Complementi.

1) Come raggruppare oggetti.

Quando consideriamo il problema di raggruppare oggetti, in realtà affrontiamo problemi di tipo assai diverso.
A volte dobbiamo distribuire degli oggetti in certe posizioni, tenendo conto dell'ordine e permettendo che uno stesso oggetto si ripeta anche più di una volta, come nel caso della colonna del totocalcio

1 X 2 1 1 X X 1 1 2 1 1 X

In questo caso gli oggetti sono 3 (1 X 2), i posti sono 13 (sono rimasto al vecchio totocalcio) e l'ordine è fondamentale! Poiché in ogni casella si può inserire uno dei tre simboli 1 X 2 in modo (in teoria) del tutto arbitrario, il numero delle possibili colonne del totocalcio è dato da $3^{13} = 1594323$.
Se i tre simboli fossero equiprobabili (in questo caso non è vero, ma, altrimenti ci complicheremmo troppo la vita) la probabilità di uscita della colonna scritta sarebbe data da

$$\frac{1}{3^{13}} = 0{,}00000063\ldots .$$

Il caso considerato è quello delle *disposizioni* (conta l'ordine) con *ripetizione* (i simboli possono, in questo caso devono, ripetersi) di 3 oggetti in 13 posti.
Un caso particolarmente importante che si può affrontare con le disposizioni con ripetizione è quello di determinare il numero di tutti i possibili sottoinsiemi di un insieme finito. Supponiamo infatti che l'insieme A che stiamo considerando contenga r elementi $a_1, a_2, ..., a_r$: un sottoinsieme di A è determinato dalla scelta di alcuni (al limite, nessuno (sottoinsieme vuoto, $\varnothing$) o tutti (sottoinsieme totale, A)) elementi di A. Un qualsiasi sottoinsieme di A può, dunque essere identificato, scrivendo "sì" sopra gli elementi che ad esso appartengono e "no" sopra gli altri

$$\{a_1, a_2, a_4\} = \overset{si}{a_1}, \overset{si}{a_2}, \overset{no}{a_3}, \overset{si}{a_4}, \overset{no}{a_5} ..., \overset{no}{a_{r-1}}, \overset{no}{a_r} \;;$$

si tratta cioè di considerare tutte le possibili disposizioni di due simboli {sì, no} in r posizioni, ciascuna corrispondente ad un elemento di A (tutti "sì": (sotto)insieme totale; tutti "no": (sotto)insieme vuoto). Dunque il numero di tutti i sottoinsiemi di un insieme contenente r elementi è 2^r.
In generale il numero delle disposizioni di n elementi in r posizioni è dato da

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

Se invece passiamo alle estrazioni del lotto e consideriamo i numeri estratti (tenendo conto dell'ordine di estrazione) in una certa ruota

1° estratto: 23 2° estratto: 7 3° estratto: 82 4° estratto: 49 5° estratto: 19

siamo di fronte ad una situazione in cui gli oggetti sono 90 (tutti i numeri interi da 1 a 90), non possono ripetersi (una volta estratto, il numero non viene reinserito nell'urna) e i posti a disposizione sono 5.
Il fatto di dover evitare ripetizioni porta ad un cambiamento rispetto all'esempio precedente: se il primo estratto può ancora essere uno qualsiasi degli oggetti a disposizione (90), per il secondo le possibilità si restringono agli 89 numeri diversi dal primo estratto e, per il terzo, ne restano a disposizione solo 88. In definitiva il numero delle possibili estrazioni dei cinque numeri del lotto su una ruota, tenendo conto dell'ordine, è dato da

$$90\times 89\times 88\times 87\times 86=5273912160.$$

Il caso considerato è quello delle *disposizioni* (conta l'ordine) *semplici* (gli oggetti non possono ripetersi) di 90 oggetti in 5 posti.

In realtà ad uno scommettitore che ha giocato una cinquina non interessa l'ordine in cui i numeri sono stati estratti, ma solo il fatto che essi siano stati estratti. L'estrazione 23, 7, 82, 49, 19 è per lui del tutto equivalente ad una qualsiasi altra che contenga quei cinque numeri: ma quanti sono i possibili ordinamenti di 23, 7, 82, 49, 19? Si tratta di un problema di disposizioni semplici dei 5 precedenti numeri in 5 posti (quando il numero di oggetti è pari al numero di posti e non sono ammesse ripetizioni si parla di *permutazioni semplici*) e sappiamo che il numero di tali disposizioni è dato da

$$5\times 4\times 3\times 2\times 1 = 5! = 120.$$

Raggruppamenti come quello delle possibili cinquine su una ruota, in cui non conta l'ordine degli oggetti, prendono il nome di *combinazioni*. Quindi ogni combinazione di 5 numeri , dà luogo a 120 disposizioni. Ne segue che il numero di tutte le possibili combinazioni di 5 numeri, estratti tra i 90 numeri del lotto, è dato dal numero delle disposizioni (in cui conta l'ordine) diviso per 5!, cioè

$$\frac{90\times 89\times 88\times 87\times 86}{5!}=\frac{90\times\ldots\times 86\times 85\times 84\ldots\times 2\times 1}{5!\,85\times 84\times\ldots\times 2\times 1}=\frac{90!}{5!\,85!}=43949268$$

(a titolo di curiosità i minuti che compongono la vita di un uomo di 80 anni sono pochi di meno del precedente numero).

Come esempio, se considero i quattro oggetti a, b, c, d, e tre celle, ho i seguenti casi:

disposizioni semplici (24) a, b, c a, c, b b, a, c b, c, a c, a, b c, b, a a, b, d a, d, b
b, a, d b, d, a d, b, a d, a, b a, c, d a, d, c c, a, d c, d, a
d, a, c d, c, a b, c, d b, d, c c, b, d c, d, b d, b, c d, c, b

combinazioni semplici (4) a, b, c a, b, d a, c, d $b, c, d.$

Consideriamo ora un problema apparentemente diverso. Supponiamo di avere n oggetti di cui k di un tipo (chiamiamoli a) e $n-k$ di un altro (chiamiamoli b). Ci chiediamo quante siano le permutazioni di questi oggetti (conta dunque solo l'ordine). Ad esempio, nel caso dei 5 oggetti a, a, a, b, b, le permutazioni sono 10:

a, a, a, b, b a, a, b, a, b a, b, a, a, b b, a, a, a, b a, a, b, b, a
a, b, a, b, a b, a, a, b, a a, b, b, a, a b, a, b, a, a b, b, a, a, a

In generale il numero di tali permutazioni è dato da $\dfrac{n!}{k!(n-k)!}$.

La dimostrazione può essere svolta con l'artificio di differenziare gli elementi dei due tipi (a e b) con l'aggiunta di un indice, scrivendo, cioè,

$$a_1, a_2,\ldots, a_k, b_1, b_2,\ldots, b_{n-k} \quad \text{invece di} \quad \underbrace{a, a,\ldots, a}_{k\ \text{volte}};\ \underbrace{b, b,\ldots, b}_{n-k\ \text{volte}}$$

Le permutazioni di questi n oggetti, così diversificati, sono, come visto, $n!$ In realtà tutte le permutazioni che si ottengono da una di queste permutando a loro volta, in tutti i modi possibili, gli indici di a e di b, sono, in realtà, la stessa permutazione, una volta che eliminiamo tali indici posticci. Poiché le permutazioni dei k oggetti a_j sono $k!$ e quelle degli $n-k$ oggetti b_i sono $(n-k)!$ il numero delle permutazioni distinte degli n oggetti in questione è dato da $\frac{n!}{k!(n-k)!}$ (per essere più espliciti, nel caso dei 5 oggetti a, a, a, b, b, divenuti, con l'aggiunta degli indici a_1, a_2, a_3, b_1, b_2, le $3!\times 2! = 12$ permutazioni (distinte, finchè si mantengono gli indici ausiliari)

$$\begin{array}{llll} a_1, a_2, a_3, b_1, b_2, & a_1, a_3, a_2, b_1, b_2, & a_2, a_1, a_3, b_1, b_2, & a_2, a_3, a_1, b_1, b_2, \\ a_3, a_1, a_2, b_1, b_2, & a_3, a_2, a_1, b_1, b_2, & a_1, a_2, a_3, b_2, b_1, & a_1, a_3, a_2, b_2, b_1, \\ a_2, a_1, a_3, b_2, b_1, & a_2, a_3, a_1, b_2, b_1, & a_3, a_1, a_2, b_2, b_1, & a_3, a_2, a_1, b_2, b_1, \end{array}$$

si riducono alla sola

$$a, a, a, b, b$$

quando gli indici vengono tolti; si osservi anche che il numero di tutte le permutazioni di 5 oggetti, di cui 3 di un tipo e due dell'altro, è dato da $\frac{5!}{3!(5-3)!} = \frac{120}{6\times 2} = 10$, esattamente come visto nell'esempio della pagina precedente).

