a 1. Ossia dice che è sempre più probabile che il numero di successi si avvicini al numero atteso; ma è solo *probabile*, non certo, anche se la probabilità tende a 1.

È abbastanza naturale pensare che se un evento ha probabilità molto vicina a 1, sia quasi certo che esso si verifichi; e viceversa, se la probabilità si avvicina a 0, si tratterà di un'eventualità pressoché irrilevante. Ma è importantissimo aver chiaro che *questo non è detto nel teorema*! Ci torneremo fra poco.

* * *

Possiamo ora applicare quanto precede al caso da cui siamo partiti, ossia dai ritardi al Lotto. Abbiamo già detto che la probabilità che un certo numero esca in una singola estrazione è $p = 1/18$; quindi in $n = 180$ estrazioni (per fare un esempio) dovremmo aspettarci che quel numero esca $np = 10$ volte: questo è il numero atteso. Dobbiamo anche aspettarci che sia molto improbabile che esca molte più o molte meno di 10 volte. Notate che la legge dei grandi numeri dice che tale probabilità tenderà a 0 per $n \to \infty$, ma non dice niente in generale su quanto debba valere per $n = 180$ o per qualunque altro n (in realtà il calcolo delle probabilità permette di dire qualcosa di più, ma per il momento non ci serve di approfondire).

Del resto possiamo facilmente rispondere a qualche domanda particolare: per es. quale sarà la probabilità che il numero non esca *mai* in 180 estrazioni? Ecco come si fa il calcolo.

La probabilità che *non esca* in una singola estrazione è 17/18; in due estrazioni, che sono indipendenti, è $(17/18) \times (17/18)$; in n estrazioni, è $(17/18)^n$. Dobbiamo dunque calcolare $(17/18)^{180}$ che fa $3.4 \cdot 10^{-5}$, ossia circa 1/30 000. È un numero piccolo, ma cerchiamo di vedere in un altro modo che cosa significa. Immaginiamo un Lotto gigantesco, che abbia non 10 ruote, come in Italia, ma giusto 30 000; mettiamoci a guardare se in 180 estrazioni il 53 esce o no su tutte queste ruote: dobbiamo aspettarci che solo in una delle ruote esso non si faccia mai vedere.

(Si badi bene che ho detto "aspettarci": non ne possiamo essere sicuri. Se anche il 53 non si mostrasse in due ruote, o se invece uscisse in tutte, non dovremmo stupirci troppo.)

Ma in che senso l'applicazione comune della legge dei grandi numeri è sbagliata? Semplicemente perché non è vero che la legge preveda il pareggio, neppure approssimato. Come abbiamo visto, ci dice solo che con probabilità sempre maggiore il numero di successi resterà entro uno scarto *relativo* prefissato. Ho messo in corsivo la parola "relativo" perché questa è la chiave (anzi *una* chiave) del discorso. Nel nostro esempio avevamo scelto il 10%, ma avremmo anche potuto prendere l'1%, o l'uno per mille, o quello che volete: ma sempre in termini relativi. Se pensiamo ad es. all'1%, la legge dei grandi numeri ci dice che al crescere di n sarà sempre più probabile che il numero di rossi non scarti più dell'1% dalla metà del numero di lanci.

Se i lanci sono 1000, e i rossi attesi sono quindi 500, stiamo parlando di un numero di rossi compreso tra 495 e 505, ossia con una variazione di 5 in più o in meno; però su un milione di lanci l'intervallo va da 495 000 a 505 000, e la variazione è ora di 5000 in più o in meno.

In realtà la teoria matematica della probabilità ci mette a disposizione uno strumento più potente: il cosiddetto "teorema centrale del limite." Sono sicuro che non me ne vorrete (anzi!) se non ne do un preciso enunciato: mi limito solo a ricavarne una conseguenza particolare per il nostro esempio. Se il numero n di prove è molto grande, si può dire con buona approssimazione (sempre migliore quanto più grande è n) che con probabilità circa 68% il numero di rossi non scarterà da $n/2$ più di $\sqrt{n}/2$.

Vediamo qualche caso: se $n = 100$, $\sqrt{n}/2 = 5$: c'è quindi una probabilità del 68% che si vedano tra 45 e 55 rossi. Se invece $n = 1000$, $\sqrt{n}/2 \simeq 16$, per cui con la stessa probabilità ci aspettiamo tra 484 e 516 rossi. Se infine $n = 10\,000$, con la solità probabilità del 68% i rossi saranno tra 4950 e 5050. Come si vede, l'intervallo per quella data probabilità si allarga al crescere di n, ma in termini relativi si restringe: si passa da 10 su 50 a 100 su 5000.

