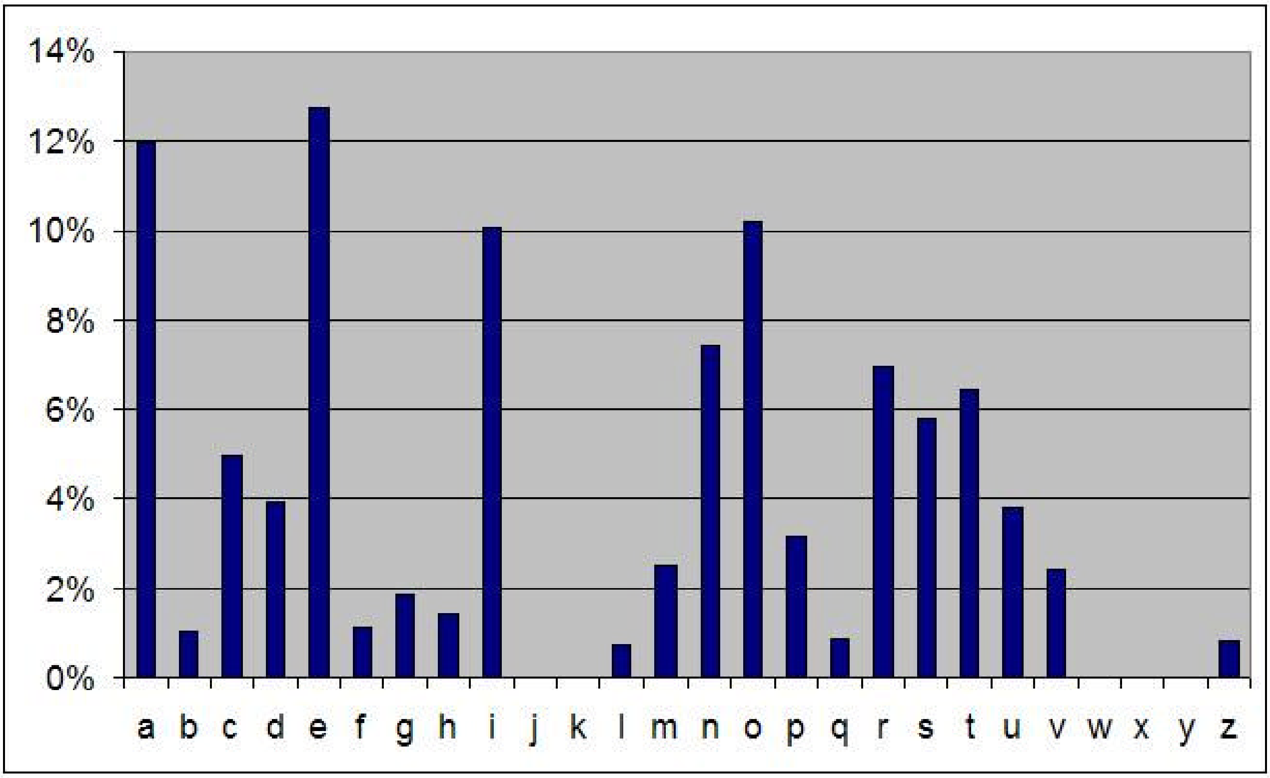
Uso di EXCEL per la modellazione e la soluzione di problemi finanziari, aziendali, economici

Università Bocconi
Anno Accademico 2006-2007
Matematica Finanziaria cod. 6008
Michele Impedovo



Introduzione

Il *Progetto di miglioramento della didattica* per il corso di Matematica Finanziaria cod. 6008 è stato preparato nel mese di dicembre 2006.
Il progetto comprende i seguenti materiali.

Testo

È questo manuale, contenuto nel file
`Progetto.pdf`,
che comprende la descrizione di tutte le attività, strutturate in 7 capitoli indipendenti, ciascuno dei quali è organizzato in paragrafi. Ogni paragrafo descrive un problema e costituisce, in linea di massima, un'unità di lavoro.
L'indice nella pagina precedente illustra la struttura del lavoro.

File Excel

In corrispondenza di ciascun paragrafo è stato costruito il relativo file di Excel illustrato nel testo. Si è utilizzata le versione di Microsoft Office 2003. Le funzioni e i comandi utilizzati sono quelli della versione italiana; per le corrispondenti funzioni della versione inglese si consulti consultare il file
`cross reference ita eng.txt`
allegato al materiale.

Tutorial

Per ogni paragrafo (e quindi per ogni file Excel) sono state costruite della animazioni audio-video che fungono da tutorial dell'intero lavoro e che illustrano passo-passo la costruzione del foglio di lavoro di Excel a cui fanno riferimento.
Ogni animazione dura pochi minuti. Ogni animazione è costituita da due file con lo stesso nome: uno in formato `.htm` (di dimensioni ridotte) e uno in formato `.swf`, che funge da libreria.
Per vedere l'animazione occorre cliccare due volte sul file `.htm` ; è sufficiente un browser qualsiasi.
Nella parte bassa del video è presente una barra di controllo che permette di interrompere, rivedere e riascoltare, avanzare veloce, saltare ad un dato punto, e così via.

Il consiglio è quello di leggere prima il testo, poi guardare il tutorial e infine aprire il relativo file di Excel.
L'obiettivo è che questo materiale possa servire:

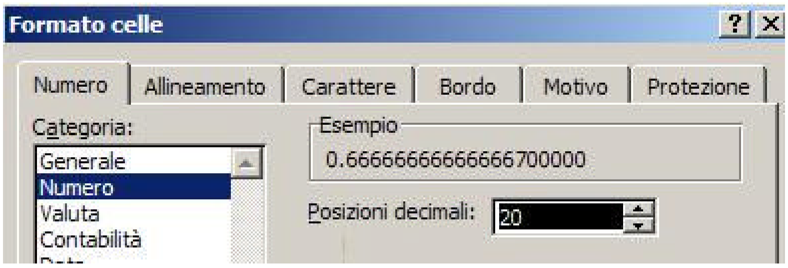
- a esplorare numericamente alcuni concetti, in modo da consolidarne la padronanza;
- ad approfondire le abilità algoritmiche e di utilizzo di strumenti automatici di calcolo;
- ad utilizzare i fogli di Excel come veri e propri strumenti di indagine e di lavoro.

È gradito qualunque feed-back; si prega di scrivere a
michele.impedovo@unibocconi.it.

1. Funzionalità di base di Excel per la matematica

1.1 Calcolo simbolico e calcolo numerico

Excel è un foglio elettronico: non è in grado di gestire simboli, ma esclusivamente numeri in forma approssimata. Ci si può rendere conto di ciò se scriviamo in una cella
=1/3
e premiamo invio. Appare 0.333333 (con 6 cifre decimali uguali a 3, oppure 9 cifre decimali se allarghiamo la cella): non è 1/3, è l'approssimazione di 1/3. Nello stesso modo se scriviamo
=2/3
otteniamo 0.666667: dunque non c'è troncamento, ma arrotondamento; l'ultima cifra è aumentata di 1 se la cifra successiva è maggiore o uguale a 5.
Se clicchiamo su Formato $\rightarrow$ Celle… $\rightarrow$ Numero osserviamo che di default Excel imposta per il risultato il formato "Generale". Scegliamo invece il formato "Numero" e aumentiamo il numero di "Posizioni decimali": possiamo osservare nella riga soprastante (Esempio) che Excel lavora con 15 cifre decimali (approssimando l'ultima): dalla sedicesima cifra decimale in poi otteniamo solo zeri.