Situazioni del tipo di cui abbiamo appena parlato, sono assai frequenti; ad esempio, se vogliamo conoscere per esteso lo sviluppo del binomio $(a + b)^n$, dobbiamo pensare a tale potenza come al prodotto di n copie del binomio $(a + b)$; i singoli monomi che compongono tale sviluppo saranno del tipo $a^k b^{n-k}$ (k=0, 1, …, n), coerentemente al fatto di aver scelto, nell'effettuare il prodotto delle n copie del binomio $(a + b)$, la lettera a in k fattori e la lettera b nei restanti $n-k$. Una tale scelta può essere visualizzata, ponendo sotto ogni fattore quale dei due addendi è stato scelto, come mostrato nel seguente esempio in cui è visualizzata uno dei possibili modi di ottenere il monomio $a^5 b^2$ nello sviluppo di $(a + b)^7$

$$\underset{a}{(a+b)}\times\underset{b}{(a+b)}\times\underset{a}{(a+b)}\times\underset{a}{(a+b)}\times\underset{a}{(a+b)}\times\underset{b}{(a+b)}\times\underset{a}{(a+b)}$$

Di conseguenza tutte le possibili scelte delle lettere a e b, atte ad ottenere il monomio $a^k b^{n-k}$, corrispondono a tutte le possibili permutazioni di n oggetti di cui k eguali ad a e $n-k$ eguali a b e dunque tale monomio compare $\frac{n!}{k!(n-k)!}$ volte nello sviluppo di $(a + b)^n$:

$$(a+b)^n = \sum_{k=0}^{n} \frac{n!}{k!(n-k)!} a^k b^{n-k}$$

(si ricorda che, per convenzione, 0!=1). Tale formula giustifica il nome di *coefficiente binomiale* con cui viene spesso chiamato il numero $\frac{n!}{k!(n-k)!}$; si ricorda inoltre la notazione assai comune $\binom{n}{k}$ adottata per tali coefficienti.

Un ultimo problema, che si riallaccia a quanto detto a proposito del numero di sottoinsiemi di un insieme finito, consiste nel determinare quanti sono i sottoinsiemi di un insieme costituito da n elementi, che contengano esattamente k elementi. Un'occhiata a quanto scritto a proposito del numero di tutti i sottoinsiemi di un insieme dato, ci convince che tale numero è pari a tutte le possibili disposizioni di n oggetti di cui k eguali a "sì", corrispondenti agli elementi che appartengono al sottoinsieme, e $n-k$ eguali a "no" (… gli altri). Quindi, per ogni k compreso tra 0 e n, esistono $\binom{n}{k}$ sottoinsiemi di un insieme di cardinalità n, che contengono esattamente k elementi (per $k = 0$ e $k = n$ il numero $\binom{n}{k}$ vale 1, ed infatti esiste un solo sottoinsieme di A privo di elementi (l'insieme vuoto) ed uno solo contenente tutti gli elementi di A (A stesso)); se sommiamo il numero di tali sottoinsiemi, riotteniamo, com'è naturale, il numero di tutti i sottoinsiemi di A

$$\sum_{k=0}^{n} \frac{n!}{k!(n-k)!} = \sum_{k=0}^{n} \frac{n!}{k!(n-k)!} 1^k 1^{n-k} = (1+1)^n = 2^n.$$

I coefficienti binomiali si distribuiscono secondo un caratteristico triangolo, detto triangolo di Tartaglia, in cui ogni coefficiente si ottiene sommando i due elementi della riga precedente immediatamente a sinistra e immediatamente a destra del coefficiente che vogliamo calcolare

$n=0$	1
$n=1$	1 1
$n=2$	1 2 1
$n=3$	1 3 3 1
$n=4$	1 4 6 4 1
$n=5$	1 5 10 10 5 1
$n=6$	1 6 15 20 15 6 1
$n=7$	1 7 21 35 35 21 7 1
$n=8$	1 8 28 56 70 56 28 8 1
$n=9$	1 9 36 84 126 126 84 36 9 1
$n=10$	1 10 45 120 210 252 210 120 45 10 1

2) Compilazione del documento Excel ***Compleanni in data fissata.xls***.

Nel primo foglio (***data fissata***) sono contenuti i risultati relativi ai valori assunti dalla probabilità $P_r(s)$ di trovare, in un gruppo di r persone, almeno s che festeggino il compleanno in una data prefissata. La formula

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

dice che tale valore corrisponde alla somma di tutti i termini della distribuzione binomiale dal posto s in avanti.
Esiste una funzione di Excel che permette di operare tale calcolo:

=DISTRIB.BINOM(num_successi;prove;probabilità_s;cumulativo)

In essa **num_successi** è il numero di successi nelle prove effettuate, dove successo significa trovare una persona che festeggia il suo compleanno alla data prefissata, cioè, per noi, k;
prove è il numero di prove effettuate, ovvero di persone considerate, per noi r;
probabilità_s è la probabilità di successo per ciascuna prova; la probabilità di trovare una persona nata in un giorno prefissato vale $\frac{1}{365}$;
cumulativo è un valore logico. Se tale valore è VERO, DISTRIB.BINOM restituirà la funzione distribuzione cumulativa, ovvero la probabilità di ottenere un numero massimo di successi pari al valore di **num_successi**. In altre parole:

"=DISTRIB.BINOM(*k* ; *r* ; 1/365 ; VERO)" $= \sum_{j=0}^{k} \binom{r}{j} \cdot \left(\frac{1}{365}\right)^j \cdot \left(\frac{364}{365}\right)^{r-j}$

se il valore cumulativo è FALSO, verrà restituita la funzione massa di probabilità, ovvero la probabilità di ottenere un numero di successi esattamente pari al valore di ***num_successi***, ovvero ancora un singolo addendo della precedente somma

$$\textbf{“=DISTRIB.BINOM(}\boldsymbol{j}\textbf{ ; }\boldsymbol{r}\textbf{ ; 1/365 ; FALSO)”} = \binom{r}{j} \cdot \left(\frac{1}{365}\right)^{j} \cdot \left(\frac{364}{365}\right)^{r-j}$$

Nel nostro caso

$$P_r(s) = \sum_{k=s}^{r} \binom{r}{k} \cdot \left(\frac{1}{365}\right)^{k} \cdot \left(\frac{364}{365}\right)^{r-k} = 1 - \sum_{k=0}^{s-1} \binom{r}{k} \cdot \left(\frac{1}{365}\right)^{k} \cdot \left(\frac{364}{365}\right)^{r-k} =$$

$$\text{“}= 1 - \text{DISTRIB.BINOM(s - 1 , r , 1/365 , VERO)”}$$

Nella prima colonna del documento Excel sono stati posti i valori di s da 1 a 25 (nelle celle **A2:A26**), mentre la cella **A27** la riserviamo ad un qualsiasi numero (possiamo ad esempio scrivere 100); nella prima riga alcuni valori di r compresi tra 1 a 1000 nelle celle **B1:AN1** e **AO1** la riserviamo di nuovo ad un qualsiasi numero (possiamo ad esempio scrivere 10000); scriviamo il valore della funzione in **B2**, ponendo “=1-DISTRIB.BINOM($A2-1;B$1;1/365;VERO)” e poi dando l'***Invio*** (la presenza del segno di $ davanti alla lettera di colonna **A** fa sì che, dovunque si copi il contenuto della cella **B2**, il riferimento sia sempre alla colonna **A**; analogamente la presenza del segno di $ davanti al numero di riga **1** permette che, dovunque si copi il contenuto della cella **B2**, il riferimento sia sempre alla riga **1**). A questo punto si ***Copia*** la cella **B2** e la si ***Incolla*** da **B2** a **AO27**.