* * *

Questo è quanto ci dicono i teoremi, ossia la matematica. Ma ci sono un paio di commenti da fare. Il primo riguarda un punto che ho già toccato, quando ho scritto che la legge dei grandi numeri (o qualunque altro teorema) non dice che se un evento ha probabilità molto vicina a 1, esso si verificherà quasi certamente, ecc.

Mi sembra quasi di sentirvi: "Come come? Ma che stai dicendo! La legge dei grandi numeri non dice che se un evento ha probabilità p, è molto probabile che in n prove si verifichi un numero di volte vicino a np? E allora, se p è molto vicino a 1, il numero di successi sarà molto vicino a n, ossia l'evento si verificherà quasi sempre!"

Piano! L'avevo detto che c'era un gioco di parole... Quello che la legge dei grandi numeri dice è che *la probabilità* che l'evento si verifichi un numero di volte vicino a n è *vicina a* 1. Ma che se un evento ha probabilità vicina a 1 sia quasi certo che lo vedremo verificarsi, *non lo dice il teorema.*

Questo è un punto centrale, e perciò occorre esaminarlo per bene. La legge dei grandi numeri, come qualsiasi altro teorema sulle probabilità, in ultima analisi fornisce (sotto certe ipotesi) dei valori o dei limiti per certe probabilità. Ma si tratta sempre e soltanto di risultati *matematici*, che in quanto tali non hanno il potere o la pretesa d'*influire sul mondo reale.*

Per realizzare il collegamento tra teoria matematica e realtà abbiamo bisogno di qualcosa di diverso: di un enunciato *a fondamento empirico*, che dica più o meno questo:

> "se in un esperimento sono verificate le condizioni che corrispondono alle ipotesi di un certo teorema di probabilità, allora ci dobbiamo aspettare che il numero di successi non si scosti in modo significativo dal valore previsto dal teorema."

È quello che a volte prende il nome di "postulato empirico del caso," e che dal punto di vista epistemologico ha il ruolo cruciale d'istituire il ponte fra teoria ed esperienza.

Ho detto che è un punto centrale, ed è un punto che assai spesso non viene capito, con la conseguenza di attribuire al calcolo matematico un potere magico: appunto la capacità d'influire sulla realtà. Posso illustrarlo con un aneddoto che ho letto di recente.

A proposito del famoso 53, un matematico racconta che un amico gli avrebbe tolto il saluto perché lui si rifiutava di ammettere che in quanto matematico *sapeva* quando il 53 sarebbe uscito. All'obiezione "ma se fosse così, tutti i matematici giocherebbero al Lotto e vincerebbero" l'amico rispondeva che evidentemente ai matematici era vietato giocare al Lotto. Insomma, una specie di setta partecipe di un sapere superiore, ma in compenso vincolata da un giuramento di astinenza...

Non so se la storia sia vera, ma come si suol dire, se non è vera è ben trovata...

* * *

E ora il secondo commento. Come tutti sanno, i teoremi hanno delle ipotesi, e nel nostro caso le abbiamo già enunciate. Ve le ripeto:

a) l'esito di ogni lancio è *casuale*, e nero e rosso sono *equiprobabili*;

b) l'esito di ciascun lancio è *indipendente* da quello degli altri lanci.

Ora dobbiamo discutere queste ipotesi, che è la parte più interessante, perché *non riguarda più la matematica.*

"Come sarebbe!" diranno i miei piccoli lettori… No, scusate, quello era Pinocchio; noi siamo seri (sebbene anche *Pinocchio* sia un libro serissimo) ma soprattutto non diciamo bugie… Comunque "come sarebbe!" qualcuno di voi l'avrà pensato di certo: "Se la probabilità è una teoria matematica, anche le sue ipotesi apparterranno alla matematica. Non può essere altrimenti… " E invece no!

Mi spiego subito: le ipotesi, intese come proposizioni che si mettono in cima al teorema, fanno certo parte della teoria; ma non fa parte della teoria, quindi della matematica, la *verifica* di quelle ipotesi.

Nel nostro solito esempio: chi ci dice che rossi e neri in un lancio sono equiprobabili? Chi ci dice che i singoli lanci sono indipendenti? Messa così, mi sembra ovvio che la risposta a queste domande dovete cercarla dal fabbricante della roulette, o dal croupier, o magari da entrambi… Oppure da qualcuno che presa nota di tutte le condizioni costruttive e operative del sistema, lo studia mettendo all'opera le sue cognizioni su quel genere di sistemi. E siccome si tratta di un sistema meccanico, è più probabile che la competenza giusta si trovi in un fisico.

