La ricerca dell'anima gemella

Eva **Filoramo**

Quando il celebre astronomo tedesco Johannes Kepler, italianizzato in Keplero, per colpa del colera perse la prima moglie, nel 1611, si trovò di fronte alla necessità di procurarsi una seconda moglie che si occupasse di lui, della casa e dei figli; si dedicò pertanto alla ricerca della migliore candidata con la stessa attenta metodicità con cui aveva esaminato i dati utilizzati per determinare la forma ellittica dell'orbita di Marte intorno al Sole.
Dal momento che il suo primo matrimonio, combinato da altri, non era stato particolarmente felice, questa volta lo scienziato era determinato a scegliere nel modo migliore possibile. Come scrisse in una lettera a un amico, il barone Strahlendorf, il processo di selezione durò quasi due anni e coinvolse, nel tempo, ben undici candidate; alla fine, nonostante il parere contrario di molti suoi amici, la prescelta fu la quinta. La donna, sebbene fosse orfana, di umili origini e del tutto priva di dote, secondo Keplero compensava tali carenze con la propria educazione; la bontà della scelta fu poi confermata dal fatto che insieme ebbero sette figli e che negli anni successivi la sua gestione della casa e della famiglia consentì all'astronomo di potersi dedicare con successo ai propri studi.

Il problema della segretaria

La scelta di Keplero avrebbe potuto costargli molto meno tempo e fatica, tuttavia, se soltanto l'astronomo fosse stato a conoscenza del cosiddetto "problema della segretaria", portato alla ribalta nel 1960 dal grande divulgatore della matematica Martin Gardner.
Possiamo riformulare il problema della segretaria in molti modi diversi; originariamente Martin Gardner lo descrisse nella rubrica *Mathematical Games* della prestigiosa rivista "Scientific American", da lui curata per quasi trent'anni; lo battezzò "il gioco del googol" (il motore di ricerca ovviamente non esisteva ancora! "Googol" è un numero e indica un 1 seguito da cento 0, ed è a questo che si sono ispirati Sergey Brin e Larry Page), ma la filosofia era molto simile a quella che andremo a descrivere adesso ispirandoci alla storia di Keplero.
Nel nostro esempio, il non più giovanissimo Giovanni vorrebbe trovare finalmente l'anima gemella, e per risparmiare tempo ha deciso di affidarsi a un sito specializzato in incontri romantici. Secondo le istruzioni ricevute al momento dell'iscrizione, Giovanni deve scegliere, arbitrariamente, un numero n di donne che è disposto a incontrare in altrettanti appuntamenti; dopo ciascun appuntamento avrà modo di dare un voto alla candidata incontrata e dovrà scegliere la propria anima gemella (la candidata migliore fra tutte) sapendo che, una volta rifiutata, una donna non potrà più essere richiamata in causa (le donne, si sa, sono molto orgogliose).
Immaginiamo allora che Giovanni vada al primo appuntamento, e che sia un vero disastro; il secondo appuntamento va leggermente meglio, ma non scocca nessuna scintilla. Anche il terzo appuntamento, pur senza essere catastrofico come il primo, non è particolarmente esaltante e il voto assegnato alla candidata è chiaramente insufficiente. A questo punto, Giovanni si chiede scoraggiato se incontrerà mai la donna giusta per lui e teme di dover effettivamente andare a tutti gli n appuntamenti, perché la sua anima gemella potrebbe celarsi addirittura all'ultimo posto dell'elenco; al contempo, tuttavia, Giovanni ha paura di andare troppo avanti nel numero di donne da incontrare e di rifiutare anche la donna più giusta per lui (non potendo, come abbiamo specificato, ritornare sui propri passi e "ripescare" una candidata in precedenza scartata). Riflettendo, Giovanni capisce che deve esistere sicuramente un numero ottimale di donne da incontrare dopo il quale può avere la maggiore probabilità di fare la scelta migliore possibile.
Fortunatamente, questo sito specializzato in incontri romantici si vanta di usare un sistema statisticamente ideale per trovare il partner migliore. Secondo questo sito, è infatti sufficiente incontrare – e respingere – il 37% delle candidate per avere una probabilità di successo del 37% di incontrare, dopo queste prime candidate, la vera anima gemella. Se allora Giovanni, per semplicità, ha stabilito inizialmente di incontrare 100 donne in altrettanti appuntamenti, significa che sarà per lui sufficiente incontrare le prime 37, scartarle tutte e, sulla base del voto dato a ciascuna di queste prime candidate, scegliere tra le restanti la prima donna a cui ha dato un voto più alto di tutti quelli assegnati finora, e tenersela ben stretta.

Perché proprio il 37%?

Per dimostrarlo, e convincerci del fatto che anche noi, se vogliamo, possiamo affidarci al metodo proposto, immaginiamo per il momento di non sapere quanto vale questo numero e definiamo $P(k)$ come la probabilità di scegliere la migliore candidata; k è, per definizione, il momento giusto per fermarsi, ossia il momento in cui diventa possibile scegliere la candidata migliore rispetto a quelle incontrate precedentemente. La probabilità $P(k)$ è data, di volta in volta, da di due probabilità: la probabilità che la prescelta sia all'*n-simo* posto nella lista di donne da incontrare e la probabilità che Giovanni scelga