Se si vuole analizzare più in dettaglio la cosiddetta "precisione di macchina", cioè l'approssimazione utilizzata da Excel, possiamo operare nel seguente modo.
Osserviamo innanzitutto che una formula come la seguente

=1+1=2

restituisce
VERO
mentre

=1+1=3 restituisce

FALSO.
Il primo "=" è una richiesta di valutazione della formula che segue, il secondo è un uguale logico; dunque viene valutata l'uguaglianza.
Ora nella colonna A scriviamo i numeri 1, 2, …, 20. A questo scopo è sufficiente scrivere 1 in A1, 2 in A2, selezionare le celle A1:A2 e trascinare verso il basso: si genera automaticamente una sequenza di valore iniziale A1 e passo uguale all'incremento A2−A1.
In B1 scriviamo la formula
=1+1/10^A1=1
che restituirà FALSO, dato che in A1 c'è 1 e $1+1/10=1.1\neq 1$.
Ora copiamo la formula della cella B1 verso il basso fino a B20 (selezionare B1, prendere con il cursore il quadratino in basso a destra della cella B1 e trascinarlo verso il basso; oppure, più semplicemente, doppio click sul quadratino).
Come è noto, copiando la formula verso il basso i riferimenti di cella (che sono relativi) si aggiornano: la formula "=1+1/10^A1=1" in B2 diventa "=1+1/10^A2=1", in B3 diventa "=1+1/10^A3=1", e così via. Nella colonna B abbiamo così costruito la successione

$1+\dfrac{1}{10^n}$, con $n = 1, 2, \ldots, 20$.

Al crescere di n l'addendo $1/10^n$ tende esponenzialmente a 0. Dal punto di vista simbolico l'uguaglianza è sempre falsa, perché per quanto grande sia n risulta comunque $1/10^n > 0$. Ma dal punto di vista numerico le cose cambiano: per $n = 15$ l'uguaglianza diventa vera: la successione

1.1
1.01
1.001
1.0001
…

sposta ad ogni passo la cifra decimale 1 verso destra, fino a che Excel sarà costretto a troncarla, ottenendo 1 e valutando VERO il contenuto della cella B15. Da B15 in poi l'uguaglianza sarà sempre vera.

	A	B
11	11	FALSO
12	12	FALSO
13	13	FALSO
14	14	FALSO
15	15	VERO
16	16	VERO
17	17	VERO
18	18	VERO
19	19	VERO
20	20	VERO

Se si ripete l'esperimento con 1/2 anziché 1/10 (ricordiamo che il calcolo interno è sempre in base 2), cioè se valutiamo la successione

$1+\dfrac{1}{2^n}$,

otteniamo che il primo VERO si ottiene per $n = 48$. Dunque vengono utilizzati 48 bit per la rappresentazione della parte decimale di un numero (*mantissa*); il numero $2^{-47} \approx 7 \cdot 10^{-15}$ viene chiamato "precisione di macchina" del sistema di calcolo e rappresenta il più piccolo numero positivo ε per il quale

$$1+\varepsilon \neq 1.$$

Ci si può chiedere: ma un'approssimazione alla quindicesima cifra decimale non è più che sufficiente per qualsiasi calcolo? Potrebbe non essere così, come mostra il seguente esempio.

Nella colonna A scriviamo i numeri 0, 1, 2, …, 30. In B1 scriviamo "=1/3", ottenendo 0.333333.

In B2 scriviamo la formula "=4*B1–1": poichè $4 \cdot \dfrac{1}{3} - 1 = \dfrac{1}{3}$, dovremmo ottenere in B2 ancora 1/3.

Ora copiamo la formula della cella B2 verso il basso fino a B31.

In questo modo abbiamo costruito i primi elementi di una successione il cui primo elemento è 1/3, e ogni successivo elemento è il quadruplo del precedente diminuito di 1:

$a_0 = 1/3$

$a_{n+1} = 4a_n + 1$

Dal punto di vista simbolico la successione che abbiamo costruito è costante di valore 1/3. Però Excel non possiede 1/3, ma solo una sua approssimazione. Osserverete che già per $n = 18$ la propagazione dell'errore è risalita alla sesta cifra decimale, e la successione, anziché rimanere costante, come dovrebbe, diverge rapidamente a $-\infty$.

	A	B
19	18	0.333332
20	19	0.333328
21	20	0.333313
22	21	0.333252
23	22	0.333008
24	23	0.332031
25	24	0.328125
26	25	0.3125
27	26	0.25
28	27	0
29	28	-1
30	29	-5
31	30	-21

Lo stesso accade per la successione

$a_0 = 1/5$

$a_{n+1} = 6a_n + 1$

nonostante il numero $1/5 = 0.2$ non abbia, in base 10, infinite cifre decimali come 1/3. Infatti il processore numerico lavora in base 2 (non in base 10), e la rappresentazione decimale di 1/5 è, in base 2, periodica quanto quella di 1/3. Invece la successione

$a_0 = 1/4$

$a_{n+1} = 5a_n + 1$

risulta effettivamente costante, perché la sua rappresentazione decimale è finita anche in base 2.

1.2 Successioni e leggi finanziarie

La struttura di Excel si presta in particolare ad analizzare, sia dal punto di vista numerico che grafico, successioni, definite sia mediante una legge generale, sia in forma ricorsiva. Vediamo entrambi questi metodi in dettaglio su un esempio finanziario: l'impiego di un capitale iniziale di 1000 €, in capitalizzazione semplice (lineare) al tasso annuo 9%, e in capitalizzazione composta (esponenziale) al tasso annuo 6%, per 20 anni.

Legge generale

Capitalizzazione semplice $C_t = 1000(1+0.09t)$

Capitalizzazione composta $C_t = 1000 \cdot 1.06^t$

Scriviamo in A1:A21 i numeri 0, 1, …, 20; in B1 scriviamo la formula

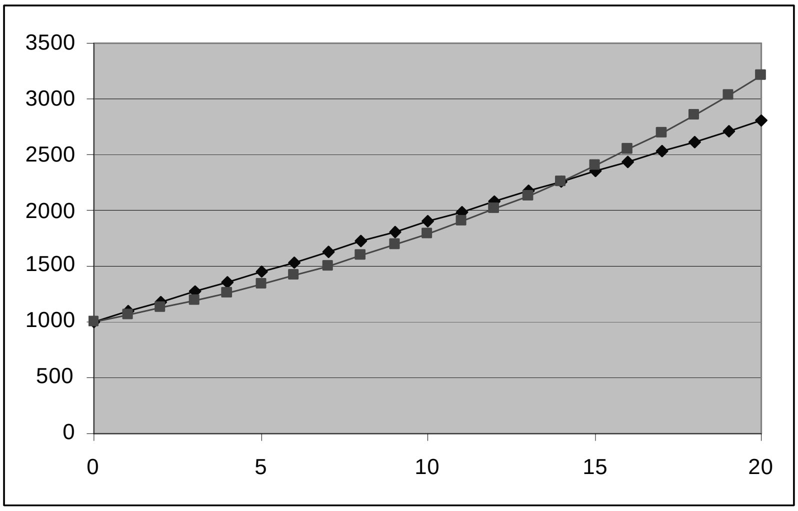
=1000*(1+0.09*A1)

e in C1 la formula

=1000*1.06^A1

	A	B	C
1	0	1000	1000
2	1	1090	1060
3	2	1180	1123.60
4	3	1270	1191.02
5	4	1360	1262.48
6	5	1450	1338.23
7	6	1540	1418.52
8	7	1630	1503.63
9	8	1720	1593.85
10	9	1810	1689.48
11	10	1900	1790.85
12	11	1990	1898.30
13	12	2080	2012.20
14	13	2170	2132.93
15	14	2260	2260.90
16	15	2350	2396.56
17	16	2440	2540.35
18	17	2530	2692.77
19	18	2620	2854.34
20	19	2710	3025.60
21	20	2800	3207.14

Selezionando le celle A1:C21, cliccando sull'icona Creazione guidata grafico, e scegliendo un grafico a dispersione (*Scatter*, in inglese), si ottiene il grafico seguente.