Compariranno i valori di probabilità cercati (le celle sono state formattate come ***Percentuale*** con 2 cifre decimali; compariranno anche tanti #NUM!: significa solo che per tali coppie di valori la distribuzione binomiale non è definita). Giocando con i valori contenuti nelle due celle **A27** e **AO1**, si possono trovare, al variare di s, i primi valori di r per cui la percentuale supera il 50% (verificare la tabella contenuta nel documento ***Compleanni coincidenti***); sempre giocando con i numeri nelle due celle indicate si può osservare che se s è molto grande, il primo valore di r per cui si supera la soglia del 50% è tale che il rapporto $\frac{r}{s}$ si avvicina sempre più a 365 (per $s = 50$, il minimo r è 18129, per $s = 100$, il minimo r è 36379)

N.B.: nonostante quanto detto a lezione Excel è in grado di svolgere tali calcoli anche per valori di r assai superiori a 1000.

Nel secondo foglio (***s date***) si calcolano i valori della probabilità di trovare, una volta fissate s date, per ognuna di esse almeno una persona che festeggia il suo compleanno esattamente in quella data. Si tratta cioè di scrivere la formula

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

Per fare ciò si scrivono nella colonna **A** i valori di s da 2 in **A2** (il caso 1 è già stato trattato) a 10 in **A10** e nella riga **1** i valori di r: ne sono stati scelti alcuni, da 10 a 400, nelle celle da **B1** a **J1**; nelle celle successive sono posti i valori che, per primi, permettono di ottenere almeno il 50% di probabilità. Per i valori di s considerati la probabilità assume i valori

$$P'_r(2) = 1 - 2 \cdot \left(\frac{364}{365}\right)^r + \left(\frac{363}{365}\right)^r ; \quad P'_r(3) = 1 - 3 \cdot \left(\frac{364}{365}\right)^r + 3 \cdot \left(\frac{363}{365}\right)^r - \left(\frac{362}{365}\right)^r$$

$$P'_r(4) = 1 - 4 \cdot \left(\frac{364}{365}\right)^r + 6 \cdot \left(\frac{363}{365}\right)^r - 4 \cdot \left(\frac{362}{365}\right)^r + \left(\frac{361}{365}\right)^r \quad \text{e così via}$$

Si tratta allora di scrivere "=1-2*(364/365)^B1+(363/365)^B1" nella cella **B2**, "=1-3*(364/365)^B1+3*(363/365)^B1-(362/365)^B1" in **B3**, e così via, fino a **B10**. Si copia poi il contenuto delle celle **B2:B10**, e lo si incolla su **C2:S10**. Si osservi, al variare di s, per quali valori di r si ottiene per la prima volta il valore del 50%.

Nel terzo foglio (***2 coincidenze***) si confrontano, al variare di r, la probabilità di trovare (almeno) una persona nata in una data prefissata (colonna **F**) con quella di trovare (almeno) una coppia di persone nate nello stesso giorno, non determinato a priori (colonna **G**).

3) Compilazione del documento Excel ***Coincidenze.xls***.

Nel primo foglio (***A***) si costruisce la soluzione al problema secondo il Metodo A.
La formula chiave per il calcolo della probabilità di trovare in un gruppo di persone almeno tre che festeggiano contemporaneamente il compleanno è dato dalla formula ricorsiva

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

che assegna la 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".
Nella colonna **A**, a partire da **A3**, si sono posti i valori di j da -1 a 250; nella riga **2**, a partire da **B2**, i valori di r pure da 1 a 250 (fino alla cella **IQ2**). Nella riga **3**, si sono messi i valori di p_{-1}^r che sono tutti nulli. Lo stesso, per motivi del tutto analoghi (si vedano le convenzioni adottate per rendere sempre valida la precedente formula), è stato fatto nella colonna **B** a partire da **B5.**
Ora sistemiamo tutte le altre probabilità, ricordandoci che il valore di p_j^r si trova all'incrocio tra la colonna corrispondente al valore di r (che è scritto nella riga **2**) e la riga corrispondente al valore di j (che è scritto nella colonna **A**), basta porre

=B4*(365-(B$2-$A4))/365+B3*(B$2-2*$A3)/365

nella cella **C4** e dare l'***Invio*** (uscirà l'ormai ben noto 0,99726 = probabilità di non trovare due persone con lo stesso compleanno in un gruppo di…2). La formula Excel è facilmente giustificabile:
in **C4** si deve porre il valore di p_0^2, che dipende da p_0^1 (che occupa la cella **B4**) e da p_{-1}^1 (che occupa la cella **B3**); si osservi anche che in **B2** vi è il valore 1 che coincide con $r-1$, in **A4** il valore 0 che coincide con j, in **A3** il valore -1 coincidente con $j-1$, e che il segno di $ ha l'effetto di bloccare il riferimento che lo segue (sia esso di riga oppure di colonna).

A questo punto basta copiare il contenuto della cella **C4** e incollarlo in **C3:IQ254**. Per ottenere la probabilità dell'evento contrario (non ci sono 3 coincidenze), si devono sommare tutti i numeri contenuti nelle colonne (nella riga **256**), mentre la probabilità dell'evento "almeno 3 persone festeggiano il compleanno nello stesso giorno" si ottiene sottraendo il precedente valore ad 1 (riga **1**).

Nel secondo foglio (***confronti***) vengono visualizzati sullo stesso grafico gli andamenti delle probabilità dei tre eventi "2 coincidenze di compleanni", "3 coincidenze di compleanni", "compleanno in data fissata". Come si vede l'ultimo evento è più probabile del secondo solo fino a $r = 50$; già a $r = 88$ le "tre coincidenze" superano il 50%, mentre il compleanno in data fissata è di poco superiore al 21%.

Nel terzo foglio controlleremo (una parte de) i risultati ottenuti nel primo foglio, utilizzando la formula del Metodo B. Per rendere più agevoli i calcoli, conviene modificare leggermente la formula

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

sostituendo al coefficiente binomiale la sua espressione in termini di fattoriali, spalmando $r-j$ tra gli r fattori 365 sotto i termini del prodotto $365\times 364\times\ldots\times(365-(r-j-1))$ e sostituendo alla variabile j (tra 0 e $[r/2]$), la variabile $m = r-j-1$ (che varia tra $r-[r/2]-1$, corrispondente a $j=[r/2]$, e $r-1$, corrispondente a $j=0$):

$$\begin{aligned}\pi_1'(r)&=\sum_{j=0}^{[r/2]}\frac{1}{365^j}\cdot\frac{1}{j!}\cdot\frac{(2j)!}{2^j}\cdot\frac{r!}{(2j)!(r-2j)!}\cdot\frac{365}{365}\cdot\frac{364}{365}\cdot\ldots\cdot\frac{365-(r-j-1)}{365}\\&=\sum_{m=r-[r/2]-1}^{r-1}\frac{r!}{730^{r-m-1}\cdot(r-m-1)!(2m-r+2)!}\cdot\left(1-\frac{1}{365}\right)\cdot\left(1-\frac{2}{365}\right)\cdot\ldots\cdot\left(1-\frac{m}{365}\right)\end{aligned}$$

A questo punto scriviamo nella colonna **A** i valori di m (partendo dal valore 0 in **A3**, fino a 150 in **A153**), quelli di k nella colonna **B** (stesso range di m), i valori di $1-\frac{k}{365}$ nella colonna **C** (scrivendo =1-B3/365 nella cella **C3**, invio, copia e incolla fino a **C153**), i prodotti $\prod_{k=0}^{m}\left(1-\frac{k}{365}\right)$ nella colonna **D**, scrivendo =C3 in **D3**, =C4*D3 in **D4** e così via (in questo modo si aggiunge ogni volta al prodotto di un certo numero di termini consecutivi del tipo $\left(1-\frac{k}{365}\right)$, il termine immediatamente successivo). A questo punto, nella colonna **E**, stimeremo la probabilità dell'evento complementare, scrivendo il valore di r in **E2** e moltiplicando il corrispondente termine della colonna **D** per il coefficiente che a lui compete $\frac{r!}{730^{r-m-1}\cdot(r-m-1)!(2m-r+2)!}$. Tale coefficiente, se scritto in **E3**, ha la seguente espressione Excel:

FATTORIALE(E$2)/(730^(E$2-$A3-1)*FATTORIALE(E$2-$A3-1)*FATTORIALE(2*$A3+2-E$2))

che, moltiplicato, sempre considerando la cella **E3**, per il termine contenuto nella cella immediatamente a sinistra (**D3**), fornisce il termine generale della somma considerata[1]. Copiando il contenuto della cella **E3** e incollando fino a **E153** si ottengono tutti gli addendi che forniscono

