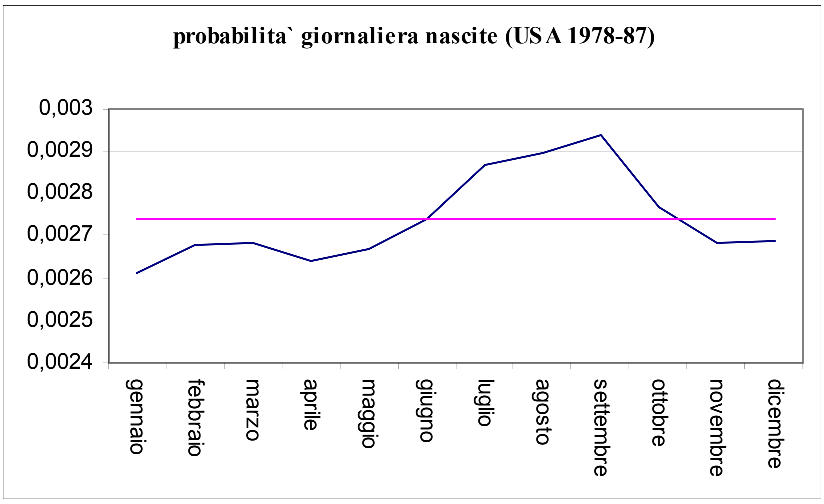
Premessa

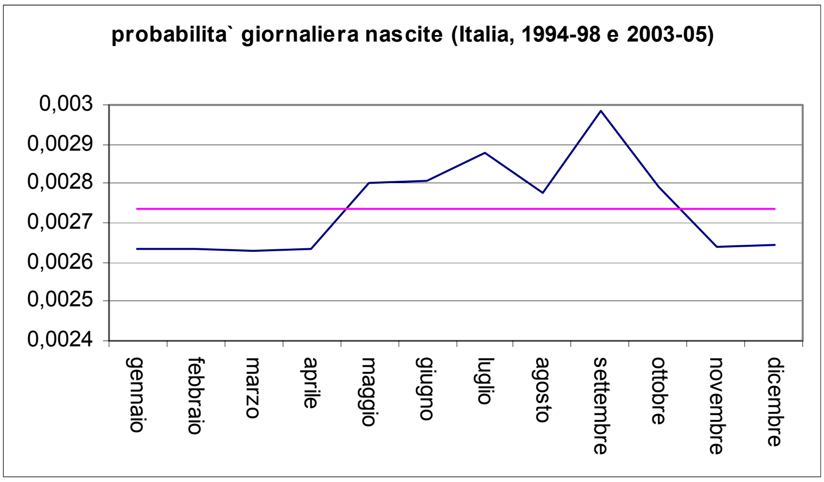
Stephen Jay Gould in [**SG**] sostiene che la mente umana ha serie difficoltà a ragionare in termini probabilistici. Forse per questo motivo, parliamo spesso di sorprendenti coincidenze di fronte ad eventi che, dal punto di vista della probabilità, sono tutt'altro che rari. Uno degli esempi di coincidenza che spesso sorprende la gente riguarda i compleanni: bastano 23 persone per superare il 50% di probabilità di trovarne due che festeggino il compleanno nella stessa data, mentre, se si vuole trovare una persona nata in un giorno prefissato (l'anno non interessa), ne servono 253 per superare lo stesso livello di probabilità. Il cosiddetto Birthday Problem, sebbene noto da tempo, e le sue possibili generalizzazioni, divennero popolari dopo la pubblicazione di *Introduction to Probability Theory and its Applications* di W. Feller [**Fe**], agli inizi degli anni '50 del secolo scorso. In queste pagine si forniscono sia gli elementi teorici, che le realizzazioni numeriche (utilizzando volutamente un software noto a tutti e non dedicato in maniera specifica alla matematica, quale Excel) per la soluzione del problema dei compleanni. Sono presenti numerosi documenti:

1) Il documento ***Compleanni coincidenti*** contiene la soluzione teorica del problema dei compleanni nei seguenti casi:
   a) trovare almeno una persona nata in una certa data prefissata;
   b) trovare (almeno) due o tre o quattro, ecc. persone nate in una data prefissata;
   c) fissate due date, trovare almeno due coppie di cui la prima festeggi il compleanno in una delle due date e la seconda nell'altra;
   d) trovare almeno due o tre o quattro persone nate tutte nella stessa data (non fissata a priori);
   e) trovare almeno due coppie di persone ciascuna con date di nascita coincidenti (non fissate a priori).

   Il documento fa riferimento ai seguenti documenti Excel in cui tali problemi vengono risolti numericamente:
2) ***Compleanni in data fissata.xls***, ***Coincidenze.xls***, ***Due coppie.xls***.
3) Il file ***Complementi*** fornisce le indicazioni per la compilazione dei tre precedenti fogli Excel nei paragrafi 2, 3 e 4; il primo paragrafo, invece, è dedicato ad una breve introduzione degli argomenti di calcolo combinatorio utilizzati nel resto di queste note; l'ultimo paragrafo tratta invece di un argomento, legato ad alcune questioni sollevate nel paragrafo 1, che conferma la scarsa attitudine della mente umana ad affrontare problemi probabilistici: la sostanziale incapacità di disporre "a mano", in modo casuale, 200 punti in un quadrato. Anche quest'ultimo problema viene svolto numericamente nel seguente documento Excel :
4) ***200 punti.xls***