Di più: ci dobbiamo addirittura chiedere se sia lecito studiare i lanci alla roulette come un fenomeno casuale. Lo so, questo sembra ovvio; ma da un punto di vista generale (e scientifico, soprattutto) bisogna pur darne una giustificazione. Il che ci porterebbe però a un problema ancora più arduo: definire chiaramente che cosa intendiamo per casuale… Tamto arduo, che per ora preferisco lasciarlo da parte.

Per risvegliare l'interesse di chi legge, che potrebbe essersi un po' assopito se non è un patito dei giochi d'azzardo, esaminiamo lo stesso problema per il più classico e banale esempio preso dalla genetica. C'è un teorema, dovuto credo a Hardy (1908), che mostra come nell'ipotesi di accoppiamento casuale le frequenze dei diversi fenotipi raggiungono un equilibrio già con la prima generazione. Immagino che questo risultato sia arcinoto a chi ha studiato genetica; io invece l'ho trovato in un classico testo di probabilità: quello di Feller.

Bene. Come sappiamo, un teorema non si discute: dall'ipotesi segue necessariamente la tesi… Ma che dire della validità dell'ipotesi? Nel nostro caso, chi ci dice se l'accoppiamento sarà casuale oppure no? Questo non ce lo dirà di certo un matematico: a seconda delle specie coinvolte e delle condizioni (sperimentali o naturali) dovremo rivolgerci a vari specialisti: agrari, etologi, ecc. Lo stesso Feller, che pur essendo un matematico aveva tutto il necessario buonsenso, lo dice chiaro: nel solito esempio mendeliano dei piselli rossi e bianchi, tutto sta a vedere come sono stati piantati. Se li abbiamo separati spazialmente: quelli bianchi in un campo, e quelli rossi ben distanti, è difficile che si possa parlare di accoppiamento casuale. (A proposito: mi accorgo di non sapere chi si fa carico dell'impollinazione dei piselli. Qualche insetto, suppongo, ma quali? E quanto volano lontano?)

Non sarà casuale l'accoppiamento di specie che sono sottoposte a selezione da parte di allevatori; ma anche nel caso vi siano preferenze sessuali diverse per i diversi fenotipi...

Insomma, queste cose le sapete meglio di me; ma il punto che volevo sottolineare è uno solo: se si possa parlare di eventi casuali, e con quali probabilità, può dircelo solo un esperto che sia informato sull'effettivo andamento del fenomeno in esame.

Detto questo, per tornare alla roulette l'ipotesi di casualità ed equiprobabilità per rosso e nero è certamente ben giustificata, per come sono costruite le roulette e per come vengono effettuati i lanci. La cosa non è ovvia, ma si può dimostrare con un adeguato studio, come ho già detto: quindi riteniamoci soddisfatti.

* * *

Avrete notato che in questa chiacchierata sulla probabilità ho sollevato alcuni problemi che mi sono guardato bene dal risolvere, anzi dall'affrontare: primo fra tutti, che cosa esattamente significa "casuale"? E a dire il vero, non ho neppure risposto esaurientemente alla domanda iniziale: che cosa c'è sotto la convinzione che un numero ritardatario al Lotto abbia maggiore probabilità di uscire? Ho solo, per così dire, preparato il terreno... Ma siccome la lunghezza di questa puntata è già sufficiente, rimandiamo il seguito a un'altra occasione.

* * *

Nella puntata precedente, quasi alla fine, è saltato un corsivo. Di conseguenza il capoverso che inizia con "Soltanto l'educazione" e finisce con "prospettiva per l'avvenire" sembra scritto da me, mentre invece è ancora di Augé. Per chiarezza riporto l'intera citazione di Augé:

[...] *Il secondo è che il suo sistema scolastico* [dell'Europa] *si presenta vieppiù non come strumento di uguaglianza delle opportunità, ma come riproduttore delle disuguaglianze.* [...] *Questo rischio di una frattura irreversibile renderebbe impossibile la costituzione di un'umanità unificata, di un'umana società, o, più specificamente, darebbe alla società planetaria in formazione un volto inquietante e profondamente antidemocratico.*

Soltanto l'educazione può scongiurare questo rischio. Ma dare la priorità all'educazione implicherebbe, a livello di ciascun paese e a livello mondiale, un investimento finanziario senza precedenti, infinitamente superiore a quello oggi esistente anche nei paesi più sviluppati. Si tratterebbe, in altre parole, di una rivoluzione a due tempi, basata sull'ipotesi che a essere in grado di aiutare i paesi meno sviluppati quanto all'educazione, sarebbero soltanto quei paesi che avessero vinto loro stessi la rivoluzione dell'educazione. L'obiettivo ultimo sarebbe aggregare l'umanità intera all'avventura scientifica, la quale, infine, cambierebbe ogni sua prospettiva per l'avvenire.