Quando si selezionano più colonne Excel di deafult mette i valori della prima colonna sull'asse x e i valori delle altre colonne sull'asse y, come grafici distinti.
Cliccando con il tasto destro del mouse sul grafico si accede ad un ricco menù di editing e di formattazione del grafico stesso.
Come si vede, mentre in capitalizzazione semplice il montante cresce in modo lineare, in capitalizzazione composta cresce via via più rapidamente e dopo 14 anni, nonostante il minor tasso, supera definitivamente l'altro.

Legge ricorsiva

Possiamo ottenere esattamente la stessa tabella se descriviamo l'andamento dei due montanti in forma ricorsiva, anziché mediante la legge generale:

Capitalizzazione semplice $C_{t+1} = C_t + 0.09 \cdot 1000$

Capitalizzazione composta $C_{t+1} = C_t + 0.06 C_t$

In D1 e in E1 scriviamo i valori iniziali delle due successioni: 1000.
In D2 scriviamo la legge ricorsiva

=D1+90

che aumenta di una quantità costante il valore della cella al tempo precedente (aumento assoluto).
In E2 scriviamo la legge ricorsiva

=E1+0.06*D1

che aumenta il valore della cella al tempo precedente di una percentuale costante (aumento relativo).
Copiando le celle D2 e E2 verso il basso otteniamo esattamente a stessa tabella costruita in A1:B21.

1.3 Parametri e riferimenti assoluti

Supponiamo ora di voler cambiare i tassi, oppure i capitali iniziali. Possiamo rendere più "elastico" il foglio di lavoro utilizzando delle celle di input con i parametri iniziali, e usare dei riferimenti assoluti ad esse.
Nelle celle A2 e A4 scriviamo rispettivamente il capitale iniziale e il tasso annuo dell'impiego in capitalizzazione semplice (per esempio 1000 e 0.06); in B2 e B4 il capitale iniziale e il tasso annuo della capitalizzazione composta (per esempio 500 e 0.14).

	A	B
1	**A[0]**	**B[0]**
2	1000	500
3	**iA**	**iB**
4	0.06	0.14

Nella colonna C scriviamo i numeri 0, 1, …, 20.
Nella cella D2 scriviamo la formula

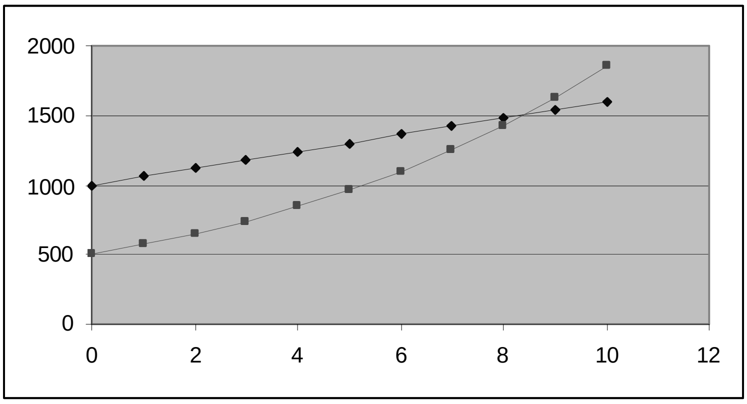
=$A$2*(1+$A$4*C2)

e nella cella E2 la formula

=$B$2*(1+$B$4)^C2

e come al solito copiamo verso il basso le due formule.

	C	D	E
1	**t**	**A[t]**	**B[t]**
2	0	1000	500
3	1	1060	570
4	2	1123.60	649.80
5	3	1191.02	740.77
6	4	1262.48	844.48
7	5	1338.23	962.71
8	6	1418.52	1097.49
9	7	1503.63	1251.13
10	8	1593.85	1426.29
11	9	1689.48	1625.97
12	10	1790.85	1853.61



Il simbolo "$" trasforma un riferimento relativo in un riferimento assoluto.
Se una formula contiene il riferimento $A$2 allora tale riferimento rimane invariato sia copiando la formula in verticale, sia in orizzontale[1].
Se il cursore si trova in un punto qualsiasi di un riferimento di cella, per esempio A2, il comando da tastiera F4 lo trasforma ciclicamente nel modo seguente:

[1] Il riferimento A$2 blocca la riga 2, cioè rimane invariato copiando in verticale (e invece si trasforma in B2, C2, … copiando in orizzontale). Il riferimento $A2 blocca la colonna A, cioè rimane invariato copiando in orizzontale (e invece si trasforma in A3, A4, … copiando in verticale).

A2 $\rightarrow$ \$A\$2 $\rightarrow$ A\$2 $\rightarrow$ \$A2 $\rightarrow$ A2 $\rightarrow$ …

Il riferimento assoluto ad una o più celle consente di aggiornare istantaneamente il foglio di lavoro modificando tali celle. Nel nostro esempio è sufficiente modificare i capitali iniziali o i tassi di impiego per ottenere la nuova tabella e il nuovo grafico. Tutto ciò rende il foglio di Excel più flessibile ed efficace.

1.4 Funzioni

Excel consente di analizzare il comportamento di una funzione in un certo intervallo, sia dal punto di vista numerico sia grafico.

Supponiamo di voler analizzare la funzione $f(x) = xe^{-x}$ nell'intervallo [0,2].

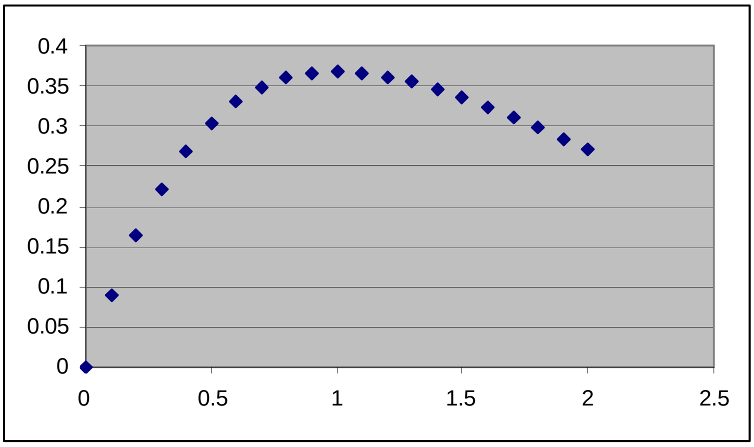
Scegliamo come passo 0.1 e nella colonna A scriviamo i numeri 0, 0.1, …, 2. In B1 scriviamo la funzione, nel linguaggio di Excel:

=A1*exp(–A1)

e copiamo verso il basso. Otteniamo la tabella seguente, che mostra un punto di massimo in 1.