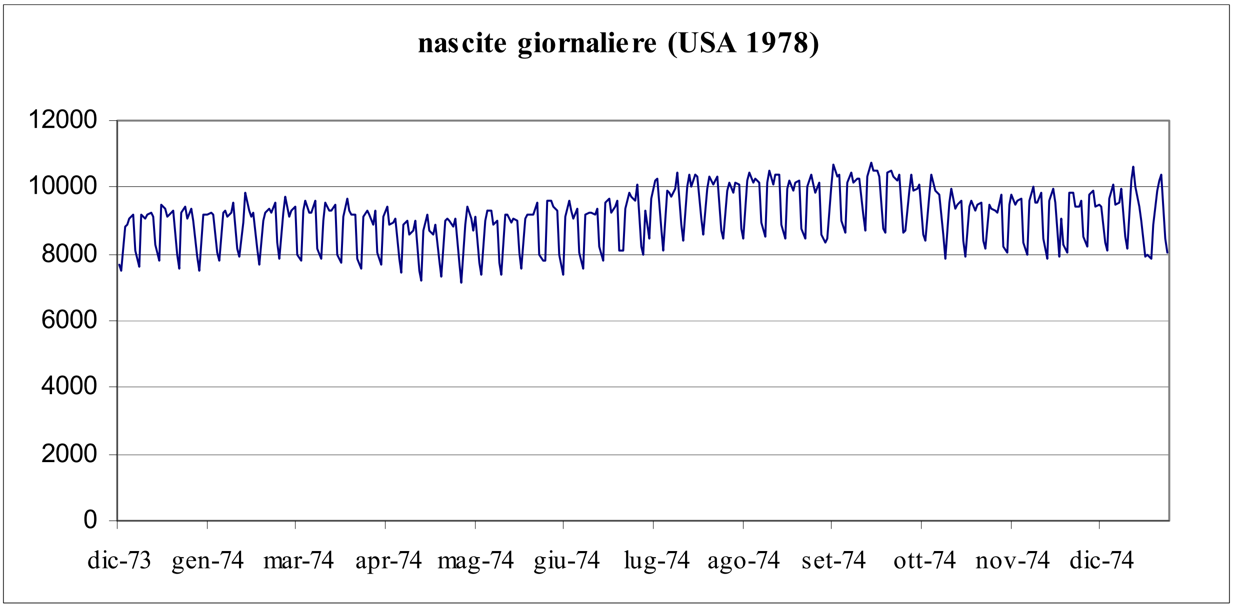
Il problema dei compleanni verrà affrontato adottando alcune necessarie semplificazioni.
Innanzitutto si supporrà che tutti gli anni comprendano 365 giorni: non ammetteremo dunque l'esistenza di anni bisestili, cosa che, d'altronde, viene già fatta in Francia dove, alle persone nate il 29 febbraio degli anni bisestili, viene attribuita come data di nascita quella del 28 febbraio (dal punto di vista della probabilità questo fatto complica un po' le cose invece di semplificarle, ma non entriamo nel dettaglio).
Supporremo poi che tutti i giorni abbiano la stessa probabilità di essere la data di nascita di qualcuno (cioè $\frac{1}{365} = 0{,}0027397...$). Questo in realtà non è vero: da un'indagine svolta negli USA nel decennio 1978-1987 risulta che i compleanni non sono uniformemente distribuiti durante l'anno; per i giorni del mese di settembre la probabilità vale 0,0029407, di oltre il 7% superiore a $\frac{1}{365}$ e di circa il 12,5% superiore alla probabilità media del mese di gennaio [**MT**].





Nel grafico a lato le probabilità giornaliere di nascita nei singoli mesi (linea blu) sono poste a confronto con la probabilità uniforme $\frac{1}{365}$ (linea rossa). La situazione è simile in Italia: esaminando i dati Istat relativi alle nascite mensili nei due periodi 1994-98 e 2003-2005, si ottiene un andamento del tutto analogo, con un picco massimo di probabilità ancora a settembre (0,002983) e un minimo a marzo (0,002627; picco massimo superiore di circa il 13,5% al minimo).

In realtà le cose sono ancora più complicate; se esaminiamo l'andamento giornaliero delle nascite negli USA durante il 1978, scopriamo un fatto curioso: tale andamento sembra mostrare una certa ciclicità che si sovrappone all'andamento mostrato nel precedente grafico:



Se, a prima vista, tale andamento può sembrare inesplicabile, basta contare il numero degli avvallamenti per trovare la spiegazione logica: ormai gran parte delle nascite avviene nelle cliniche e negli ospedali e, in tali strutture, si preferisce non programmare interventi (nascite comprese) durante il fine settimana.

Chi fosse interessato alla trattazione del problema dei compleanni nella sua generalità può consultare [**Nu**]. Noi supporremo sempre che la frequenza delle nascite sia uniforme. Buona lettura

Saverio Giulini

Compleanni coincidenti

I problemi che ci poniamo di risolvere sono del tipo: date r persone, qual è la probabilità di
a) trovare almeno una persona nata in una certa data prefissata;
b) trovare (almeno) due o tre o quattro, ecc. persone nate in una data prefissata;
c) fissate due date, trovare almeno due coppie di cui la prima festeggi il compleanno in una delle due date e la seconda nell'altra;
d) trovare almeno due o tre o quattro persone nate tutte nella stessa data (non fissata a priori);
e) trovare almeno due coppie di persone ciascuna con date di nascita coincidenti.

Bisogna partire da un problema (apparentemente) più semplice: qual è la probabilità che, ponendo a caso r palle in n celle, ne vadano a finire esattamente k in una cella fissata.
La prima palla può andare a finire in una qualsiasi delle n celle, così pure la seconda, la terza e così via. In questo caso associamo ad ogni palla la cella in cui è stata posta: la palla dunque rappresenta la "posizione" a cui l'"oggetto" cella viene associato. Si tratta dunque di disposizioni di n oggetti in r posizioni, il cui numero, come visto in "***Complementi***, 1) Come raggruppare oggetti", è

$$\underbrace{n \times n \times \ldots \times n}_{r \text{ volte}} = n^r .$$

Una delle possibili situazioni in cui $k = 4$ palle vanno a finire nella Cella 3, è illustrata nella seguente tabella:

	Cella 1	Cella 2	Cella 3	Cella 4	Cella 5
Palla 1	×				
Palla 2			×		
Palla 3		×			
Palla 4			×		
Palla 5					×
Palla 6			×		
Palla 7			×		

In quanti modi diversi può capitare che 4 palle vadano a finire nella Cella 3?
Il problema va spezzato in due sottoproblemi:

a) in quanti modi distinti posso scegliere le 4 palle che vanno a finire nella Cella 3?
b) per ognuna delle scelte operate al punto a) in quanti modi diversi le restanti 3 palle possono disporsi nelle restanti 4 celle?