[1] Uso del simbolo $. Come già accennato, tale simbolo serve a fissare i riferimenti: per essere più precisi, se, nella formula contenuta in una cella, si fa riferimento ad un'altra cella, tale riferimento deve essere considerato "relativo", nel senso che riguarda solo la posizione *relativa* della seconda cella rispetto alla prima. Ad esempio, se in una formula contenuta in **D7** si fa riferimento al contenuto della cella **C5**, in realtà ci si riferisce alla cella situata una colonna a sinistra e due righe sopra la cella di partenza; se ora si incolla il contenuto di **D7** in **F11**, la cella a cui si farà riferimento sarà **E9** (che si trova appunto una colonna a sinistra e due righe sopra **F11**). Se nella formula in **D7** si fosse scritto $C5 invece di C5, il riferimento continuava ad essere relativo per quanto riguarda l'indice di riga (due posizioni sopra), ma fisso per quel che riguarda la colonna (sempre colonna **C**): così, se si incolla il contenuto di **D7** in **F11**, la cella a cui si farà riferimento sarà **C9** (che si trova appunto due righe sopra **F11**, ma nella colonna **C**). È chiaro ciò che succede negli altri casi; riassumendo

formula in **D7** in cui vi è il riferimento C5	⇒	incollato in **F11** diviene E9;
formula in **D7** in cui vi è il riferimento $C5	⇒	incollato in **F11** diviene $C9;
formula in **D7** in cui vi è il riferimento C$5	⇒	incollato in **F11** diviene E$5;
formula in **D7** in cui vi è il riferimento C5	⇒	incollato in **F11** resta C5.

$\pi_1^{''}(r)$; bisogna però tenere presente che vanno sommati solo i termini compresi tra $r-[r/2]-1$ e $r-1$. Excel è in grado di realizzare automaticamente tale limitazione, utilizzando la funzione logica SE(*test*; *se_vero*; *se_falso*) ove *test* è un valore o un'espressione qualsiasi che può dare come risultato VERO o FALSO, *se_vero* è il valore che viene restituito se test è VERO, *se_falso* è il valore che viene restituito se test è FALSO.

Nel nostro caso dovremo fare in modo che gli addendi che intervengono nell'espressione di $\pi_1^{''}(r)$ siano nulli al di fuori dell'intervallo richiesto: per fare ciò basta usare due volte la funzione SE; si dovrà porre in **E3**:

=$D3*SE($A3>E$2-1;0;SE($A3<E$2/2-1;0;FATTORIALE(E$2)/(FATTORIALE(E$2-$A3-1)*
FATTORIALE(2*$A3+2-E$2)*730^(E$2-$A3-1))))

copiare ed incollare fino a **E153**. Si osservi infatti che nella colonna **A** si trovano i valori di *m* e che in **E2** si trova il valore di *r*; quindi la condizione $A#>E$2−1 (# sta per un qualsiasi numero, indice di riga) significa $m > r-1$ e, quando ciò è verificato, l'addendo deve essere nullo: infatti nella posizione *se_vero* c'è il valore 0; altrimenti (posizione *se_falso*) vi è una nuova funzione SE con *test* $A#<E$2/2-1, che è equivalente a $m < r - \left\lfloor \frac{r}{2} \right\rfloor - 1$ (è proprio la stessa cosa se r è pari e, se r è dispari, … pure, perché *m* è un intero); di nuovo, quando ciò è verificato, l'addendo deve essere nullo (corrisponde allo 0 come secondo argomento); altrimenti, ed ora è finalmente verificata la condizione $r - \left\lfloor \frac{r}{2} \right\rfloor - 1 \le m \le r-1$, l'addendo ha l'espressione analitica non nulla che abbiamo determinato in precedenza. Si osservi che l'ultimo addendo non nullo corrisponde al caso in cui non si presenti alcuna coincidenza di compleanni.
A questo punto, per trovare la probabilità dell'evento complementare, basta sommare i valori, ponendo =SOMMA(E3:E153) in **E154** e, quindi, per ottenere la probabilità dell'evento "3 coincidenze" si scrive =1-E154 in **E1**. Cambiando i valori di r in **E2**, si può verificare la coincidenza con i valori ottenuti nel primo foglio.

4) Compilazione del documento Excel ***Due coppie.xls***.

Nel primo foglio si costruisce la soluzione secondo il Metodo A.
Le due formule chiave sono

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

Nella colonna **A**, a partire da **A4**, si pongono i valori di $k-1$ da -1 a 250; nella riga **3**, a partire da **B3**, i valori di r pure da 1 a 250 (fino alla cella **IQ1**). Nella riga **4**, si sono messi i valori di p_{-1}^r che sono tutti nulli. Ora sistemiamo tutte le altre probabilità, ricordandoci che il valore di p_{k-1}^r si trova all'incrocio tra la colonna corrispondente al valore di r (che è scritto nella riga **3**) e la riga corrispondente al valore di $k-1$ (che è scritto nella colonna **A**):

=B4/365+B5*(365-(C$3-$A5-1))/365

da scrivere in **C5** (uscirà, come al solito, 0,99726). La formula Excel è giustificabile esattamente come quella in ***Coincidenze.xls***; bisogna solo osservare che nella colonna **A** sono presenti i valori di $k-1$ e non di k e questo giustifica il termine C$3-$A5-1, eguale a $r-(k-1)-1$. Basta copiare il contenuto di **C5** e incollarlo prima in **C5:IQ5** e poi in **C7:IQ255**. Nella riga **6** bisogna usare l'altra formula:

=(C$3-1)*B5/365+B6*(365-(C$3-$A6-1))/365

in **C6**, invio, copia e incolla su **C6:IQ6**. Per ottenere la probabilità dell'evento contrario si sommano tutti i valori contenuti nelle colonne, ponendo =SOMMA(C5:C255) nella riga **C257**, invio, copia e incolla fino a **IQ257**, mentre la probabilità dell'evento "esistono almeno 2 coppie che festeggiano il compleanno nello stesso giorno (diverso per ogni coppia)" si ottiene sottraendo il precedente valore ad 1 (riga **1**).
Nel secondo foglio vengono visualizzati sullo stesso grafico gli andamenti delle probabilità dei tre eventi "2 coincidenze di compleanni", "3 coincidenze di compleanni", "compleanno in data fissata". Come si vede l'ultimo evento è più probabile del secondo solo fino a $r = 50$; già a $r = 88$ le "tre coincidenze" superano il 50%, mentre il compleanno in data fissata è di poco superiore al 21%.
Il secondo foglio è dedicato al controllo dei risultati ottenuti nel primo foglio, utilizzando la formula del Metodo B:

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

operando il cambio di variabile $r-k=m$

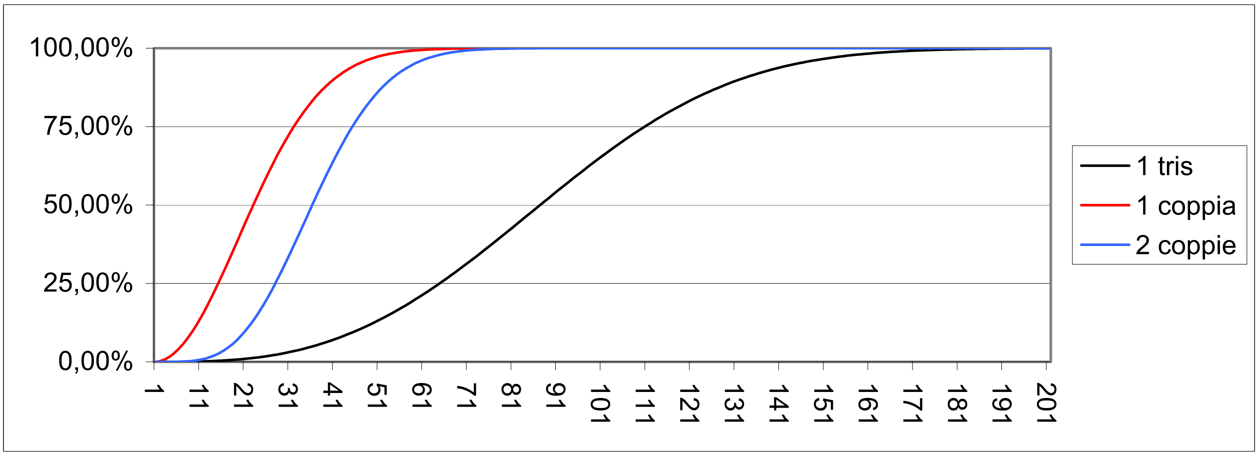
$$p^r = 1-\prod_{k=0}^{r-1}\left(1-\frac{k}{365}\right)-\sum_{m=0}^{r-2}\binom{r}{r-m}\cdot\frac{1}{365^{r-m-1}}\cdot\prod_{k=0}^{m}\left(1-\frac{k}{365}\right)$$

