Quarto incontro:

applicazioni della passeggiata casuale.

Lo scopo di questa lezione è di fare da tramite tra le passeggiate casuali (in matematica) e quei fenomeni fisici che vanno sotto il nome di diffusione.

La lezione si divide in tre momenti: il primo conclude la lezione sulle passeggiate casuali, il secondo riprende il teorema centrale del limite ed il terzo momento introduce al fenomeno fisico noto come moto browniano.

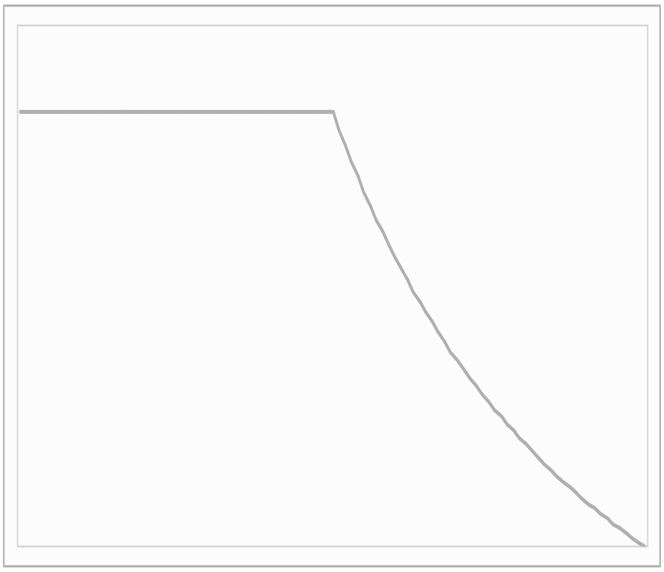
1. Rovina del giocatore

Un ubriaco si trova ad un passo dal burrone: un solo passo verso sinistra e si ritrova in fondo alle rocce. Ad ogni passo, ha una probabilità q (diciamo, ad esempio 1/3) di andare verso sinistra e p (nell'esempio, sarà 2/3) di andare verso destra. Qual è la probabilità che si salvi, se continua a camminare per un tempo lunghissimo?
Un possibile approccio è quello di considerare l'albero delle possibili traiettorie, che assomiglia ad una passeggiata casuale con tutti i rami a sinistra di 0 "potati". Questo si analizza bene per un numero limitato di passi

vi è una sola traiettoria che cade dopo un passo, ed ha probabilità q
vi è una sola traiettoria che cade al terzo passo (destra, sinistra e sinistra) ed ha probabilità pq^2
vi sono due traiettorie che cadono al quinto passo ed ognuna contribuisce con una probabilità p^2q^3
…..

Possiamo continuare su questa strada (difficile) oppure tovare un modo alternativo per risolvere il problema.
Indichiamo con P_1 (l'indice in basso rappresenta il punto da cui iniziamo la passeggiata) la probabilità di arrivare a 0 ("cadere nel burrone"). Questo numero deve dipendere da p: se ad esempio scegliamo $p = 0$, ossia andiamo sicuramente verso il burrone, la probabilità di cadere sarà 1 (qualsiasi sia il punto di partenza). Viceversa, se scegliamo $p = 1$, andiamo sicuramente verso destra, quindi non potremo mai cadere, ossia $P_1 = 0$. Ora vogliamo trovare un modo per esprimere P_1 rispetto a p.
Le passeggiate che partono da 1 e arrivano, prima o poi, a 0 (le indicheremo con $(1 \Longrightarrow 0)$) si dividono in due classi disgiunte: quelle che ci arrivano al primo passo (useremo la freccia singola per indicare cosa succede in un solo passo) e quelle che al primo passo sono in 2, quindi arrivano, prima o poi, a 0: in simboli, $(1 \Longrightarrow 0) = (1 \longrightarrow 0) \cup (1 \longrightarrow 2, 2 \Longrightarrow 0)$. Guardiamo il secondo evento: una volta che la passeggiata è arrivata in 2, possiamo supporre che si dimentichi da dove è arrivata e che proceda in modo indipendente dal passato. Allora otteniamo la seguente relazione per le probabilità: $P_1 = q + p\, P_2$.
Vediamo come è fatta una passeggiata che parte dal punto 2 e arriva in 0. Questa passaggiata passa, sicuramente, dal punto 1, quindi possiamo scrivere $(2 \Longrightarrow 0) = (2 \Longrightarrow 1, 1 \Longrightarrow 0)$. La prima parte è una passeggiata che – se fosse traslata di un passo a sinistra – porterebbe $1 \Longrightarrow 0$; la seconda parte, come abbiamo già osservato in precedenza, è indipendente dalla prima (una volta arrivati in 1, vado avanti senza ricordare da dove sono venuto. Allora risulta essere $P_2 = (P_1)^2$, ed allo stesso modo $P_k = (P_1)^k$.
Possiamo allora impostare l'equazione del secondo ordine $P_1 = q + p\, (P_1)^2$, da cui si ottengono le soluzioni 1 e q/p. Non possiamo avere due numeri per una sola probabilità: quale delle due è quella