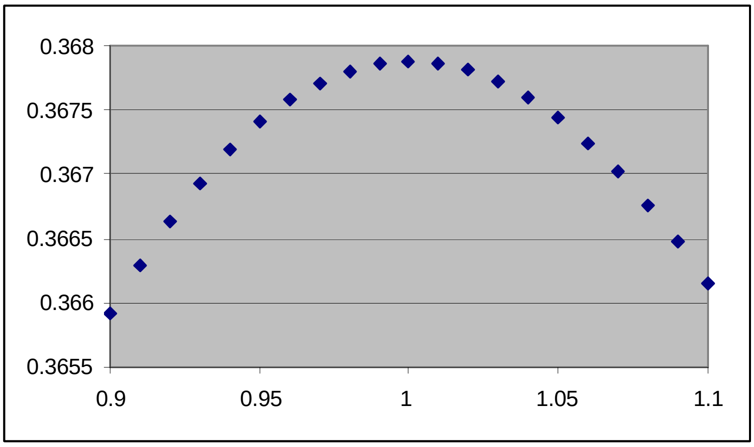
	A	B
1	0	0
2	0.1	0.090484
3	0.2	0.163746
4	0.3	0.222245
5	0.4	0.268128
6	0.5	0.303265
7	0.6	0.329287
8	0.7	0.34761
9	0.8	0.359463
10	0.9	0.365913
11	1	0.367879
12	1.1	0.366158
13	1.2	0.361433
14	1.3	0.354291
15	1.4	0.345236
16	1.5	0.334695
17	1.6	0.323034
18	1.7	0.310562
19	1.8	0.297538
20	1.9	0.28418
21	2	0.270671

Selezioniamo la tabella A1:B21, clicchiamo su Inserisci, Grafico, Dispersione. Otteniamo il grafico di $f(x)$, a passi discreti di 0.1, da 0 a 2.



Naturalmente è un grafico per punti: in linea di principio non sappiamo che cosa accade tra 0 e 01, tra 0.1 e 0.2, e così via. Tuttavia se la funzione è regolare possiamo avere un'idea sufficientemente precisa dell'andamento della funzione. Soprattutto possiamo raffinare l'analisi quanto vogliamo. Ecco il comportamento di $f(x)$ tra 0.9 e 1.1 con passo 0.01.

	A	B
1	0.9	0.365913
2	0.91	0.366297
3	0.92	0.366638
4	0.93	0.366935
5	0.94	0.36719
6	0.95	0.367404
7	0.96	0.367577
8	0.97	0.367711
9	0.98	0.367805
10	0.99	0.367861
11	1	0.367879
12	1.01	0.367861
13	1.02	0.367807
14	1.03	0.367717
15	1.04	0.367593
16	1.05	0.367435
17	1.06	0.367243
18	1.07	0.367019
19	1.08	0.366763
20	1.09	0.366476
21	1.1	0.366158



1.5 Funzioni in due variabili e tabelle a doppia entrata

Possiamo sfruttare quanto visto in 1.3 a proposito dei riferimenti assoluti per costruire una tabella a doppia entrata.
Supponiamo di voler analizzare la funzione in due variabili
$f(i, t) := (1+i)^t$
che fornisce, in funzione del tasso composto annuo i e del tempo di impiego t, il montante in t di un capitale iniziale unitario.
Stabiliamo di analizzare la funzione f per i compreso tra 3% e 10% con passo 1% e per t compreso tra 1 e 8 anni, con passo 1. Il capitale iniziale C_0 è considerato un parametro, e perciò utilizzato nel foglio di lavoro, nella cella A2, come riferimento assoluto: è sufficiente modificare A2 per aggiornare la tabella a C qualsiasi.

	A	B	C	D	E	F	G	H	I	J
1	**C[0]**		**3%**	**4%**	**5%**	**6%**	**7%**	**8%**	**9%**	**10%**
2	1	**1**	1.03	1.04	1.05	1.06	1.07	1.08	1.09	1.1
3		**2**	1.0609	1.0816	1.1025	1.1236	1.1449	1.1664	1.1881	1.21
4		**3**	1.09273	1.12486	1.15763	1.19102	1.22504	1.25971	1.29503	1.331
5		**4**	1.12551	1.16986	1.21551	1.26248	1.3108	1.36049	1.41158	1.4641
6		**5**	1.15927	1.21665	1.27628	1.33823	1.40255	1.46933	1.53862	1.61051
7		**6**	1.19405	1.26532	1.3401	1.41852	1.50073	1.58687	1.6771	1.77156
8		**7**	1.22987	1.31593	1.4071	1.50363	1.60578	1.71382	1.82804	1.94872
9		**8**	1.26677	1.36857	1.47746	1.59385	1.71819	1.85093	1.99256	2.14359

La formula su cui si basa l'intera tabella è nella cella C2:
=$A$2*(1+C$1)^$B2
Si osservi la presenza dei riferimenti *misti* (né assoluti né relativi) dei tre tipi: C$1 blocca la riga 1, $B2 blocca la colonna B, $A$2 blocca la cella A2.

1.6 Struttura a termine dei tassi di interesse

Sul sito www.euribor.org è disponibile, quotidianamente aggiornata, la struttura a termine dei tassi *spot* (con scadenze che vanno da 1 settimana a 1 anno), che i mercati finanziari europei forniscono come previsione di massima. Sono i tassi che le banche tipicamente utilizzano come riferimento per la concessione di finanziamenti a lungo termine.
Utilizziamo i dati forniti da Euribor per costruire la tabella a doppia entrata dei corrispondenti tassi *forward*.
In data 11/12/06 i tassi spot $h(0,t)$ (cioè i tassi per impieghi certi attivati oggi e con scadenza t) sono i seguenti (il tempo è misurato in mesi).

	A	B	C
1	**Date**	**scadenza**	**Tasso spot**
2	1 week	**0.25**	3.551%
3	2 week	**0.5**	3.583%
4	3 week	**0.75**	3.643%
5	1 month	**1**	3.634%
6	2 month	**2**	3.645%
7	3 month	**3**	3.669%
8	4 month	**4**	3.710%
9	5 month	**5**	3.742%
10	6 month	**6**	3.767%
11	7 month	**7**	3.795%
12	8 month	**8**	3.819%
13	9 month	**9**	3.840%
14	10 month	**10**	3.852%
15	11 month	**11**	3.873%
16	12 month	**12**	3.884%

Vogliamo ricavare, per ogni coppia di tassi spot $h(0,s)$ e $h(0,t)$, con $s<t$, il corrispondente tasso forward da s a t: $h(s,t)$.
Poiché in capitalizzazione composta risulta

$$(1+h(0,s))^s(1+h(s,t))^{t-s} = (1+h(0,t))^t$$

allora per il tasso forward risulta

$$h(s,t) = \left(\frac{(1+h(0,t))^t}{(1+h(0,s))^s}\right)^{\frac{1}{t-s}} - 1.$$

In Excel scriviamo sia nelle celle C1:Q1 sia nelle celle A3:A17 le scadenze (in mesi) 1/4, 2/4, 3/4, 1, 2, …, 12.
Sia nelle celle C2:Q2 sia nelle celle B3:B17 riportiamo i tassi spot $h(0,t)$.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1		**t**	**0.25**	**0.5**	**0.75**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
2	**t**	**h(0,t)**	**3.551%**	**3.583%**	**3.643%**	**3.634%**	**3.645%**	**3.669%**	**3.710%**	**3.742%**	**3.767%**	**3.795%**	**3.819%**	**3.840%**	**3.852%**	**3.873%**	**3.884%**
3	**0.25**	**3.551%**															
4	**0.5**	**3.583%**															
5	**0.75**	**3.643%**															
6	**1**	**3.634%**															
7	**2**	**3.645%**															
8	**3**	**3.669%**															
9	**4**	**3.710%**															
10	**5**	**3.742%**															
11	**6**	**3.767%**															
12	**7**	**3.795%**															
13	**8**	**3.819%**															
14	**9**	**3.840%**															
15	**10**	**3.852%**															
16	**11**	**3.873%**															
17	**12**	**3.884%**															

Vogliamo ora riempire la tabella in modo da leggere in una cella qualsiasi, il tasso forward; per esempio nella cella K7 (evidenziata nella figura precedente), leggeremo il tasso forward $h(2,6)$, cioè il tasso che il mercato implicitamente prevede per impieghi futuri che vengano attivati fra 2 mesi, con scadenza tra 6 mesi. Poiché il tasso $h(s, t)$ è definito soltanto se $s < t$, utilizzeremo nella formula la struttura di controllo SE, la cui sintassi è

SE(condizione; *se vero*; *se falso*)

dove *se vero* indica il comando che verrà eseguito nel caso la condizione sia vera, *se falso* indica il comando che verrà eseguito nel caso la condizione sia falsa. La condizione è che la scadenza nella colonna A sia minore della scadenza nella riga 1; poiché in caso contrario vogliamo che la cella resti vuota, il comando *se falso* nel nostro caso consisterà in una coppia di doppie virgolette (""), che indicano, in Excel, la cosiddetta *stringa vuota*, cioè un testo … senza alcun carattere.
Nella cella C3 scriviamo la formula che copieremo su tutta la tabella.