Scriviamo nella colonna **A** i valori di m (partendo dal valore 0 in **A3**, fino a 150 in **A153**), quelli di k nella colonna **B** (stesso range di r), quelli di $1-\frac{k}{365}$ nella colonna **C** (scrivendo =1-B3/365 in **C3**, invio, copia e incolla fino a **C153**), i prodotti $\prod_{k=0}^{m}\left(1-\frac{k}{365}\right)$ nella colonna **D**, scrivendo =C3 in **D3**, =C4*D3 in **D4** e così via. A questo punto, nella colonna **F**, stimeremo la probabilità dell'evento complementare, scrivendo il valore di r in **F2** e moltiplicando il corrispondente termine della colonna **D** per il coefficiente che a lui compete $\binom{r}{r-m}\cdot 365^{r-m-1}$. Quindi bisognerebbe scrivere in **F3** la seguente espressione Excel:

=D3*FATTORIALE(F$2)/(FATTORIALE(A3)*FATTORIALE(F$2-A3)*365^(F$2-1-A3))

In realtà bisogna considerare solo i valori di m da 0 a $r-2$, mentre, se $m = r-1$, bisogna considerare solo il prodotto $\prod_{k=0}^{r-1}\left(1-\frac{k}{365}\right)$; basta allora scrivere in **F3**
=SE($A3<F$2-1;$D3*FATTORIALE(F$2)/(FATTORIALE($A3)*FATTORIALE(F$2-$A3)*365^(F$2-1-$A3));SE($A3<=F$2-1;D3;0))

Come negli altri casi la probabilità dell'evento contrario si ottiene sommando tutti gli addendi in **F155** (=SOMMA(F3:F153)), e quella dell'evento considerato in **F1**, ponendo =1−F155.
Il risultato ottenuto, sintetizzato nel grafico all'inizio della pagina seguente, mostra che è estremamente più facile superare il 50% di probabilità nel caso della doppia coppia (basta $r = 36$), che quello del tris (basta $r = 36$).



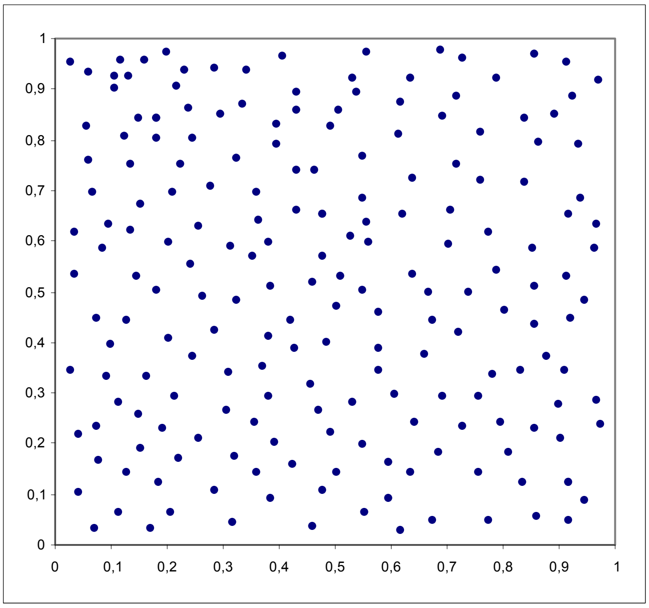
5) Il problema di inserire punti casuali in un quadrato.

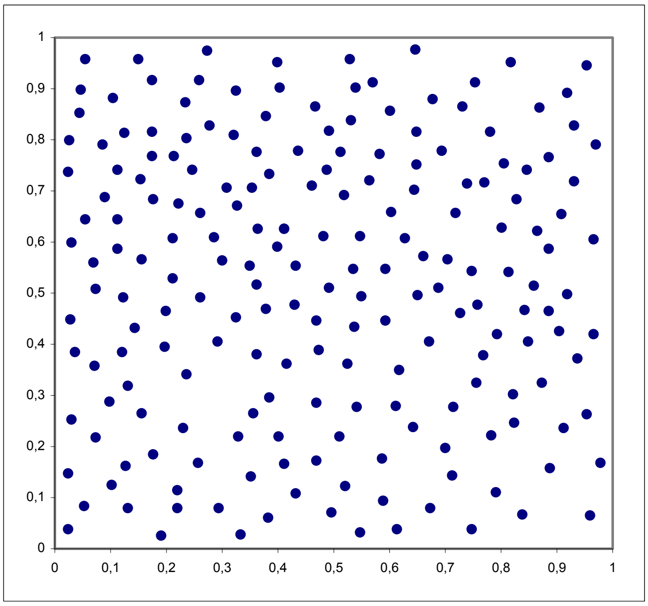
Abbiamo visto all'inizio del documento ***Compleanni coincidenti*** che, nell'inserimento casuale di r oggetti in n celle, la probabilità di trovarne k nella stessa cella segue la distribuzione binomiale, relativa alla probabilità $\frac{1}{n}$ (poiché supponiamo che sia equiprobabile per un oggetto finire in una cella piuttosto che in un'altra)

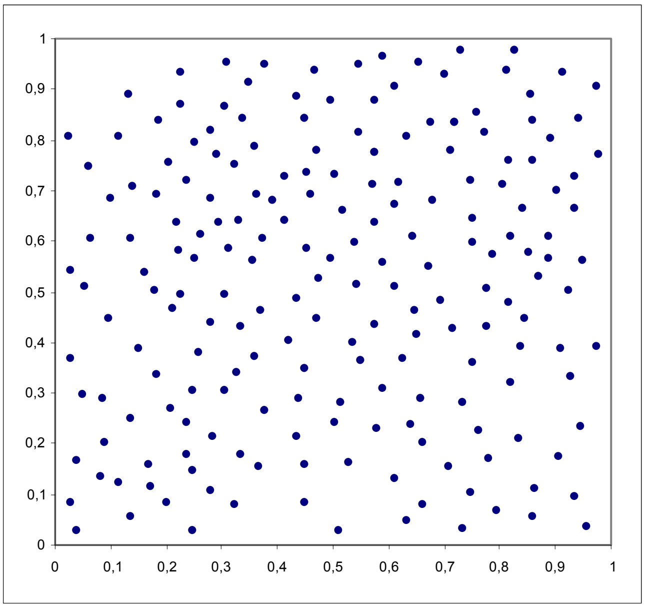
$$p_r^n(k) = \binom{r}{k} \cdot \left(\frac{1}{n}\right)^k \cdot \left(1 - \frac{1}{n}\right)^{r-k}$$