giusta? Si vede che se $p < ½$, q/p è maggiore di 1, quindi possiamo escludere questo ramo e dobbiamo accettare che $P_1 = 1$. Se $p = ½$, $q/p = 1$, quindi ancora la soluzione vale $P_1 = 1$. Infine, se accettiamo che P_1 sia una funzione continua di p (si può dimostrare, in realtà, che questo è il caso), la soluzione corretta per $p > ½$ deve essere q/p (che vale 0 per $p = 1$). In definitiva, il grafico della funzione P_1 in termini di p è dato nel seguente grafico



Se $p \leq ½$ si ottiene $P_k = 1$ per ogni k: qualunque sia il punto di partenza, certamente arrivo a 0 prima o poi. Che questo valga anche per una passeggiata simmetrica è sorprendente! Se $p > ½$, si ottiene $P_2 = (q/p)^2$ ed in generale $P_k = (q/p)^k$.

A questo punto possiamo chiederci cosa succede se la porta di casa si trova da qualche parte, ossia se esiste una barriera superiore. Questo cambia la nostra probabilità di salvarci?
La formula è piuttosto complicata, ma sapere che esiste è comodo. Possiamo usarla in alcuni esercizi, ad esempio: se dobbiamo raddoppiare il nostro capitale, conviene giocare un punto alla volta oppure tutto insieme?
Supponiamo che la porta si trovi a una distanza N dal burrone. Supponiamo di partire da un punto a distanza x e ci chiediamo quanto vale la probabilità R_x di arrivare al burrone, ovvero quanto vale la probabilità $S_x = 1 - R_x$ di arrivare alla porta. Si ha $R_0 = 1$, $S_N = 1$; come calcolare cosa succede negli altri casi? Prendiamo le traiettorie ($x \Longrightarrow 0$): queste si possono dividere in due classi, quelle che prima di arrivare al burrone passano dal punto N e quelle che non ci arrivano mai, ossia ($x \Longrightarrow 0$) = ($x \Longrightarrow N$, $N \Longrightarrow 0$) ∪ ($x \Longrightarrow 0$ senza passare da N). In termini di probabilità scriviamo allora

$$P_x = S_x P_N + R_x = S_x P_N + (1 - S_x)$$

Se P_N è diverso da 1, ossia se $p > ½$, otteniamo

$$S_x = \frac{1-(q/p)^x}{1-(q/p)^N}$$

Se p < ½, un ragionamento basato sul problema inverso (scambiando la posizione della porta e del burrone e facendo attenzione a spostare il punto di partenza!) porta alla formula

$$S_x = \frac{(p/q)^{N-x}-(p/q)^N}{1-(q/p)^N}$$

che si dimostra essere – dopo un poco d'algebra – uguale a quella precedente. Rimane fuori il caso simmetrico, in cui si ha invece $S_x = x/N$.