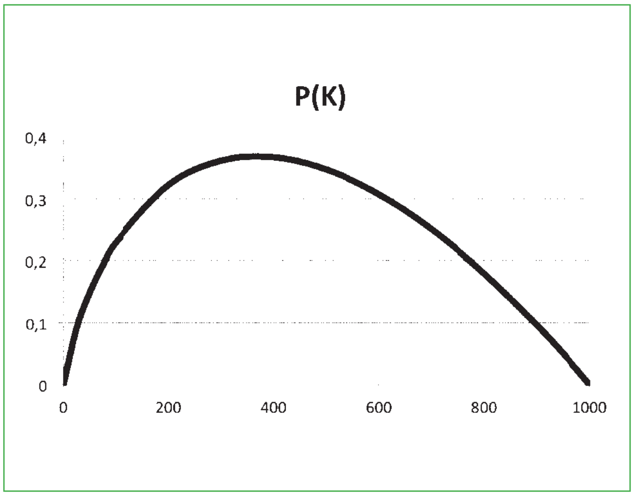
proprio quella donna tra tutte le candidate. Ricordiamo che, secondo la formula proposta del sito, Giovanni deve incontrare k donne e valutarle, dopodiché scegliere la prima donna che, dal $k+1$-*simo* appuntamento compreso in poi, ottiene un voto migliore di tutte quelle incontrate fino ad allora.
Su n candidate, la probabilità a priori che la prescelta si trovi all'n-*simo* posto della lista è sempre pari a $1/n$; la probabilità di scegliere proprio quella all'n-*simo* posto, invece, è 0 fino al k-*simo* appuntamento, perché per ipotesi fino al k-*simo* appuntamento Giovanni non sceglie nessuna candidata. Se Giovanni sceglie la donna incontrata all'appuntamento $k+1$, la probabilità che, trovandosi in quella posizione, sia la prescelta, è pari a 1, sempre perché fino al k-*simo* appuntamento Giovanni non ha scelto nessuna donna. Una volta arrivato all'appuntamento $k+2$, tuttavia, questa probabilità decresce, perché Giovanni potrebbe aver scelto la donna dell'appuntamento $k+1$ con probabilità pari a $1/k+1$; la probabilità che scelga la donna dell'appuntamento $k+2$, allora, è pari a $(1\text{-}1/k+1)=(k/k+1)$ e così via, fino ad arrivare all'n-*simo* appuntamento e a una probabilità pari a $(k/n\text{-}1)$. Componendo, appuntamento dopo appuntamento, le due probabilità, si ottiene la seguente formula:

$$P_n(k)=\left(\underbrace{\frac{1}{n}\times 0+\ldots+\frac{1}{n}\times 0}_{primi\_k\_appuntamenti}+\underbrace{\frac{1}{n}\times 1}_{k+1-esimo}+\underbrace{\frac{1}{n}\times\frac{k}{k+1}}_{k+2-esimo}+\ldots+\underbrace{\frac{1}{n}\times\frac{k}{n-1}}_{n-esimo}\right)=\frac{k}{n}\sum_{j=k+1}^{n}\left(\frac{1}{j-1}\right)$$

Il numero k è, ricordiamolo, quel numero ideale di appuntamenti dopo cui Giovanni smette di scartare le candidate e dà inizio al processo di selezione vero e proprio, scegliendo come partner ideale quella che ottiene il voto migliore rispetto alle donne precedentemente incontrate. Perché sia di qualche utilità, dobbiamo trovare il modo di capire quanto può valere numericamente k sotto certe condizioni.
Quella ricavata sembra una formula molto complicata, ma in realtà è abbastanza facile da calcolare per valori di n bassi; quando n aumenta, la sommatoria diventa approssimabile a un integrale, la cui risoluzione dà come risultato che:

$$P_n(k)\cong\frac{k}{n}\ln\left(\frac{n}{k}\right)$$

Semplificando e ponendo $x=k/n$, sfruttando le proprietà dei logaritmi si ottiene $P(x)\cong -x\ln x$. Abbiamo trovato il modo di semplificare al massimo la funzione di probabilità; vediamo ora come questo può aiutare Giovanni nella difficile scelta.
Torniamo un attimo indietro e, per un momento, dimentichiamo la regole del gioco: se per caso Giovanni, indipendentemente da voti e confronti, avesse scelto la prima donna incontrata dopo il fatidico k-*simo* appuntamento (la donna $k+1$), quale sarebbe stata la probabilità che proprio questa donna fosse la sua anima gemella? Abbastanza piccola, ovviamente. Se avesse rifiutato questa $k+1$-*sima* donna e avesse scelto la $k+2$-*sima*, invece, quale sarebbe stata la probabilità che costei fosse la sua anima gemella? Un po' più alta della precedente, ma sempre abbastanza piccola.



Se proviamo a trasformare questo ragionamento qualitativo in un grafico come quello mostrato in figura 1, emerge chiaramente che, tra il $k+1$-*simo* e l'n-*simo* appuntamento, da qualche parte l'anima gemella di Giovanni deve esistere e, a seconda del valore di k, la probabilità che lui la incontri aumenta fino a raggiungere un picco, per poi ridiscendere inesorabilmente verso lo 0.
Differenziando la funzione di probabilità $P(x)$ e ponendo il risultato uguale a zero, si ottiene che $x=1/e$ ossia circa 0,37; in altri termini, ricordando che $x=k/n$: $k=0{,}37\times n$ proprio come avevamo anticipato. Ecco quindi che la funzione di probabilità raggiunge il valore massimo, prima di iniziare a decrescere, in corrispondenza della candidata numero "trentasette per cento del totale".
Tornando all'incipit di questo articolo, possiamo meravigliarci ancora una volta dell'ingegno di Keplero che, sebbene all'oscuro di questo ragionamento, aveva fatto proprio la scelta giusta: nel caso di 11 candidate, k è infatti pari a 4, e scegliendo la prima candidata migliore delle prime k l'astronomo si comportò in modo da ottimizzare le proprie *chances* anche secondo la statistica.