=SE($A3<C$1;((1+C$2)^C$1/(1+$B3)^$A3)^(1/(C$1-$A3))-1;"")

Si osservi il ruolo fondamentale (tipico delle tabelle a doppia entrata) dei riferimenti misti: nel riferimento alle celle A3 e B3 occorre, quando si copierà verso destra, bloccare la colonna (questo spiega i riferimenti $A3 e $B3) mentre nelle celle C1 e C2 occorre, quando si copierà verso il basso, bloccare la riga (questo spiega i riferimenti C$1 e C$2).
Naturalmente la cella C3 resta vuota, dato che la condizione A3<C1 risulta falsa. Copiamo la cella C3 verso destra fino a Q3. Poi selezioniamo la riga C3:Q3 e copiamola verso il basso (oppure doppio click sul quadratino in basso a destra di Q3) fino alla riga 17.
Ecco il risultato.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1		t	0.25	0.5	0.75	1	2	3	4	5	6	7	8	9	10	11	12
2	t	h(0,t)	3.551%	3.583%	3.643%	3.634%	3.645%	3.669%	3.710%	3.742%	3.767%	3.795%	3.819%	3.840%	3.852%	3.873%	3.884%
3	0.25	3.551%		3.615%	3.689%	3.662%	3.658%	3.680%	3.721%	3.752%	3.776%	3.804%	3.828%	3.848%	3.860%	3.881%	3.891%
4	0.5	3.583%			3.763%	3.685%	3.666%	3.686%	3.728%	3.760%	3.784%	3.811%	3.835%	3.855%	3.866%	3.887%	3.897%
5	0.75	3.643%				3.607%	3.646%	3.678%	3.725%	3.759%	3.785%	3.813%	3.837%	3.858%	3.869%	3.890%	3.900%
6	1	3.634%					3.656%	3.687%	3.735%	3.769%	3.794%	3.822%	3.845%	3.866%	3.876%	3.897%	3.907%
7	2	3.645%						3.717%	3.775%	3.807%	3.828%	3.855%	3.877%	3.896%	3.904%	3.924%	3.932%
8	3	3.669%							3.833%	3.852%	3.865%	3.890%	3.909%	3.926%	3.931%	3.950%	3.956%
9	4	3.710%								3.870%	3.881%	3.908%	3.928%	3.944%	3.947%	3.966%	3.971%
10	5	3.742%									3.892%	3.928%	3.947%	3.963%	3.962%	3.982%	3.986%
11	6	3.767%										3.963%	3.975%	3.986%	3.980%	4.000%	4.001%
12	7	3.795%											3.987%	3.998%	3.985%	4.010%	4.009%
13	8	3.819%												4.008%	3.984%	4.017%	4.014%
14	9	3.840%													3.960%	4.022%	4.016%
15	10	3.852%														4.083%	4.044%
16	11	3.873%															4.005%
17	12	3.884%															

La tabella è per metà vuota: la condizione $s < t$ è vera solo per le celle che stanno sopra la diagonale C3, D4, E5, …, Q17.
Nella cella K7, per esempio, leggiamo il tasso forward $h(2, 6) = 3.828\%$.

1.7 Somme: il VAN di un'operazione finanziaria

Excel possiede un comando che permette di sommare tutti i valori di un intervallo di celle; per esempio la formula
=SOMMA(A1:E10)
restituisce la somma delle 50 celle da A1 fino a E10 (le 5 colonne A-E e la 10 righe 1-10).
Questo comando consente di calcolare il valore attuale netto (VAN), ad un certo tasso di valutazione, di un'operazione finanziaria.
Supponiamo ad esempio di voler valutare il VAN dell'operazione schematizzata nella seguente tabella:

Tempi	0	1	2	3	4	5
Flussi	−1000	200	200	200	300	300

Il VAN di tale operazione, in funzione del tasso x di valutazione, è per definizione:

$$\mathrm{VAN}(x) = -1000 + \frac{200}{1+x} + \frac{200}{(1+x)^2} + \frac{200}{(1+x)^3} + \frac{300}{(1+x)^4} + \frac{400}{(1+x)^5}\,.$$

Inseriamo come parametro il tasso i di valutazione, per esempio 6.5%, nella cella A2, i tempi in B2:B7, i flussi in C2:C7.
Ora in D2 scriviamo la formula

=C2/(1+$A$2)^B2

che calcola il valore attuale netto del primo flusso mediante il tasso in A2 (riferimento assoluto). Copiamo da D2 fino a D7.
Possiamo ottenere il VAN totale in E2 mediante la formula

=SOMMA(D2:D7)

	A	B	C	D	E
1	**tasso**	**t**	**a[t]**	**VAN[t]**	**VAN totale**
2	0.065	**0**	-1000	-1000	-18.14
3		**1**	200	187.79	
4		**2**	200	176.33	
5		**3**	200	165.57	
6		**4**	300	233.20	
7		**5**	300	218.96	

Come si vede il VAN è negativo: se posso investire al 6.5% quell'investimento non mi conviene, corrisponde ad una perdita (valutata oggi) di 18.14.
Modificando il tasso in A2, per esempio $i = 3\%$, il VAN cambia e diventa positivo: se il tasso al quale posso accedere è solamente del 3% allora l'operazione mi produce un utile (valutato oggi) di 91.05.

	A	B	C	D	E
1	**tasso**	**t**	**a[t]**	**VAN[t]**	**VAN totale**
2	0.03	**0**	-1000	-1000	91.05
3		**1**	200	194.17	
4		**2**	200	188.52	
5		**3**	200	183.03	
6		**4**	300	266.55	
7		**5**	300	258.78	

Vedremo più avanti come calcolare il TIR (Tasso Interno di Rendimento) dell'operazione finanziaria.

1.8 Matrici

Excel possiede alcuni interessanti comandi per le operazioni tra vettori e matrici. Iniziamo dal **prodotto** di matrici; supponiamo che sia definito il prodotto **AB** tra le matrici **A** e **B**, e cioè il numero di colonne di **A** sia uguale al numero di righe di **B**. Per esempio, vogliamo calcolare il prodotto **AB** tra le matrici

$$\mathbf{A} = \begin{pmatrix} 6 & 2 & 8 \\ 2 & 0 & 2 \\ 0 & 4 & 0 \end{pmatrix}, \mathbf{B} = \begin{pmatrix} 1 & 3 & 1 \\ 4 & 3 & 6 \\ 3 & 8 & 3 \end{pmatrix}$$

Scriviamo la matrice **A** nelle celle A1:C3 e la matrice **B** nelle celle E1:G3.

	A	B	C	D	E	F	G
1	6	2	8		1	3	1
2	2	0	2		4	3	6
3	0	4	0		3	8	3

Poichè **A** ha dimensioni 3×3 e **B** ha dimensioni 3×3, il prodotto **C** = **AB** ha dimensioni 3×3. Selezioniamo un intervallo di celle di dimensione 3×3, per esempio le celle I1:K3, che conterranno il risultato. Con le celle selezionate scriviamo la formula:

=MATR.PRODOTTO(A1:C3;E1:G3)

(MMULT per la versione inglese) e premiamo CTRL+SHIFT+INVIO (che sostituisce INVIO per qualunque funzione che si applica a più di una cella) .