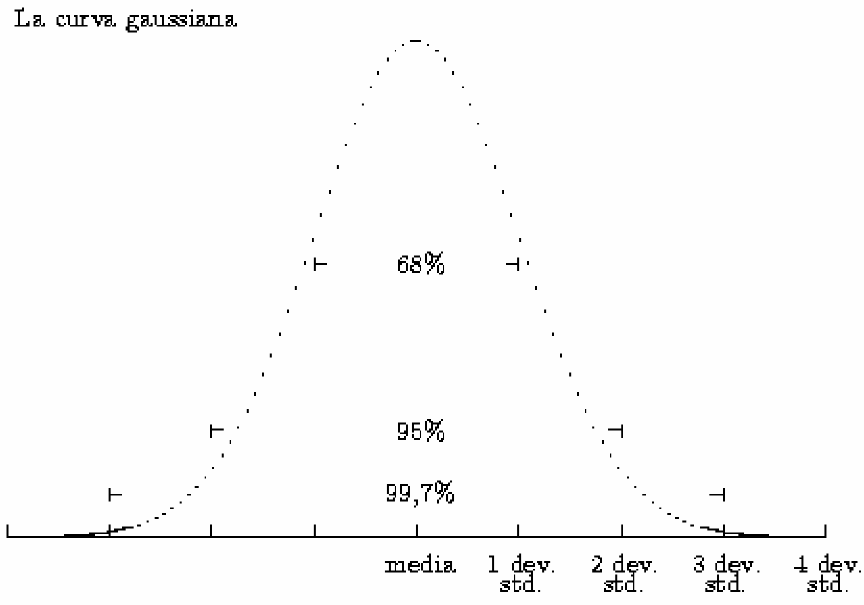
Siamo a Las Vegas, è notte ed abbiamo in tasca solo 1000€. Solo?!? Sì, perché domani abbiamo appuntamento con una banda di gangster e, se non gli consegnamo 2000€, ci uccideranno. Davanti a

noi, le porte di un casinò. Dobbiamo entrare e provare a giocare, ma un dubbio ci assale: conviene giocare 1€ alla volta, oppure tutto insieme?
È un problema di passeggiata casuale con barriere, in cui p (la probabilità di vittoria) è 9/19, $x = 1000$, $N = 2000$ e quindi la probabilità di raddoppiare il capitale si calcola con la formula precedente e si ottiene praticamente 0 (siamo nell'ordine di 10^{-45}). Dobbiamo confrontare questo numero con la probabilità di raddoppiare il capitale in un'unica puntata che è 9/19 ossia circa il 47%. Questo risultato è sorprendente per molti, che ritengono che il gioco conservativo sia più conveniente del gioco veloce.

2. Distribuzione gaussiana

La curva a campana: si chiama anche curva di Gauss, ma effettivamente risale al secolo precedente; fu un matematico francese, fuoriuscito in Inghilterra per motivi religiosi, di nome Abraham De Moivre, a scoprire la "legge normale degli errori": una somma di piccoli errori casuali dà come prodotto un errore la cui distribuzione si può approssimare con una legge normale (o gaussiana).
Cosa dobbiamo sapere della curva a campana? Che si determina tramite due parametri (media e varianza), che la varianza determina di quanto mi posso allontanare dalla media (in un certo senso). Così, dobbiamo ricordare che

entro una deviazione standard si concentra il 68% della probabilità,
entro due deviazioni standard si concentra il 95% della probabilità, e
entro tre deviazioni standard si concentra il 99,5% della probabilità.



2.1 Capire i sondaggi.

Se dovessimo definire un sondaggio, dovremo parlare di un'indagine attuata su un *campione* di persone per arrivare a conoscere l'opinione dell'intera popolazione su una determinata questione.
Quali sono le caratteristiche del campione? Deve essere rappresentativo (fotografa bene la popolazione nel complesso) ed eterogeneo (vi è una buona variabilità all'interno del campione). Il campione inoltre deve essere abbastanza numeroso da metterci al riparo da errori troppo grandi.