La risposta alla prima domanda è: in tutti i modi in cui si possono scegliere 4 elementi estratti da 7 assegnati e tale numero è dato da

$$\binom{7}{4} = \frac{7!}{4! \cdot 3!} = 35$$

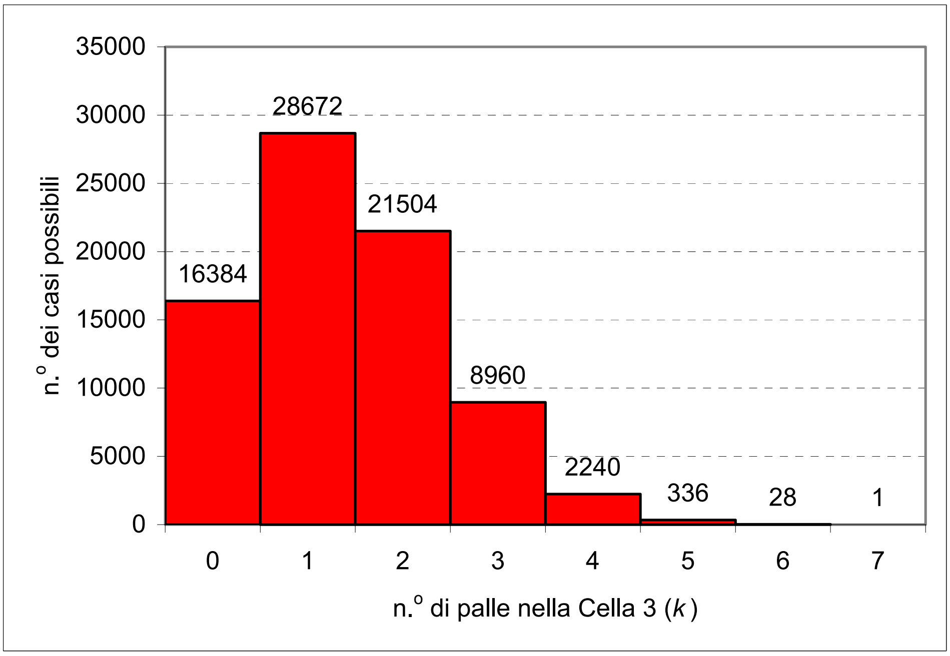
Per ognuna di tali scelte delle palle che vanno nella Cella 3, le altre 3 palle possono essere messe in modo del tutto arbitrario nelle $5 - 1 = 4$ celle rimanenti e le possibilità sono

$$(5\text{-}1)^3 = 4^3 = 64.$$

Dunque 4 palle possono finire nella cella 3 in $35 \times 64 = 2240$ modi differenti sulle $5^7 = 78125$ possibili disposizioni di tutte le palle in tutte le celle e la probabilità dunque di trovare 4 palle nella cella 3 è data dal quoziente tra il numero di casi in cui ciò capita e il numero di tutti i casi possibili

$$\frac{2240}{78125} = 0{,}028672...$$

Nel successivo grafico è stato rappresentato il numero, al variare di k, di tutti i casi in cui k palle vanno a finire nella Cella 3.



Il discorso fatto in questo caso particolare si può estendere al caso generale. Basta sostituire r a 7, n a 5 e k a 4: il risultato diviene

$$\binom{r}{k}(n-1)^{r-k}$$

per il numero dei casi in cui esattamente k palle scelte tra r vanno a finire in una cella prefissata;

$$\frac{\binom{r}{k}\cdot(n-1)^{r-k}}{n^r} = \frac{\binom{r}{k}\cdot n^{r-k}\left(1-\frac{1}{n}\right)^{r-k}}{n^r} = \binom{r}{k}\cdot\left(\frac{1}{n}\right)^k\cdot\left(1-\frac{1}{n}\right)^{r-k}$$

rappresenta quindi la probabilità che in una data cella finiscano esattamente k palle (quest'ultima è la cosiddetta *distribuzione binomiale*):

$$p_r^n(k) = \binom{r}{k}\cdot\left(\frac{1}{n}\right)^k\cdot\left(1-\frac{1}{n}\right)^{r-k}.$$

Attenzione, nell'esempio precedente, si erano associata ad ogni palla l'unica cella in cui tale palla andava a finire. Nell'esempio delle date di nascita ad ognuna delle r persone veniva associata la sua (unica!) data di nascita; quindi ora $n = 365$.

Se $k = 0$ (cioè nessuna persona ha data di nascita coincidente con quella assegnata), la probabilità è data da

$$p_r^{365}(0) = \left(1 - \frac{1}{365}\right)^r = \left(\frac{364}{365}\right)^r$$

Tale evento (nessuna persona ha data di nascita coincidente con quella assegnata), è esattamente il complementare di quello che ci interessa (almeno una persona nata in tale data) e dunque la probabilità di trovare almeno un persona su r che festeggia il compleanno in una data prefissata è

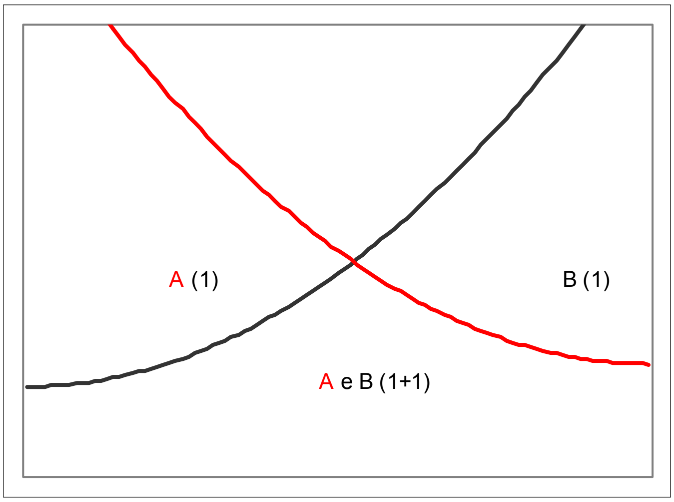
$$P_r(1) = 1 - \left(\frac{364}{365}\right)^r = \sum_{k=1}^{r} \binom{r}{k} \cdot \left(\frac{1}{365}\right)^k \cdot \left(\frac{364}{365}\right)^{r-k} = \frac{1}{365^r} \sum_{k=1}^{r} \binom{r}{k} \cdot 364^{r-k}$$

Ogni addendo ha una sua ben chiara spiegazione: si ha almeno una persona nata in una certa data sia quando se ne trova esattamente una (k =1), sia quando se ne trovano esattamente 2, 3, 4 fino al caso estremo in cui tutte le persone risultano nate proprio in quel giorno ($k = r$).