	A	B	C	D	E	F	G	H	I	J	K
1	6	2	8		1	3	1		38	88	42
2	2	0	2		4	3	6		8	22	8
3	0	4	0		3	8	3		16	12	24

È possibile con Excel ottenere anche la **matrice trasposta** $\mathbf{A}^{\mathbf{T}}$ di una matrice **A**, cioè la matrice che scambia le colonne con le righe. Il comando è MATR.TRASPOSTA (TRANSPOSE nella versione inglese). Per esempio, data la matrice 3×4 delle celle A1:D3, si selezionano le celle F1:H4 e il comando

=MATR.TRASPOSTA(A1:D3)

seguito da CTRL+SHIFT+INVIO, fornisce la matrice trasposta.

	A	B	C	D	E	F	G	H
1	1	2	3	4		1	5	9
2	5	6	7	8		2	6	10
3	9	10	11	12		3	7	11
4						4	8	12

Ricordiamo che il **determinante** di una matrice quadrata **C** è diverso da 0 se e solo se i vettori colonna che formano la matrice **C** sono linearmente indipendenti. Calcoliamo il determinante della matrice **C** appena ottenuta. In una cella qualsiasi, per esempio A5, scriviamo la formula

=MATR.DETERM(I1:K3)

(MDETERM per la versione inglese) e premiamo CTRL+SHIFT+INVIO.

	A	B	C	D	E	F	G	H	I	J	K
1	6	2	8		1	3	1		38	88	42
2	2	0	2		4	3	6		8	22	8
3	0	4	0		3	8	3		16	12	24
4											
5	32										

Risulta det(**C**) = 32.
Spesso ciò che importa del determinante di una matrice è che sia uguale a 0 oppure diverso da 0. Ricordiamo che Excel è un sistema di calcolo numerico, e quindi il calcolo del determinante è necessariamente approssimato, e difficilmente il risultato dell'algoritmo applicato (che è sostanzialmente l'algoritmo di Gauss) darà esattamente 0. Per esempio, il determinante della matrice

$$\mathbf{A} = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}$$

è 0 (la terza colonna è uguale al doppio della seconda meno la prima); Excel restituisce nella cella E1 un numero molto piccolo, circa $6 \cdot 10^{-16}$, ma non nullo. In questo caso non c'è dubbio che si tratti di un'approssimazione di 0: il determinante di una matrice di numeri interi è un numero intero.

	A	B	C	D	E
1	1	2	3		6,66134E-16
2	4	5	6		
3	7	8	9		

Ricordiamo che ammettono **matrice inversa** solo le matrici quadrate con determinante non nullo (si chiamano anche *singolari*). Calcoliamo la matrice inversa $\mathbf{C}^{-1}$ di $\mathbf{C}$. Selezioniamo un intervallo di celle di dimensione 3×3, per esempio I5:K7, e scriviamo la formula

=MATR.INVERSA(I1:K3)

(MINVERSE nella versione inglese) e premiamo CTRL+SHIFT+INVIO.

	A	B	C	D	E	F	G	H	I	J	K
1	6	2	8		1	3	1		38	88	42
2	2	0	2		4	3	6		8	22	8
3	0	4	0		3	8	3		16	12	24
4											
5									13,5	-50,25	-6,875
6									-2	7,5	1
7									-8	29,75	4,125

Excel restituisce la matrice inversa in forma numerica; come abbiamo visto prima, Excel non riconosce le matrici con determinante nullo. Ecco che cosa accade se tentiamo di invertire una matrice singolare.

	A	B	C	D	E	F	G
1	1	2	3		-4,5036E+15	9,0072E+15	-4,5036E+15
2	4	5	6		9,0072E+15	-1,80144E+16	9,0072E+15
3	7	8	9		-4,5036E+15	9,0072E+15	-4,5036E+15

Ovviamente il risultato è sbagliato: dividendo per il determinante, circa 10^{-16}, si ottengono elementi della matrice inversa circa pari a 10^{16}.

1.9 Sistemi lineari

Con le funzionalità appena viste è possibile con Excel risolvere un sistema lineare *di Cramer*, cioè un sistema di n equazioni in n incognite la cui matrice dei coefficienti è non singolare.
Per esempio, si debba risolvere il sistema

$$\begin{cases} x + 2y + 3z = 4 \\ x + y = 6 \\ y + z = 8 \end{cases}$$

che in forma matriciale, $\mathbf{Ax} = \mathbf{b}$, è il seguente:

$$\begin{pmatrix} 1 & 2 & 3 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 4 \\ 6 \\ 8 \end{pmatrix}.$$

Poiché $\det(\mathbf{A}) = 2 \neq 0$, si risolve il sistema semplicemente calcolando $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$.
Scriviamo la matrice **A** nelle celle A1:C3, e il vettore dei termini noti **b** nelle celle D1:D3. Selezioniamo le celle F1:F3, scriviamo il comando

=MATR.PRODOTTO(MATR.INVERSA(A1:C3);D1:D3)

e premiamo come al solito CTRL+SHIFT+INVIO.

	A	B	C	D	E	F
1	1	2	3	4	x =	-7
2	1	1	0	6	y =	13
3	0	1	1	8	z =	-5
4						
5	det(A) =		2			

Otteniamo così la soluzione: $x = -7$, $y = 13$, $z = -5$.

2. Modelli dinamici lineari e non lineari

Excel offre la possibilità di copiare una formula aggiornando i riferimenti relativi: si presta dunque in modo naturale a costruire successioni, di numeri o di vettori, definite in modo ricorsivo. Mentre i modelli lineari, cioè del tipo $\mathbf{x}_{t+1} = \mathbf{A}\mathbf{x}_t$ oppure lineari affini, del tipo $\mathbf{x}_{t+1} = \mathbf{A}\mathbf{x}_t + \mathbf{b}$, si possono trattare agevolmente dal punto di vista teorico, per i modelli non lineari le difficoltà possono essere insormontabili. Con uno strumento di calcolo numerico, invece, non c'è alcuna differenza sostanziale. Presentiamo di seguito due classici modelli, uno lineare e uno quadratico.

2.1 Catene di Markov

Si tratta di un classico modello lineare che, mediante una successione ricorsiva di vettori, descrive l'evoluzione di un sistema chiuso, all'interno del quale la popolazione può assumere diversi stati. Descriviamolo mediante un esempio.
Supponiamo che in un certo supermercato ciascuno dei 1000 clienti acquisti ogni settimana un detersivo, scegliendolo tra le marche 1 e 2. La matrice

$$\mathbf{A} = \begin{pmatrix} 0.8 & 0.3 \\ 0.2 & 0.7 \end{pmatrix}$$

descrive gli spostamenti relativi da marca a marca ogni settimana:

- $a_{11} = 0.8$: l'80% dei clienti che acquista la marca 1 in una certa settimana, conferma la scelta della marca 1 la settimana successiva (passa da 1 a 1);
- $a_{21} = 0.2$: il rimanente 20% dei clienti che acquista la marca 1 in una certa settimana, cambia marca e acquista la marca 2 (passa da 1 a 2);
- $a_{12} = 0.3$: il 30% dei clienti che acquista la marca 2 in una certa settimana, la settimana successiva cambia marca e acquista la marca 1 (passa da 2 a 1);
- $a_{22} = 0.7$: il rimanente 70% dei clienti che acquista la marca 2 in una certa settimana, conferma la scelta della marca 2 la settimana successiva (passa da 2 a 2).