Generalmente, le possibili risposte sono due: *si* oppure *no*. Dovremo allora mettere in relazione la proporzione campione (il numero di *si* ricevuto nel sondaggio diviso per la numerosità del campione) con la proporzione reale di *si* nella popolazione.
Se il campione è rappresentativo, in media (ossia, aumentando la numerosità del campione) i due valori si avvicinano. Noi però ci fermiamo a un campione limitato (mica possiamo intervistare tutti!) e dobbiamo chiederci: qual è l'errore che compiamo?
Se il campione è grande, ma ancora piccolo rispetto alla popolazione, allora la deviazione standard della proporzione campione è pari a $\sqrt{\frac{p(1-p)}{N}}$ dove p è la proporzione campione, N la numerosità del campione. Il numeratore si può massimizzare prendendo $p = q = ½$, e così faremo (per semplicità); allora la deviazione standard è pari a $1/2\sqrt{N}$. Questo significa che con una probabilità pari al 95% (questo valore si chiama livello di fiducia, o affidabilità, del sondaggio), il valore della proporzione reale è compreso nell'intervallo

$$\text{proporzione campione} \pm 1/\sqrt{N}$$

Per comprendere bene un sondaggio, quindi, dovremo sapere: qual è il valore stimato, qual è l'ampiezza dell'intervallo di confidenza, qual è il livello di fiducia del sondaggio (oppure, qual è la numerosità del campione, in modo da ricavare da soli questo valore).

Prendiamo ad esempio il sondaggio, annunciato pochi giorni fa sul Corriere della Sera, secondo cui il centrosinistra ha un vantaggio di 4 punti percentuali, con un errore di ± 1%. In altre parole, la proporzione campione per il centrosinistra è $p = 0{,}52$ con un errore di ± 0,5%. In questo caso, manca un dato! Supponiamo che la società di sondaggi sia convinta di avere ragione con una probabilità del 95%: in tal caso, con quel livello di fiducia, possiamo dire che la proporzione reale nella popolazione è compresa tra il 51,5% e il 52,5%. Possiamo anche dire qualcosa sulla numerosità del campione: se il livello è pari al 95%, allora dalla relazione $\sqrt{N} = 200$ risulta $N = 40000$ (a me sembra eccessivo: probabilmente hanno scelto un livello di fiducia più basso data la distanza dalle elezioni?!?)

3. Introduzione al moto browniano.

Il moto di polveri (resine o pollini) sospese in un liquido, osservato già nel XVII secolo, viene comunemente associato al nome del botanico scozzese Robert Brown, che lo descrisse nel 1827, avendolo studiato utilizzando un microscopio ad alta risoluzione.
Inizialmente, venne attribuito alla presenza di materia organica, vivente, nel liquido; in seguito, ci si accorse che questa spiegazione non reggeva, in quanto ogni corpo inanimato, purché sufficientemente piccolo, acquistava un moto simile a quello del polline. Ci si convinse, quindi, che era necessaria una spiegazione meccanica del fenomeno ed effettivamente diverse teorie apparvero nelle riviste scientifiche dell'epoca. Poi, nel 1905, Albert Einstein pubblicò la sua teoria sul moto browniano (questo venne pubblicato sulla stessa rivista, gli Annalen der Physik, in cui comparvero sia l'articolo sulla relatività ristretta, sia quello sull'effetto fotovoltaico che gli portò il Nobel). La presenza di questa numerosa letteratura sull'argomento rende ancor più sorprendente il fatto che Einstein non fosse in precedenza a conoscenza del fenomeno: nelle sue note autobiografiche, lo stesso Einstein scrive «in questo lavoro il mio obiettivo principale era di trovare fatti che avrebbero garantito il più possibile l'esistenza di atomi di dimensione finita. Nel fare questo scoprii che, in accordo con la teoria atomistica, avrebbe dovuto essere possibile l'osservazione del movimento di particelle microscopiche sospese, senza sapere che quell'osservazione, che riguardava il moto browniano, era nota da molto tempo».

Le fluttuazioni del moto browniano possono essere rivelate dal movimento di particelle immerse in un fluido, a patto che le loro dimensioni non siano troppo grandi rispetto al cammino medio delle molecole del fluido. Se si verifica questa condizione, la particella subirà un numero casuale di urti che la costringeranno a muoversi seguendo una traiettoria irregolare e non prevedibile. Il moto diventerà tanto più caotico quanto più alta è la temperatura del sistema e tanto più piccole sono le dimensioni delle particelle; viceversa, il movimento sarà minore in un fluido con alta viscosità.