Il problema generale di trovare almeno s persone nate in una data prefissata ha ora una soluzione immediata: basta sommare le singole probabilità di trovare k persone nate in quella data, facendo variare k da un minimo di s a un massimo di r.

$$P_r(s) = \sum_{k=s}^{r} \binom{r}{k} \cdot \left(\frac{1}{365}\right)^k \cdot \left(\frac{364}{365}\right)^{r-k} = \frac{1}{365^r} \sum_{k=s}^{r} \binom{r}{k} \cdot 364^{r-k}$$

Consultare ora il documento ***Compleanni in data fissata.xls*** da cui si evince che occorrono più di 250 persone per avere più del 50% di probabilità di trovare una persona nata in una data prefissata, (esattamente r = 253), più di 600 (esattamente 613) per ottenere, sempre con il 50% di probabilità, due persone nate in quella data, quasi 1000 (esattamente 976) per trovarne 3 (vedere anche "***Complementi***, secondo paragrafo (2) Compilazione del documento Excel ***Compleanni in data fissata.xls***.")).



Affrontiamo ora il problema di fissare due date e di trovare (almeno) una persona nata in ciascuna di esse.
Conviene affrontare il problema opposto: in quanti casi capita che non si trova una coppia di persone con le precedenti caratteristiche?
Per comodità fissiamo le due date del 16/8 e del 23/9. Noi sappiamo che, se consideriamo tutti i possibili gruppi composti da r persone, ce ne sono 364^r che non contengono persone nate il 16/8 (insieme A, al di sotto della linea rossa) ed altrettanti casi in cui non ci sono persone nate il 23/9 (insieme B, al di sotto della linea nera). Nell'unione $A \cup B$ ci possono essere gruppi contenenti persone che festeggiano il compleanno il 16/8, ma non persone che lo festeggiano il 23/9 (zona denotata da A(1)) oppure gruppi contenenti persone che lo festeggiano il 23/9, ma non persone che lo festeggiano il 16/8 (zona denotata da B(1)) o infine gruppi in cui tutti gli appartenenti festeggiano il compleanno in date diverse dalle due prefissate (zona denotata dalle lettere A e B(1+1)). Quanti elementi contiene tale unione? Se sommiamo il numero di tutti gli elementi che stanno in A e di tutti gli elementi che stanno in B, gli elementi dell'intersezione sono

contati due volte (ecco spiegato l'1+1 accanto all'indicazione A e B nella figura precedente), quindi dobbiamo sottrarre una volta il numero degli elementi contenuti nell'intersezione A ∩ B.

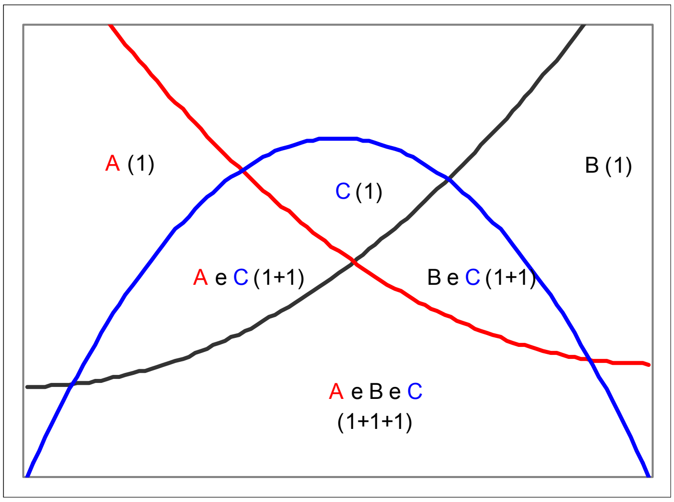
$$\#(A \cup B) = \#(A) + \#(B) - \#(A \cap B)$$

ove #(E) indica il numero degli elementi dell'insieme E.
Gli elementi dell'intersezione sono tutti i gruppi in cui sono vietate entrambe le date considerate e quindi il loro numero è 363^r. Ne segue che i casi in cui non si trova alcuna coppia del tipo richiesto sono

$$364^r + 364^r - 363^r$$

e la probabilità è data da

$$P_r^{'}(2) = 1 - 2 \cdot \left(\frac{364}{365}\right)^r + \left(\frac{363}{365}\right)^r$$



Abbiamo appena visto che

$$\#(A\cup B)=\#A + \#B - \#(A\cap B);$$

questo risultato si può estendere al caso di tre o più insiemi. Nella figura a lato i tre insiemi A, B, C sono rappresentati dalle porzioni di piano contenute sotto le linee rossa, nera e blu rispettivamente. Si deduce facilmente che

$$\#(A\cup B\cup C) = \#A + \#B + \#C - \#(A\cap B) - \#(A\cap C) - \#(B\cap C) + \#(A\cap B\cap C)$$

da cui si può dedurre la probabilità di trovare *almeno* 3 persone nate ciascuna in una tra 3 date fissate

$$P_r^{'}(3) = 1 - 3 \cdot \left(\frac{364}{365}\right)^r + 3 \cdot \left(\frac{363}{365}\right)^r - \left(\frac{362}{365}\right)^r$$

e, in generale,

$$P_r^{'}(s) = \sum_{j=0}^{s} (-1)^j \cdot \binom{s}{j} \cdot \left(\frac{365-j}{365}\right)^r$$

La dimostrazione di questo fatto non è difficile: infatti, in analogia con quanto appena visto, la formula generale per la cardinalità dell'unione è

$$\#\left(\bigcup_{i=1}^{s} A_i\right) =$$

$$\sum_{i=1}^{s} \#(A_i) - \sum_{\substack{i_1, i_2 = 1,..s \\ i_1 \neq i_2}} \#\left(A_{i_1} \cap A_{i_2}\right) + \sum_{\substack{i_1, i_2, i_3 = 1,..s \\ i_1 \neq i_2 \neq i_3}} \#\left(A_{i_1} \cap A_{i_2} \cap A_{i_3}\right) - \ldots + (-1)^{s-1} \#\left(\bigcap_{i=1}^{s} A_i\right)$$

ove, al secondo membro, alla somma delle cardinalità dapprima si sottrae la somma delle cardinalità delle intersezioni degli insiemi, presi a due a due; poi si aggiunge la somma delle cardinalità delle intersezioni degli insiemi, presi a tre a tre, e così via, alternando segni + e segni –, fino all'ultimo termine, in cui si considera la cardinalità dell'intersezione di tutti gli insiemi A_i, preceduta dal segno − oppure + a seconda che s sia pari, oppure dispari. A questo punto noi sappiamo che ogni insieme del tipo $A_{i_1} \cap A_{i_2} \cap ... \cap A_{i_j}$ rappresenta la collezione di tutti i gruppi composti da r persone che non sono nate in certe j date comprese tra le s date prefissate e il numero di tali gruppi è $(365\text{-}j)^r$. Il problema del numero di modi in cui si possono scegliere j date, da un insieme di r date prefissate è già stato risolto e la soluzione è data dal coefficiente binomiale $\binom{s}{j}$ e questo dimostra la formula enunciata.