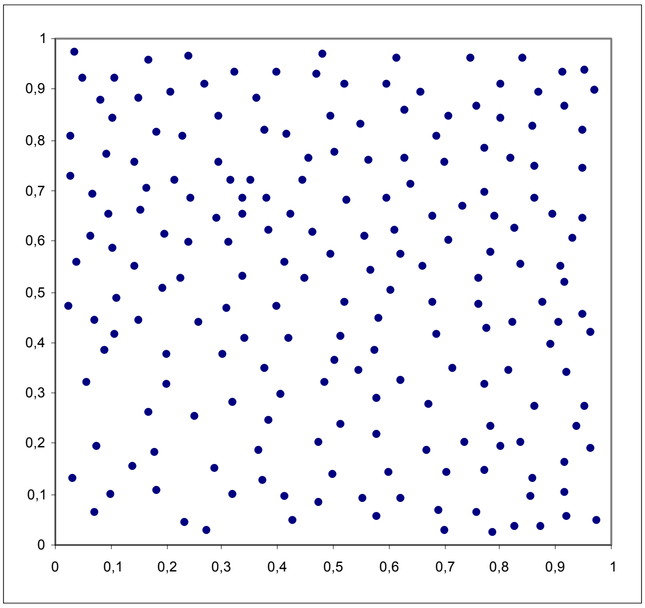
Questo risultato può permetterci di capire se r punti sono stati distribuiti casualmente oppure no in un quadrato. Supponendo che il quadrato abbia lato unitario, possiamo infatti dividere il quadrato in tanti quadratini di lato $\frac{1}{m}$ e contare quanti punti sono caduti in ogni singolo quadratino e, dopo aver visto in quanti quadratini sono caduti 0, 1, 2, 3, ecc. punti, confrontare la distribuzione ottenuta con quella teorica data dalla distribuzione binomiale con $n = m^2$ (che è, per l'appunto, il numero dei quadratini). La cosa è meno banale di quanto possa sembrare a prima vista, perché il risultato dipende in modo assai forte dal rapporto tra r e m^2; infatti, nel caso banale in cui $m = 1$, comunque si distribuiscano i punti, tutti vanno a finire nello stesso (unico!) quadratino; se invece, tenendo fisso r, si fa crescere m in modo che m^2 risulti molto più grande di r, quasi tutti i quadratini risulteranno vuoti, circa r conterranno un solo punto e solo pochissimi ne conterranno più di uno: quindi, anche in questo secondo caso, le varie distribuzioni di punti risulteranno indistinguibili.
Il test che è stato fatto si ispira ad una curiosità (vedi [**SG**]): l'essere umano è incapace di mettere veramente a caso dei punti in un quadrato.
Con AutoCad ho inserito 200 punti casualmente (o, almeno, così pensavo di avere fatto) in un quadrato, che ho poi suddiviso in 16, 25, 36 e 100 quadratini eguali; ho contato i punti che cadevano in ogni quadratino e ho confrontato con la distribuzione binomiale (teorica) con $r = 200$ e $m^2 = 16$, 25, 36 e 100. Per valutare il grado di affinità tra distribuzioni osservate e teoriche ho usato il test del χ^2 che fornisce la probabilità che la distribuzione osservata differisca da quella teorica, unicamente per effetto del caso.
In un secondo tempo l'operazione di disporre 200 punti a caso è stata affidata ad Excel, utilizzando la funzione =CASUALE() (questa in realtà genera numeri pseudocasuali, perché nessun algoritmo può veramente produrre numeri casuali).
Le disposizioni di punti ottenute in queste due serie di esperimenti sono rappresentate nelle seguenti pagina (si osservi come nella seconda serie appare l'effetto "cielo stellato" tipico delle distribuzioni veramente casuali, mentre le prime sono una perturbazione di una distribuzione uniforme).

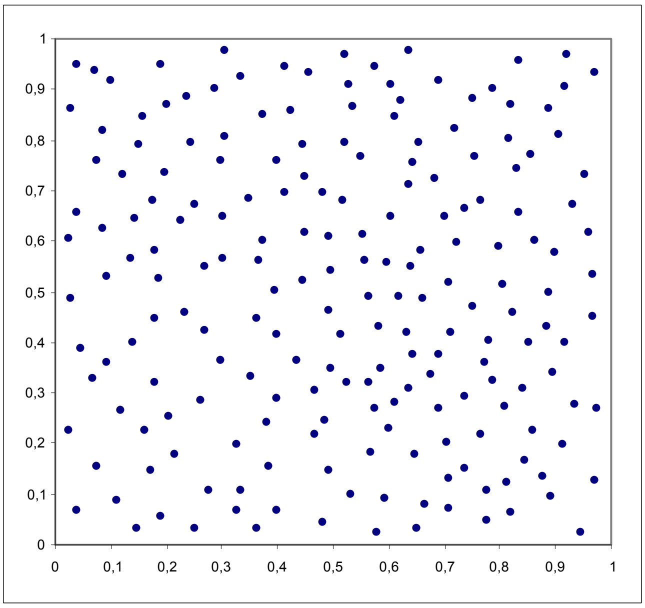
Serie di 200 punti disposti “casualmente” a mano