In generale l'elemento a_{ij} della matrice fornisce la percentuale di clienti che, da una settimana all'altra, passa dalla marca j alla marca i.
Viene effettuata una rilevazione dalla quale risulta che in una certa settimana, che chiameremo settimana 0, 200 clienti hanno acquistato la marca 1 e 800 la marca 2. Quale sarà la distribuzione delle marche alla settimana 1, 2, …?
Nelle celle A1:B2 scriviamo la matrice **A**. Nella colonna C i valori del tempo, misurato in settimane. In D2 e E2 la distribuzione iniziale, $\mathbf{x}_0 = [200, 800]$. La distribuzione al tempo 1 sarà uguale al prodotto $\mathbf{A}\mathbf{x}_0$. In D3 scriviamo la formula

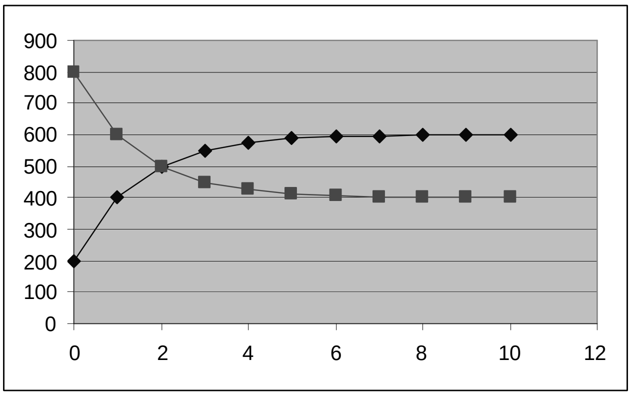
=$A$1*D2+$B$1*E2

e in E3 la formula

=$A$2*D2+$B$2*E2.

Copiamo le celle D3:E3 verso il basso.

	A	B	C	D	E
1	**0,8**	**0,3**	**t**	**x[t]**	**y[t]**
2	**0,2**	**0,7**	0	200	800
3			1	400	600
4			2	500	500
5			3	550	450
6			4	575	425
7			5	587,5	412,5
8			6	593,75	406,25
9			7	596,875	403,125
10			8	598,4375	401,5625
11			9	599,2188	400,7813
12			10	599,6094	400,3906



Si osserva, sia dalla tabella che dal grafico, che la distribuzione converge rapidamente al vettore [600, 400]. Si verifica anche che, modificando i valori iniziali nelle celle D2:E2 (ma mantenendo costante la somma 1000), tale vettore di equilibrio non dipende da $\mathbf{x}_0$.

2.2 Epidemia SIR

Questo è un classico modello non lineare (quadratico) che simula, in una popolazione chiusa di numerosità *N*, l'epidemia di una malattia infettiva che può portare alla morte (con probabilità *p*) oppure alla guarigione con immunità permanente (con probabilità $1-p$):

Suscettibili $\rightarrow$ Infetti $\rightarrow$ Rimossi.

L'unità di misura del tempo è il periodo che mediamente intercorre tra il contagio e la "rimozione" dall'insieme dei suscettibili (cioè la guarigione oppure la morte). Il vettore di stato
$[\mathrm{inf}_t, \mathrm{sus}_t, \mathrm{mor}_t, \mathrm{gua}_t]$
descrive, al tempo t, il numero di infetti, di suscettibili, di morti e di guariti. La legge del moto che presentiamo è la seguente:

$$\begin{pmatrix} \mathrm{inf}_{t+1} \\ \mathrm{sus}_{t+1} \\ \mathrm{mor}_{t+1} \\ \mathrm{gua}_{t+1} \end{pmatrix} = \begin{pmatrix} a \cdot \mathrm{inf}_t \cdot \mathrm{sus}_t \\ \mathrm{sus}_t - a \cdot \mathrm{inf}_t \cdot \mathrm{sus}_t \\ \mathrm{mor}_t + p \cdot \mathrm{inf}_t \\ \mathrm{gua}_t + (1-p) \cdot \mathrm{inf}_t \end{pmatrix}$$

dove a è un parametro che descrive la probabilità del contagio nell'incontro tra un suscettibile e un infetto. Poniamo $a = 0.0001$, $p = 0.3$, $N = 20000$; partiamo dalla situazione iniziale in cui ci sono 50 infetti:

$\mathbf{x}_0 = [50, 19950, 0, 0]$.

Nella colonna A scriviamo i tempi, da 0 a 15, per esempio. Nelle celle B2:E2 i valori iniziali e in F2 la loro somma. Nelle celle B3:E3 dobbiamo immettere la legge ricorsiva.

B3: =0.0001*B2*C2

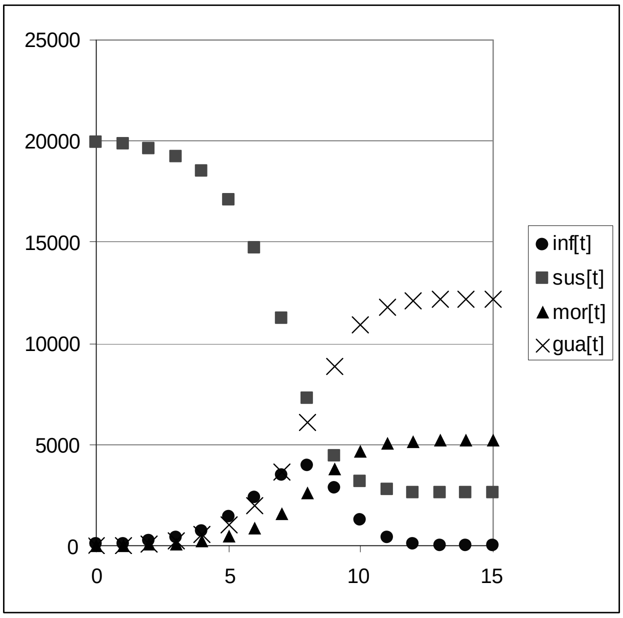
C3: =C2–B3

D3: =D2+0.3*B2

E3: =E2+0.7*B2

In F3 controlliamo che la somma rimanga costante.
Selezioniamo le celle B3:F3 e copiamo verso il basso.

	A	B	C	D	E	F
1	**t**	**inf[t]**	**sus[t]**	**mor[t]**	**gua[t]**	**tot[t]**
2	0	50	19950	0	0	20000
3	1	100	19850	15	35	20000
4	2	198	19652	45	105	20000
5	3	389	19263	104	243	20000
6	4	750	18514	221	516	20000
7	5	1388	17126	446	1041	20000
8	6	2377	14749	862	2012	20000
9	7	3505	11244	1575	3676	20000
10	8	3941	7303	2627	6129	20000
11	9	2878	4424	3809	8888	20000
12	10	1273	3151	4673	10903	20000
13	11	401	2750	5055	11794	20000
14	12	110	2639	5175	12075	20000
15	13	29	2610	5208	12152	20000
16	14	8	2603	5217	12173	20000
17	15	2	2601	5219	12178	20000
18	16	1	2600	5220	12180	20000
19	17	0	2600	5220	12180	20000



Al tempo $t = 17$ la situazione si è completamente stabilizzata: l'epidemia si è conclusa con un totale di 5220 morti e 12180 immunizzati: della popolazione iniziale solo 2600 non hanno contratto la malattia. Il sistema ammette come equilibrio il vettore $[0, N, 0, 0]$: in assenza di infetti non c'è contagio. Ma non è stabile: se la prima componente è positiva, inizia il contagio e l'evoluzione dell'epidemia segue sempre lo stesso corso qualitativo.