Queste formule portano a risultati assolutamente sorprendenti: la seguente tabella contiene, nella seconda colonna, il minimo valore di r (=numero di persone considerate) per cui si supera il 50% di probabilità di trovare s persone nate in un giorno prefissato e, nella terza colonna, lo stesso valore minimo, nel caso di trovare (almeno) una persona nata in ciascuna delle s date distinte prefissate.

s	r (s persone 1 giorno)	r (s persone s giorni)
1	253	253
2	613	448
3	976	576
4	1340	671
5	1705	746
6	2070	809
7	2435	862
8	2800	909
9	3164	950
10	3529	987

Sostanzialmente dunque trovare 3 persone nate nella stessa data fissata è altrettanto difficile che trovarne 10 ciascuna delle quali nata in ciascuna tra 10 date fissate.

Il prossimo problema è totalmente diverso e consiste nel determinare quanto spesso capita, sempre in un gruppo di r persone, che ce ne siano 2 o più che festeggiano il compleanno contemporaneamente, ma non in date prefissate.

Il problema, nel caso più semplice, è di soluzione elementare, qualora si consideri l'evento complementare, cioè quello di non trovare due persone che festeggino contemporaneamente il compleanno:

la prima persona può festeggiare il compleanno in un giorno qualsiasi, la seconda deve festeggiarlo in un giorno diverso dal precedente, quindi ci sono 364 date possibili, il terzo deve festeggiarlo in un giorno diverso dai due precedenti, quindi 363 date possibili e così via; in definitiva ci sono

$$365 \times 364 \times 363 \times \ldots \times (365 - (r - 1))$$

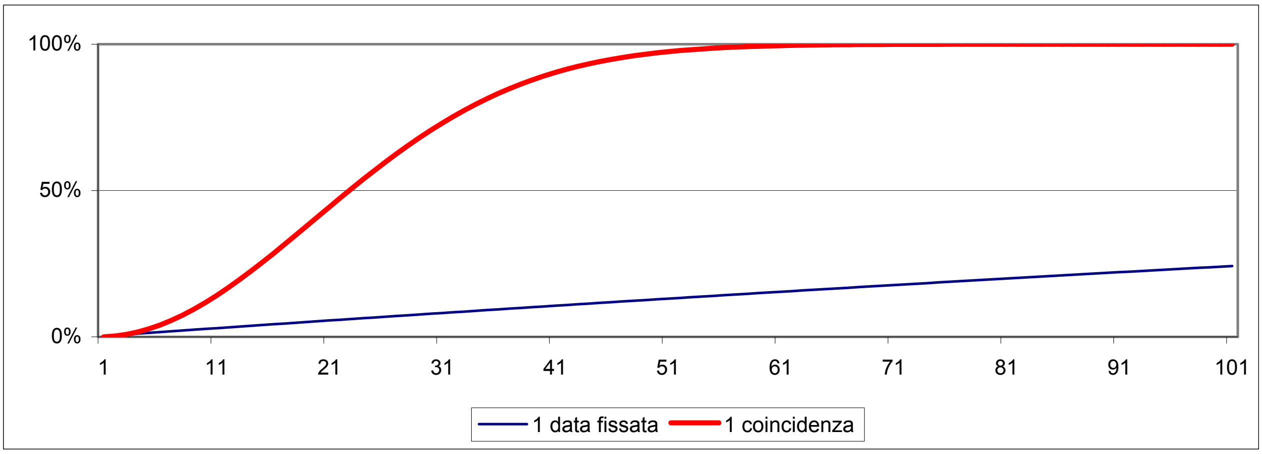
casi possibili di r date di compleanno che non presentano coincidenze; poiché il numero di tutte le possibili disposizioni di r date di compleanno è dato da 365^r, si deduce che la probabilità di non riscontrare coincidenze di compleanni in gruppi composti da r persone è dato da

$$\pi_0(r) = \frac{365 \times 364 \times ... \times (365 - (r-1))}{365^r}$$

e la probabilità dell'evento opposto è dato da

$$\pi_1(r) = 1 - \pi_0(r)$$

Tale numero tende assai rapidamente a 1 (=100%) come dimostra il seguente grafico in cui si confrontano gli andamenti della probabilità di trovare una persone nata in una data prefissata (linea blu), con quella di trovare due persone nate lo stesso giorno (linea rossa). N.B.: i valori di r sono posti sull'asse delle ascisse.



In particolare vale la seguente (sorprendente) tabella, in cui sono riportate, a seconda dei valori di r (prima colonna), le probabilità di trovare almeno una persona che festeggi il compleanno in una data prefissata (seconda colonna) e almeno una coincidenza di compleanni (terza colonna).

r	prob. compleanno fissato	prob. compleanni coincidenti
10	2,17%	11,69%
20	5,34%	41,14%
23	6,12%	50,73%
30	7,90%	70,63%
40	10,39%	89,12%
50	12,82%	97,04%
60	15,18%	99,41%
80	19,71%	99,99%

Se le coincidenze aumentano il problema si complica notevolmente: affronteremo solo i casi in cui si vogliano almeno 3 persone con compleanni coincidenti o almeno due coppie di persone con compleanni a due a due coincidenti, ma in date differenti per le due coppie.
Entrambi i problemi verranno affrontati con due (apparentemente) distinti approcci.

I problema: tre coincidenze

Metodo A

Consideriamo il problema complementare: se non vogliamo che in un insieme di r date ce ne siano 3 coincidenti, ci potranno essere j coppie di date coincidenti a due a due (quindi $2j$ persone che hanno a due a due compleanno in comune) e $r-2j$ date distinte sia tra loro che dalle j precedenti (corrispondono alle altre $r-2j$ persone) .