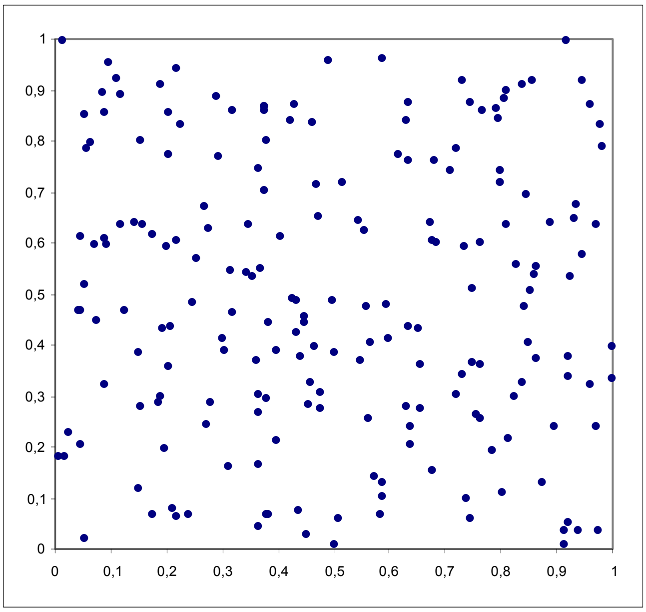


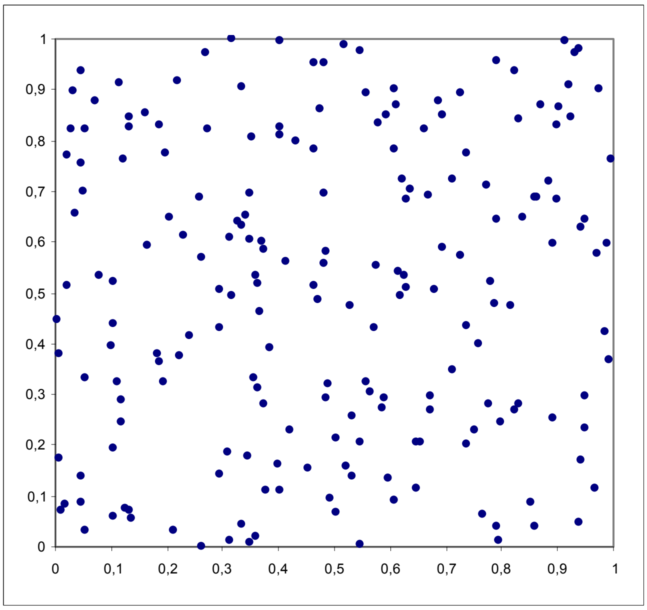


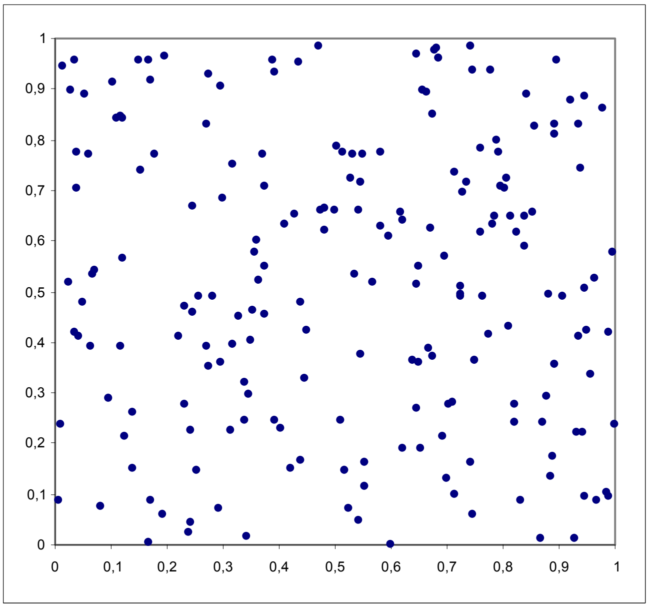


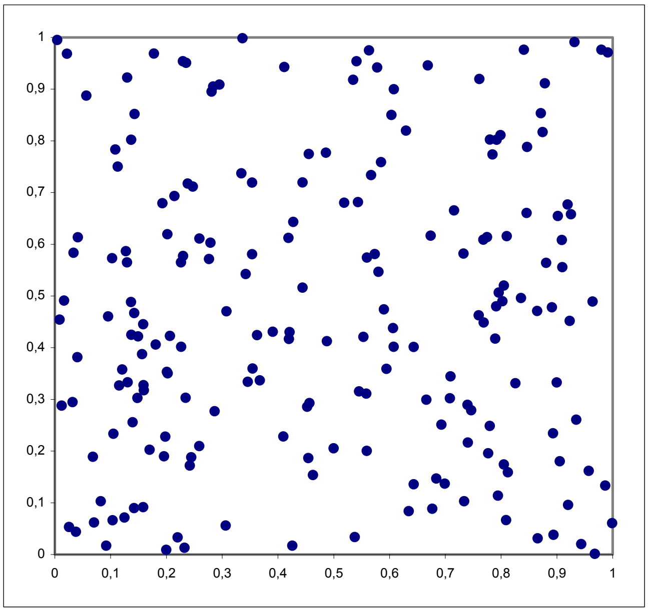


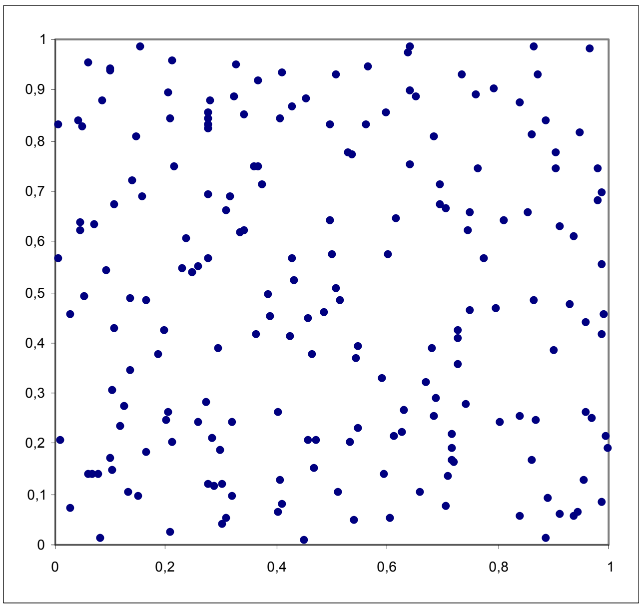
Serie di 200 punti disposti “casualmente” da Excel



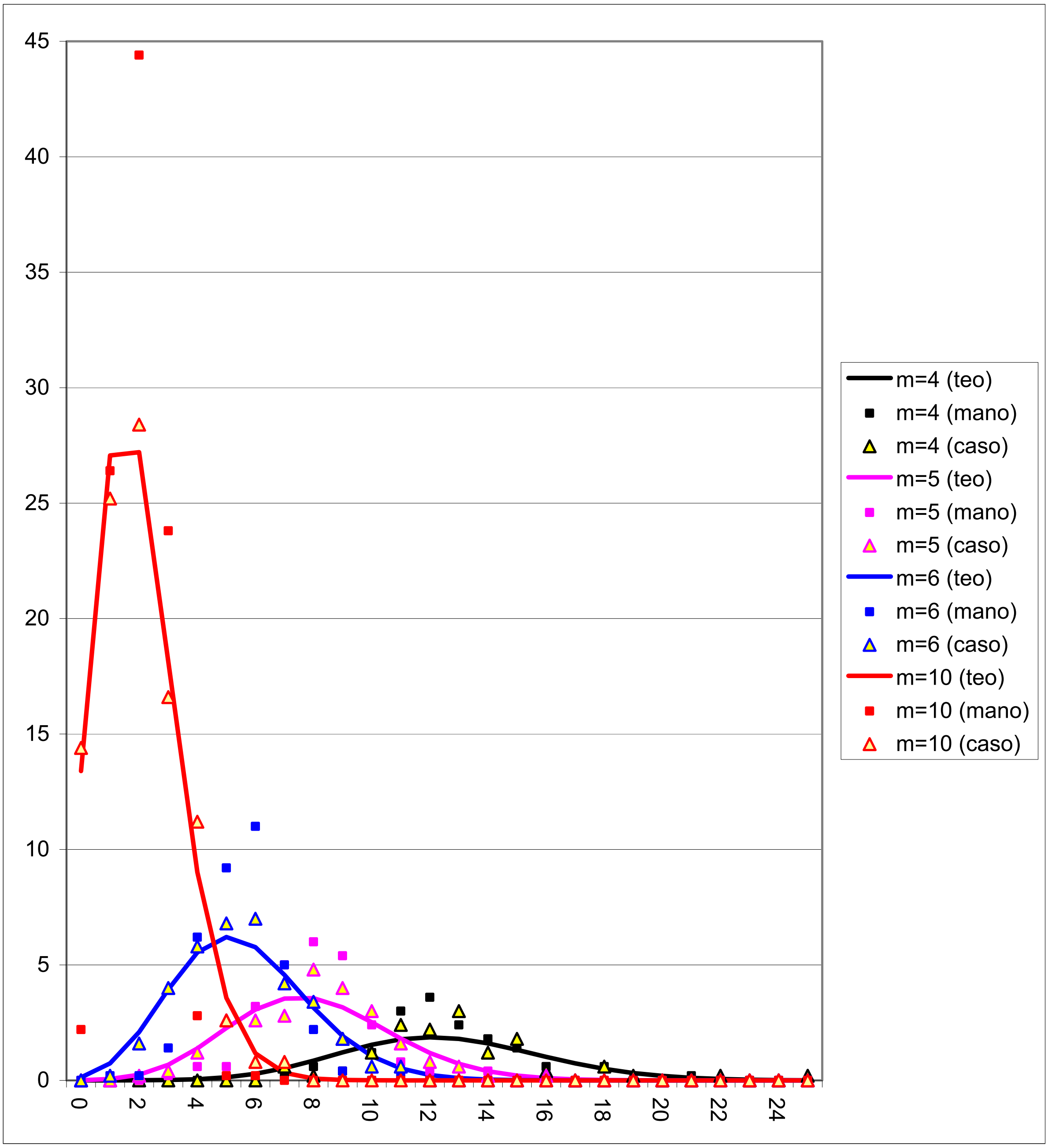








Nella tabella ***Confronti.pdf*** viene realizzato un confronto analitico tra le due serie di distribuzioni di punti. I valori della probabilità ricavati da χ^2 sono calcolati considerando solo frequenze nelle distribuzioni teoriche non troppo inferiori al 5% (limitazione del tutto standard, che corrisponde nel caso m = 4 al valore 0,8, nel caso m = 5 a 1,25, nel caso m = 6 a 1,8 nel caso m = 10 al valore 5).
Come si può osservare, mentre per le serie ottenute con Excel, utilizzando la funzione =CASUALE(), la probabilità che possano essere imputate al caso le differenze tra distribuzione teorica e media delle distribuzioni osservate, si mantiene grosso modo costante e superiore al 95%, per le serie "ottenute a mano" tale probabilità decresce rapidamente al crescere di m fino a raggiungere il valore nullo per m = 10. Non sorprenda poi il fatto che le medie siano spesso molto più vicine alla distribuzione teorica che non i singoli casi; nelle medie, gli scostamenti dal comportamento previsto teoricamente tendono a bilanciarsi a vicenda. Il seguente grafico fornisce un'idea di come le serie casuali (caso) si adattino al comportamento teorico (teo) assai meglio di quelle tracciate ad occhio (mano).



Come si ottengono i risultati esposti in questo paragrafo?
Basta aprire il documento ***200 punti.xls***:

1) nelle colonne **A** e **B** sono contenute le coordinate dei punti (nel documento in esame in tutte le celle vi è la funzione =CASUALE() che fornisce un numero (pseudo)casuale compreso tra 0 e 1; bisogna fare attenzione che, ad ogni invio la funzione =CASUALE() cambia valori: quindi se si vuole conservare un file con una determinata scelta dei 200 punti casuali, si devono copiare le due colonne **A** e **B** ed incollarle sopra sé stesse con ***Incolla speciale > Valori*** (in questo modo vengono conservati solo i valori della funzione come numeri e non come risultato di una formula);

2) le colonne **C, D** ed **E** sono dedicate alla determinazione del quadratino in cui il punto considerato va a finire: ad esempio, nel caso $m = 4$, in cui si è diviso il quadrato in 16 quadratini, numeriamo questi ultimi da 0 a 15;

2a) il quadratino 0 conterrà tutti i punti con $0 \le x < 0{,}25$ e $0 \le y < 0{,}25$: moltiplicando entrambe le coordinate per m (4, nel nostro caso), otteniamo $0 \le 4x < 1$ e $0 \le 4y < 1$ e quindi la parte intera (cioè davanti alla virgola) sia di $4x$ che di $4y$ sarà nulla e così pure la somma della prima parte intera, moltiplicata per m, più la seconda varrà 0, che è proprio il numero del quadratino;

2b) il quadratino 2 conterrà tutti i punti con $0 \le x < 0{,}25$ e $0{,}25 \le y < 0{,}5$: moltiplicando entrambe le coordinate per m, otteniamo $0 \le 4x < 1$ e $1 \le 4y < 2$ e quindi la parte intera di $4x$ è 0, mentre quella di $4y$ vale 1: quindi 4 volte la prima parte intera più la seconda dà come risultato 1, che è proprio il numero del quadratino;

2c) il quadratino 6 conterrà tutti i punti con $0{,}25 \le x < 0{,}5$ e $0{,}5 \le y < 0{,}75$: moltiplicando ancora per m, otteniamo $1 \le 4x < 2$ e $2 \le 4y < 3$ e quindi la parte intera di $4x$ è 1, mentre quella di $4y$ vale 2: quindi la somma del quadruplo della prima di tali parti intere con la seconda vale 6, che è proprio il numero del quadratino;

2d) e così via.