3.1 Distanza media dall'origine.

Supponiamo che la polvere in sospensione percorra sempre la stessa distanza nel periodo tra due urti; ci chiediamo se sia possibile simulare il suo cammino in modo da poterne ricavare un modello. Supporremo che le collisioni avvengano in maniera indipendente le une dalle altre e che il movimento lungo l'asse orizzontale sia indipendente da quello lungo l'asse verticale. Possiamo allora mettere in relazione questo modello con quello della passeggiata casuale.
Tracciamo un reticolo quadrato, in cui ci muoveremo lungo le diagonali. Ad ogni passo, lanciamo due dadi, che ci daranno informazioni sui movimenti, rispettivamente, dei due assi (ad esempio, pari indica salita e dispari discesa). Dopo N passi, avremo raggiunto un punto P che dista D dall'origine; il nostro scopo è di mettere in relazione D con N.
Se si eseguono un elevato numero di prove è possibile farci un'idea empirica del valore medio di questa distanza, che segue la legge $<D> \approx \sqrt{N}$ (su un foglio a quadretti da 5mm, proviamo a fare 10, 20 e 40 passi: la media dovrebbe solo raddoppiare dalla prima alla terza misura).
Possiamo dimostrare questa legge? Conviene studiare la distanza quadrata; è facile vedere che $D^2 = S_X(N)^2 + S_Y(N)^2$ e che quindi vale $<D^2> = 2 <S(N)^2> = 2\,N$.

3.2 Probabilità di ritorno all'origine.

Per una passeggiata casuale in una dimensione, qual è la probabilità di ritornare all'origine in un qualche momento del futuro?
Vediamo cosa succede al primo passo: andiamo in 1 con probabilità p e in -1 con probabilità q. Da 1, torniamo nell'origine con probabilità 1, se $p \leq q$, con probabilità q/p altrimenti; da -1, la situazione è simmetrica, e la probabilità di ritorno è 1 se $q \leq p$, p/q altrimenti. Allora, se $p = q = ½$, la probabilità di ritorno all'origine è $½ \cdot 1 + ½ \cdot 1 = 1$; se $p < q$, la probabilità è $p \cdot 1 + q \cdot p/q = 2p < 1$; se $p > q$ risulta infine $2q < 1$.
Ora siamo interessati alla passeggiata casuale in dimensione 2, e la precisione del risultato precedente (è ricorrente se e solo se $p = q = ½$) dà sicuramente adito a molti dubbi che possa essere così anche in questo caso.
Avremo bisogno di tutti gli strumenti introdotti finora. Dalla *legge geometrica* (il tempo di attesa del primo successo) otteniamo che la probabilità di successo è l'inverso della media. Dalla passeggiata casuale, la probabilità di essere in 0 dopo 2n passi è Bin($2n,n$) $(1/2)^{2n}$ e questo valore, per n grande, si approssima con $1/\sqrt{\pi n}$.
Indichiamo con P la probabilità di ritornare all'origine: allora la probabilità di tornare x volte è P^xQ, il numero medio di ritorni μ è uguale a $1/P$, quindi tutto ciò che rimane da fare è calcolare μ.
Il numero di ritorni a 0 di una traiettoria è data dalla somma $X_2 + X_4 + \ldots$, dove le X_{2n} contano se siamo passati da 0 al passo n: valgono 1 in quel caso e 0 altrimenti.
Vediamo che la probabilità di essere in 0 dopo 2, 4,… passi è pari al prodotto delle probabilità di ritorno in 0 delle due passeggiate casuali (che sono indipendenti). Quindi la probabilità di ($X_{2n} = 1$) – che è uguale alla sua media – è pari a $[\text{Bin}(2n,n)\,(1/2)^{2n}]^2$. Quello che rimane da fare è la somma, senza spaventarci: a noi basta sapere se la somma è finita o infinita, quindi basta verificarlo per la somma dei valori approssimati; ma la serie con termine $1/n$ diverge a infinito, quindi il numero medio di ritorni è infinito e $P = 1$.