Chiamiamo una qualsiasi configurazione di questo tipo C_j^r e vediamo in quali casi, aggiungendo una nuova data, rimaniamo in una situazione in cui non si presentano tre coincidenze.

Due sono i casi possibili:

a) esce una data diversa dalle $r - j$ precedenti ($365 - (r - j)$ possibilità):
conseguenza: il numero (j) di coppie con compleanni coincidenti non varia: si ottiene un C_j^{r+1}
b) esce una delle $r - 2j$ date singole ($r - 2j$ possibilità):
conseguenza: il numero (j) di coppie con compleanni a due a due coincidenti aumenta di un'unità, quindi si ottiene un C_{j+1}^{r+1} (attenzione questo secondo caso è possibile solo se $j < \frac{r}{2}$).

Facendo lo stesso discorso, ma a ritroso, un C_j^r (attenzione! adesso c'è r e non $r + 1$) si genera in due possibili modi:

da un C_j^{r-1} in $365 - (r - 1 - j)$ modi

da un C_{j-1}^{r-1} in $r - 1 - 2(j - 1)$ modi

(nel caso $j = 0$, non ho coincidenze di date in C_0^r, quindi non ne posso neppure avere, quando tolgo una delle date in esso contenute: ne segue che solo il primo caso ha senso, in accordo con il fatto che C_{-1}^{r-1} non esiste; se $j = \frac{r}{2}$, mi trovo in una configurazione composta tutta da coppie di date e quindi essa può provenire solo da una configurazione con un'unica data isolata, a cui si è aggiunta proprio quella data: ne segue che solo il secondo caso ha senso, in accordo con il fatto che $C_{r/2}^{r-1}$ non esiste in quanto $\frac{r}{2} > \frac{r-1}{2}$)

Se allora indico con N_j^r il numero di configurazioni del tipo C_j^r, ottengo che

$$N_j^r = (365 - (r - 1 - j))N_j^{r-1} + (r - 1 - 2(j - 1))N_{j-1}^{r-1}$$

(in accordo con quanto detto sopra questa formula, a rigore, è inesatta se $j = 0$ o $j = \frac{r}{2}$; in realtà il problema non si pone, perché basta porre convenzionalmente $N_{-1}^r = 0$ e $N_j^r = 0$ se $j > \frac{r}{2}$)

La stessa relazione vale per le probabilità di ottenere una simile configurazione:

$$p_j^r = \frac{365 - (r - 1 - j)}{365} p_j^{r-1} + \frac{r - 1 - 2(j - 1)}{365} p_{j-1}^{r-1}$$

con la convenzione che $p_j^r = 0$ se $j > \frac{r}{2}$ oppure $j < 0$. Si osservi inoltre che, con la precedente convenzione, essendo noto $p_0^1 = 1$ (se ho un'unica persona sono sicuro che non ci possono essere 2 coincidenze di compleanno!!), la formula precedente mi fornisce tutti i valori di p_j^r (provare per credere, oppure si consulti il foglio Excel ***Coincidenze.xls*** o la guida per la sua compilazione in "***Complementi***, 3) Compilazione del documento Excel ***Coincidenze.xls***").

Il valore trovato corrisponde alla probabilità di trovare, in un gruppo di r persone, j coppie con compleanni a due a due coincidenti e tutti gli altri con compleanni "single"; basta ora sommare tali probabilità per tutti i possibili valori ammissibili di j, cioè da 0 (nessuna coincidenza di compleanni) al valore massimo possibile, che è dato da $\frac{r}{2}$ se r è pari e da $\frac{r-1}{2}$ se r è dispari: tale numero non è altro che il più grande intero che non supera $\frac{r}{2}$ e viene chiamato *parte intera* di $\frac{r}{2}$ e indicato con $[r/2]$. Quindi la probabilità, in un gruppo di r persone, di <u>non</u> <u>trovarne</u> <u>3</u> che abbiano il compleanno in comune è:

$$\pi_1'(r) = \sum_{j=0}^{[r/2]} p_j'^{\,r}$$

e la probabilità invece di <u>trovarne</u> <u>3</u> è, come è ovvio,

$$\pi_2(r) = 1 - \pi_1'(r)$$

Metodo B

Anche in questo caso consideriamo il problema complementare: devo escludere sia il caso in cui le date siano tutte distinte, sia tutti i possibili casi in cui si hanno solo coppie di compleanni coincidenti. Quanti sono questi ultimi casi? Ovvero quante sono le possibili configurazioni di r date di cui $2j$ siano a due a due coincidenti e le restanti $r-2j$ distinte tra loro e dalle precedenti? Comincio dunque con lo scegliere le $2j$ persone che avranno compleanni a due a due coincidenti; esistono

$$\binom{r}{2j} \text{ possibilità;}$$

mi chiedo ora in quanti modi si possano dividere queste persone in j gruppi di due persone ciascuno. Ne scelgo due tra i $2j$ e questo lo posso fare in $\binom{2j}{2}$ modi possibili; per ognuna di queste scelte ne devo scegliere altri 2 tra i $2(j-1)$ rimanenti, il che si può fare in $\binom{2(j-1)}{2}$ modi, e così via; in questo modo ho ottenuto tutti i possibili raggruppamenti a 2 a 2 che sono dunque

$$\binom{2j}{2} \times \binom{2(j-1)}{2} \times \ldots \times \binom{4}{2} \times \binom{2}{2} = \frac{(2j)!}{2!(2(j-1))!} \times \frac{(2(j-1))!}{2!(2(j-2))!} \times \ldots \times \frac{4!}{2! \cdot 2!} \times 1 = \frac{(2j)!}{2^j}$$

ma in questo modo abbiamo distinto anche gruppi di j coppie che differiscono solo per l'ordine con cui le coppie si presentano (ad esempio {(1,2),(3,4)} e {(3,4),(1,2)}; poiché gli ordinamenti di j oggetti (le coppie) sono $j!$, devo dividere per $j!$ il precedente risultato e in definitiva, il numero totale di possibili configurazioni di r persone, di cui $2j$ hanno compleanni a due a due coincidenti e le restanti $r-2j$ compleanni distinti sia tra loro sia dalle precedenti persone, è dato da

$$\underbrace{\frac{1}{j!}\frac{(2j)!}{2^j}}_{\substack{\text{modi di dividere } 2j \text{ persone} \\ \text{in gruppetti di 2}}} \times \underbrace{\binom{r}{2j}}_{\substack{\text{modi di scegliere } 2j \text{ persone} \\ \text{in un gruppo di } r}}$$