Quanto descritto è realizzato nelle colonne **C, D** ed **E** del foglio ***200 punti.xls***: in **C2** si scrive =ARROTONDA.DIFETTO(C1*A2;1) che fornisce proprio la parte intera di $4x$ (si ricorda che in **C1** vi è il valore $m = 4$); analogamente in **D2** con =ARROTONDA.DIFETTO(C1*B2;1); la somma

$$4\times[4x]+[4y]$$

è realizzata da =C2*C1+D2 in **E2**. Quindi nella colonna **E** è indicato il numero del quadratino a cui il punto appartiene;

3) la colonna **F** contiene i numeri dei singoli quadratini; nella colonna **G** è indicato il numero di volte in cui i numeri corrispondenti ai vari quadratini compaiono nella colonna **E** (cioè il numero di punti contenuto in ogni quadratino); per far ciò si è usata la funzione =CONTA.SE(*intervallo*;*criteri*), che conta il numero di celle in un intervallo che soddisfano i criteri specificati. Nel nostro caso poniamo =CONTA.SE(E2:E201;F2) in **G2** che restituisce il numero di volte nell'intervallo **E2:E201** in cui compare il valore contenuto in **F2**, che è 0; copiando e incollando si ottiene il numero di punti contenuti in ogni quadratino.

4) la colonna **H** contiene un possibile range di valori per il numero di punti che cadono nei vari quadratini: si è scelto l'intervallo da 0 a 25; nella colonna **I** è calcolato il numero di quadratini in cui cade un numero di punti eguale al valore contenuto nella cella immediatamente a sinistra; così in **I2** è contenuta l'espressione

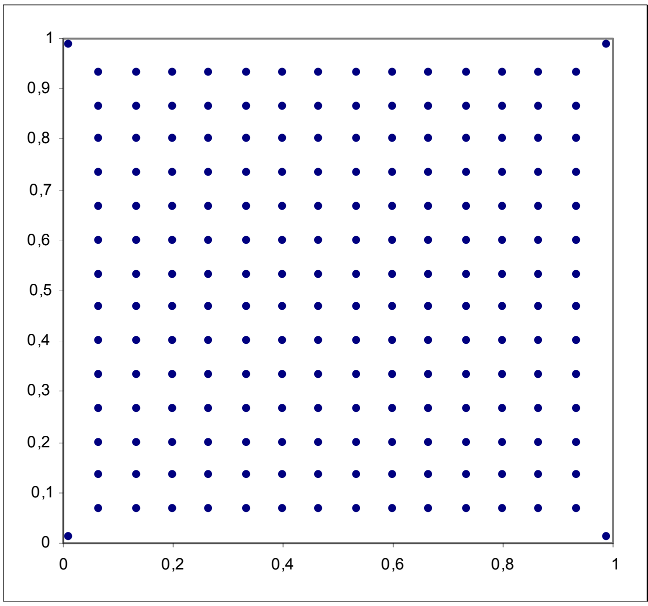
=CONTA.SE(G2:G17;H2)

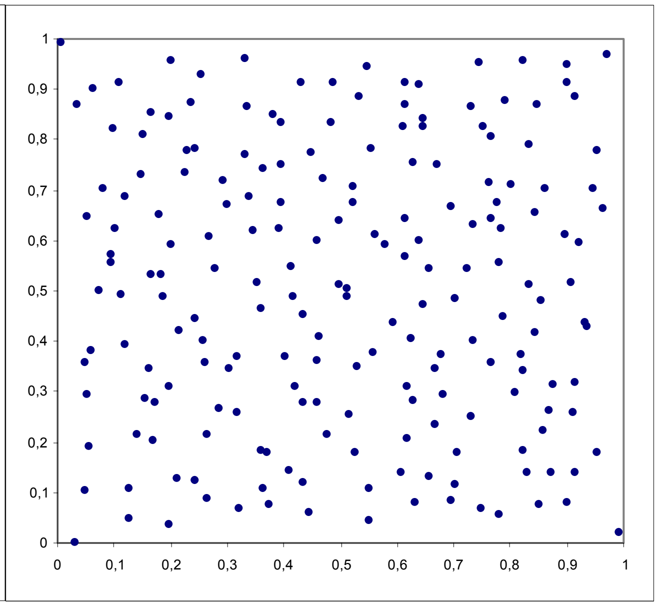
che mi restituisce il numero di volte in cui, nella colonna **G**, compare il valore contenuto in **H2**, che è 0 (ovvero ottengo il numero dei quadratini in cui non cade alcun punto); copia e incolla ci permette di sapere in quanti quadratini non cadono punti, in quanti ne cade uno, in quanti due e così via.

5) le successive colonne sono organizzate nello stesso modo, ma per differenti valori di m.

6) il secondo foglio (***confronti***) permette di confrontare le distribuzioni osservate con quella teorica tramite l'applicazione del test χ^2 alle frequenze più significative.

Per completare il quadro si consideri una distribuzione uniforme (a sinistra) ed una ottenuta da questa tramite una permutazione casuale di entità pari al massimo alla metà della distanza tra i punti della distribuzione di partenza (a destra). Nella tabella si vede come, nel primo caso, per tutti i valori di *m*, la probabilità che tale distribuzione capiti casualmente è nulla, mentre per la seconda si passa da un valore del 54% per $m = 4$ ad una inferiore allo 0,05‰ per $m = 10$.





	m=4			*m*=5			*m*=6			*m*=10		
k	teo	unif	pert	teo	unif	pert	teo	unif	pert	teo	unif	pert
0	0,00	0	0	0,01	0	0	0,13	0	0	13,40	0	0
1	0,00	0	0	0,06	0	0	0,74	0	0	27,07	32	30
2	0,00	0	0	0,25	0	0	2,09	0	0	27,20	52	47
3	0,02	0	0	0,68	0	0	3,94	0	0	18,14	0	16
4	0,05	0	0	1,39	0	0	5,55	12	5	9,02	16	7
5	0,13	0	0	2,27	1	1	6,21	4	12	3,57	0	0
6	0,29	0	0	3,07	6	4	5,77	16	14	1,17	0	0
7	0,53	0	0	3,54	2	5	4,57	0	4	0,33	0	0
8	0,85	0	0	3,56	0	8	3,15	0	1	0,08	0	0
9	1,21	0	0	3,17	15	2	1,92	4	0	0,02	0	0
10	1,54	4	2	2,52	1	3	1,05	0	0	0,00	0	0
11	1,78	0	3	1,81	0	1	0,52	0	0	0,00	0	0
12	1,87	8	4	1,19	0	0	0,23	0	0	0,00	0	0
13	1,80	0	3	0,72	0	1	0,10	0	0	0,00	0	0
14	1,60	0	1	0,40	0	0	0,04	0	0	0,00	0	0
15	1,33	0	2	0,21	0	0	0,01	0	0	0,00	0	0
16	1,02	4	1	0,10	0	0	0,00	0	0	0,00	0	0
17	0,74	0	0	0,04	0	0	0,00	0	0	0,00	0	0
18	0,50	0	0	0,02	0	0	0,00	0	0	0,00	0	0
19	0,32	0	0	0,01	0	0	0,00	0	0	0,00	0	0
20	0,19	0	0	0,00	0	0	0,00	0	0	0,00	0	0
21	0,11	0	0	0,00	0	0	0,00	0	0	0,00	0	0
22	0,06	0	0	0,00	0	0	0,00	0	0	0,00	0	0
23	0,03	0	0	0,00	0	0	0,00	0	0	0,00	0	0
24	0,02	0	0	0,00	0	0	0,00	0	0	0,00	0	0
25	0,01	0	0	0,00	0	0	0,00	0	0	0,00	0	0
$\chi^2\rightarrow$		0,00%	54,04%		0,00%	22,64%		0,00%	0,04%		0,00%	0,00%