Terminiamo con l'assegnare, sia ad ogni coppia sia ad ogni singolo, $r-j$ date distinte: come al solito, la prima può essere scelta arbitrariamente tra le 365 a disposizione, la seconda tra le 364 restanti e così via; in totale le possibilità sono, come sappiamo, $365 \times 364 \times 363 \times \ldots \times (365-(r-j-1))$. In definitiva il numero di tutte le possibili configurazioni in cui, tra r persone, si hanno esattamente j coppie di compleanni coincidenti è dato da

$$\frac{1}{j!}\frac{(2j)!}{2^j} \cdot \binom{r}{2j} \times 365 \times 364 \times \ldots \times (365-(r-j-1))\,;$$

Se nella precedente formula consideriamo il caso (per ora escluso) j = 0, otteniamo il prodotto di tutti gli interi compresi tra 365 e 365−(r −1), che fornisce, come già visto, esattamente il numero dei casi in cui non si abbia alcuna coincidenza di compleanni (è sorprendente, ma questa affermazione non compare nei testi consultati). Sommando quindi la precedente quantità, per tutti i valori ammissibili di j (da 0 = "nessuna coincidenza" a $\left[\frac{r}{2}\right]$= "massimo numero di coppie che possono avere compleanni coincidenti"), si ottiene il numero di tutti i possibili casi in cui non si hanno terne di compleanni coincidenti:

$$\sum_{j=0}^{[r/2]} \frac{1}{j!}\frac{(2j)!}{2^j}\cdot\binom{r}{2j}\times 365\times 364\times\ldots\times(365-(r-j-1))$$

Per ottenere la loro probabilità complessiva basta dividere per tutte le possibili configurazioni che, come sappiamo, sono 365^r

$$\pi_1^{'}(r)=\frac{1}{365^r}\sum_{j=0}^{[r/2]} \frac{1}{j!}\frac{(2j)!}{2^j}\cdot\binom{r}{2j}\times 365\times 364\times\ldots\times(365-(r-j-1))$$

Finalmente, per ottenere la probabilità del nostro evento (3 persone con compleanni coincidenti), bisogna sottrarre ad 1 la precedente probabilità

$$\pi_2(r)=1-\pi_1^{'}(r)$$

Si può osservare che già con 88 persone la probabilità di trovarne almeno 3 persone con il compleanno coincidente supera il 50%.

La realizzazione numerica di quanto sopra descritto si trova nel documento Excel ***Coincidenze.xls***. (si veda anche "***Complementi***, 3) Compilazione del documento Excel ***Coincidenze.xls***").

Se si richiedono più di 3 coincidenze la formula diviene quasi ingestibile; se indichiamo con $\pi_{s-1}(r)$ la probabilità, in un gruppo di r persone, di trovarne almeno s con la stessa data di compleanno, vale la formula

$$\pi_{s-1}(r)=1-\sum_{j=1}^{s-1}\pi_j^{''}(r)$$

ove $\pi_j^{''}(r)$ è la probabilità che il numero massimo di compleanni coincidenti sia esattamente j (attenzione! la definizione fornita in [**MW**] non è esatta; osserviamo che $\pi_1^{''}(r)$ coincide con il nostro $\pi_0(r)$, perché denota la probabilità di non trovare coincidenze di compleanno, e $\pi_2^{''}(r)$ coincide con $\pi_1^{'}(r)-\pi_0(r)$). In [**#28**] viene anche fornita una formula ricorsiva per $\pi_j^{''}(r)$, che è troppo complessa per essere qui riportata: alcuni tra i risultati che si possono dedurre da essa sono sintetizzati nella seguente tabella in cui si mostra il primo valore di r (=numero di persone considerate) per cui si supera il 50% di probabilità di trovare, nel caso della seconda colonna, s persone nate in un giorno prefissato e nel caso della terza colonna, s persone che festeggiano il compleanno nello stesso giorno (non fissato a priori).

s	r (in 1 giorno fissato)	r (nello stesso giorno)
1	253	1
2	613	23
3	976	88
4	1340	187
5	1705	313
6	2070	460
7	2435	623
8	2800	798
9	3164	985
10	3529	1181

Può essere interessante ricordare che esistono numerose forme semplificate per determinare, con un errore ragionevole, il numero minimo di persone contenuto nella terza colonna della precedente colonna; in [**DM**] si fornisce la seguente approssimazione :

$$r = 47\sqrt{\left(s - \frac{3}{2}\right)^3}$$

i cui valori sono, per s = 3, 4, 5, 6, 7, 8, 9, 10, rispettivamente 86, 185, 307, 448, 606, 778, 965, 1164.

II problema: trovare (almeno) due coppie in ciascuna delle quali si ha coincidenza di compleanni (ovviamente le due coppie festeggiano il compleanno in giorni differenti). Di questo problema non ho trovato traccia in letteratura anche se è analogo, ma più semplice del precedente.

Metodo A

Come al solito consideriamo il problema complementare: si ammette che ci siano k persone che festeggiano il compleanno nello stesso giorno, ma tutti gli altri devono avere compleanni in date differenti tra loro e dalla precedente. Chiamiamo una qualsiasi configurazione di questo tipo C_{k-1}^{r} e vediamo in quali casi, aggiungendo una nuova data, rimaniamo in una situazione in cui non si presentano due coppie di coincidenze (la notazione adottata, cioè l'indice $k-1$, può sembrare strana, ma è giustificata dal fatto che, se k = 1, si hanno $k - 1 = 0$ coincidenze, e, in generale, in C_{k-1}^{r} si trovano $k - 1$ persone con il compleanno coincidente con un altro); osserviamo che le date coinvolte in una tale configurazione sono 1 (la data condivisa da k persone) + ($r - k$) (le date singole).

Se $k \geq 2$ (cioè si ha già una data che si presenta più di una volta), ci sono due casi possibili:
a) esce una data diversa dalle $r - k + 1$ già coinvolta (365 – ($r - k + 1$) possibilità):
conseguenza: il numero (k) di persone con lo stesso compleanno non varia: si ottiene un C_{k-1}^{r+1}
b) esce proprio la data singola (1 possibilità):
conseguenza: il numero (k) di persone con lo stesso compleanno aumenta di un'unità, quindi si ottiene un C_{k}^{r+1}.

Se k = 1 (cioè tutti i compleanni sono singoli), ci sono di nuovo due casi possibili:
a) esce una data diversa dalle r già coinvolta (365 – r possibilità):
conseguenza: continuano a non esserci coincidenze di date, quindi si ottiene un C_{0}^{r+1}

b) esce una delle r date già coinvolte (r possibilità):

conseguenza: nasce una coppia di persone con lo stesso compleanno, quindi si ottiene un C_1^{r+1}.

Procedendo a ritroso, un C_{k-1}^r (attenzione! adesso c'è r e non $r+1$) si genera in diversi possibili modi:

$k=1$	C_0^r nasce da	C_0^{r-1}	in $365-(r-1)$ modi
$k=2$	C_1^r nasce da	C_0^{r-1}	in $(r-1)$ modi
	nasce da	C_1^{r-1}	in $365-(r-1-1)$ modi
$2<k<r$	C_{k-1}^r nasce da	C_{k-2}^{r-1}	in 1 solo modo
	nasce da	C_{k-1}^{r-1}	in $365-(r-k)$ modi
$k=r$	C_{r-1}^r nasce da	C_{r-2}^{r-1}	in 1 solo modo

da cui, indicata con p_{k-1}^r la probabilità che, in un gruppo di r persone, si presenti una configurazione C_{k-1}^r, dal precedente schema si ottiene

$$k=1 \qquad p_0^r = p_0^{r-1}\left(1-\frac{r-1}{365}\right);$$

$$k=2 \qquad p_1^r = p_0^{r-1}\frac{r-1}{365} + p_1^{r-1}\left(1-\frac{r-2}{365}\right)$$

$$2<k<r \qquad p_{k-1}^r = p_{k-2}^{r-1}\frac{1}{365} + p_{k-1}^{r-1}\left(1-\frac{r-k}{365}\right)$$

$$k=r \qquad p_{r-1}^r = \frac{p_{r-2}^{r-1}}{365};$$

in realtà, se, come nel caso precedente, poniamo $p_{k-1}^s = 0$ se $k>s$ oppure se $k=0$, le precedenti formule si riducono solo a due:

$$k=2 \qquad p_1^r = p_0^{r-1}\frac{r-1}{365} + p_1^{r-1}\left(1-\frac{r-2}{365}\right)$$

$$k\neq 2 \qquad p_{k-1}^r = p_{k-2}^{r-1}\frac{1}{365} + p_{k-1}^{r-1}\left(1-\frac{r-k}{365}\right)$$

A questo punto la probabilità dell'evento contrario è data $\sum_{k=1}^{r} p_{k-1}^r$ e quello del nostro evento è

$$p^r = 1-\sum_{k=1}^{r} p_{k-1}^r$$

Metodo B

In questo caso fornisce la soluzione più semplice. Come sappiamo, il problema complementare: consiste nel considerare i casi in cui tutte le date siano distinte, oppure al più una possa presentarsi più volte. Ci chiediamo quante siano le possibili configurazioni di r date di cui una si presenti k volte e le restanti $r-k$ siano distinte tra loro e dalla precedente.

Le k persone con compleanni coincidenti possono essere scelte in $\binom{r}{k}$ modi; ora assegniamo le possibili date distinte, che sono in numero di r–k+1: come al solito, la prima può essere scelta arbitrariamente tra le 365 a disposizione, la seconda tra le 364 restanti e così via; in totale le possibilità sono, come sappiamo, 365×364×363×…×(365 – (r – k)). In definitiva il numero di tutte le possibili configurazioni del tipo preso in esame sono

$$\binom{r}{k} \times 365 \times 364 \times \ldots \times (365-(r-k))\,;$$

e la loro probabilità è data da

$$p_{k-1}^{r} = \binom{r}{k} \times \frac{365 \times 364 \times \ldots \times (365-(r-k))}{365^{r}} = \binom{r}{k} \cdot 1 \cdot \left(1-\frac{1}{365}\right) \cdot \ldots \cdot \left(1-\frac{r-k}{365}\right) \cdot 365^{k-1}$$

se k > 1, mentre, nel caso di date tutte diverse (k = 1), vale la solita formula

$$p_{0}^{r} = \left(1-\frac{1}{365}\right) \cdot \ldots \cdot \left(1-\frac{r-1}{365}\right)$$

Quindi

$$p^{r} = 1-\left(1-\frac{1}{365}\right) \cdot \ldots \cdot \left(1-\frac{r-1}{365}\right) - \sum_{k=2}^{r} \binom{r}{k} \cdot \left(1-\frac{1}{365}\right) \cdot \ldots \cdot \left(1-\frac{r-k}{365}\right) \cdot 365^{1-k}$$

La realizzazione numerica di quanto sopra descritto si trova nel documento Excel ***Due coppie.xls***. (si veda anche "***Complementi***, 3) Compilazione del documento Excel ***Due coppie.xls***").

Bibliografia

[**SG**] Stephen J. Gould, *Sorridi, grande lucciola*, Feltrinelli (1994), §1.
[**Fe**] Feller, W. *An Introduction to Probability Theory and its Applications*, Vol. I, Wiley (New York, 1968), p. 31-32.
[**DW**] Diaconis, P. and Mosteller, F. *Methods of studying coincidences*, J. Amer. Statist. Assoc. **84** (1989), 853-861.
[**#28**] Finch, S. *Puzzle #28 [June 1997] Coincident Birthdays* (1997), http://www.mathcad.com/library/LibraryContent/puzzles/soln28/sol28.html
[**#28a**] Finch, S. *Exact Solutions for the Generalized Birthday Problem* (1997), http://www.mathcad.com/library/LibraryContent/puzzles/soln28/exact28.html
[**MW**] Weisstein, W. *Birthday Problem*, MathWorld – A Wolfram Web Resource (1999), http://mathworld.wolfram.com/BirthdayProblem.html
[**MT**] Peterson, I. *MathTrek: Birthday Surprises* (1998) http://www.sciencenews.org/sn_arc98/11_21_98/mathland.htm
[**Nu**] Nunnikhoven, T.S. *A Birthday Problem Solution for Nonuniform Birth Frequencies*, American Statistician **46**: November: 270 (1992)