3. Risolvere equazioni

3.1 Due classici algoritmi per la risoluzione di equazioni

Excel possiede un potente strumento per risolvere equazioni (vedi il successivo paragrafo "Il solver di Excel"). Illustrare algoritmi per la risoluzione di equazioni ha il solo scopo di dare un'idea di quali strumenti possa utilizzare un software di calcolo numerico per questo obiettivo.
Mostreremo il *metodo decimale* e l'*algoritmo di Newton.*

Metodo decimale

Lavoriamo sull'equazione seguente:

$x^3+x = 100.$

La soluzione che cerchiamo, chiamiamola c, è certamente compresa tra 4 e 5: infatti, posto $f(x)=x^3+x$, risulta

$f(4)=64+4=68$
$f(5)=125+5=130$

e poiché f è continua, $c\in(4,5)$; il primo passo consiste nel determinare la prima cifra decimale di c: calcoleremo $f(x)$ da 4 a 5 con passo $\Delta x=0.1$.
In Excel impostiamo nella cella A2 il valore Δx, cioè inizialmente 0.1. Nella cella B2 scriviamo l'estremo sinistro dell'intervallo a cui appartiene c; in B3 scriviamo la formula

=B2+$A$2

e copiamo B2 verso il basso fino a B12, costruendo così la sequenza 4, 4.1, 4.2, …, 5.
In C2 scriviamo la formula

=B2^3+B2

e la copiamo verso il basso fino a C12, calcolando così i valori $f(4)$, $f(4.1)$, …, $f(5)$.

	A	B	C
1	**dx**	**x**	**f(x)**
2	0,1	4	68
3		4,1	73,021
4		4,2	78,288
5		4,3	83,807
6		4,4	89,584
7		4,5	95,625
8		4,6	101,936
9		4,7	108,523
10		4,8	115,392
11		4,9	122,549
12		5	130

Si osserva che $f(4.5)<100$ mentre $f(4.6)>100$. Quindi $c\in(4.5,4.6)$ e la prima cifra decimale è dunque 5. Proseguiamo ora nello stesso modo per la seconda cifra decimale: cambiamo il passo in A2 scrivendo 0.01, e in B2 scriviamo 4.5. Il foglio si aggiorna nel seguente modo.

	A	B	C
1	**dx**	**x**	**f(x)**
2	0,01	4,5	95,625
3		4,51	96,24385
4		4,52	96,86541
5		4,53	97,48968
6		4,54	98,11666
7		4,55	98,74637
8		4,56	99,37882
9		4,57	100,014
10		4,58	100,6519
11		4,59	101,2926
12		4,6	101,936

Con lo stesso ragionamento di prima scopriamo che c è compreso tra 4.56 e 4.57. In A2 scriviamo 0.001 e in B2 4.56, e proseguiamo così fino alla precisione desiderata, cambiando ad ogni iterazione i valori di A2 e B2.

	A	B	C
1	**dx**	**x**	**f(x)**
2	0,001	4,56	99,37882
3		4,561	99,44221
4		4,562	99,50563
5		4,563	99,56908
6		4,564	99,63256
7		4,565	99,69606
8		4,566	99,75959
9		4,567	99,82315
10		4,568	99,88674
11		4,569	99,95035
12		4,57	100,014

	A	B	C
1	**dx**	**x**	**f(x)**
2	0,0001	4,569	99,95035
3		4,5691	99,95671
4		4,5692	99,96308
5		4,5693	99,96944
6		4,5694	99,97581
7		4,5695	99,98217
8		4,5696	99,98853
9		4,5697	99,9949
10		4,5698	100,0013
11		4,5699	100,0076
12		4,57	100,014

	A	B	C
1	**dx**	**x**	**f(x)**
2	0,00001	4,5697	99,9949
3		4,56971	99,99553
4		4,56972	99,99617
5		4,56973	99,99681
6		4,56974	99,99744
7		4,56975	99,99808
8		4,56976	99,99872
9		4,56977	99,99935
10		4,56978	99,99999
11		4,56979	100,0006
12		4,5698	100,0013

	A	B	C
1	**dx**	**x**	**f(x)**
2	0,000001	4,56978	99,99999
3		4,569781	100,0001
4		4,569782	100,0001
5		4,569783	100,0002
6		4,569784	100,0002
7		4,569785	100,0003
8		4,569786	100,0004
9		4,569787	100,0004
10		4,569788	100,0005
11		4,569789	100,0006
12		4,56979	100,0006

Se ci fermiamo alla sesta cifra decimale otteniamo

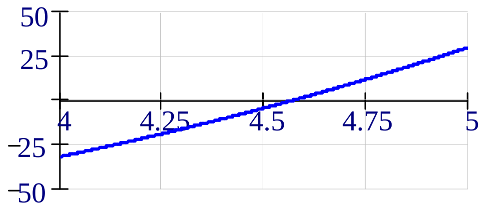
$c = 4.569780\ldots$

Algoritmo di Newton

L'algoritmo di Newton è uno dei più potenti metodi di approssimazione per le soluzioni di un'equazione. Lavoriamo sull'equazione

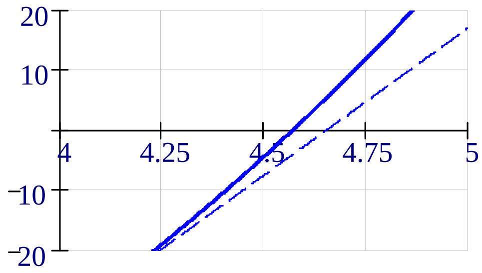
$$x^3+x-100=0$$

Posto $f(x) = x^3+x-100$, risulta $f(4)=-32$, $f(5)=30$. La soluzione c è dunque compresa tra 4 e 5, e si può interpretare come l'ascissa del punto in cui il grafico di $f(x)$ interseca l'asse x.



Prendiamo, come primo tentativo, $x_0 = 4$. In x_0 mandiamo la retta tangente a $f(x)$ ($f(x)$ deve dunque essere derivabile), la cui equazione è

$y = f(4)+f'(4)(x-4) = -228+49x.$



Tale retta interseca l'asse x in un punto $x_1 = 228/49 \approx 4.65$ che in generale è più vicino a c di x_0. Si prosegue nello stesso modo partendo da x_1, costruendo così una successione
$x_0, x_1, x_2, \ldots$
che in generale converge (con straordinaria rapidità) a c.
Fatti i conti in generale, la successione ricorsiva che fa passare da x_n a x_{n+1} è la seguente:

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$$

Nel nostro esempio, poiché $f'(x) = 3x^2+1$, la successione è così definita:

$$x_{n+1} = x_n - \frac{x_n^3 + x_n - 100}{3x_n^2 + 1}$$

Non ci resta che implementare in Excel questa successione. Nella colonna A costruiamo la successione dei numeri naturali 0, 1, 2, …, 20. In B2 scriviamo il valore iniziale $x_0 = 4$.
In B3 la formula
=B2–(B2^3+B2–100)/(3*B2^2+1)
che copiamo verso il basso fino alla riga 12. Ecco la tabella che otteniamo.

	A	B
1	**n**	**x[n]**
2	0	4
3	1	4,6530612244898
4	2	4,5712393790314
5	3	4,5697806213763
6	4	4,5697801629327
7	5	4,5697801629327
8	6	4,5697801629327
9	7	4,5697801629327
10	8	4,5697801629327
11	9	4,5697801629327
12	10	4,5697801629327

Come si vede in tabella, alla quarta iterazione sono già state approssimate le prime 13 cifre decimali di c.

3.2 Il "solver" di EXCEL

Excel possiede un ottimo "risolutore" (o "solver" nella versione inglese) di equazioni. Lo si trova in una voce del menù Strumenti. Se non c'è, occorre caricarlo in memoria (dopodiché rimane nel menù all'avvio di Excel) cliccando su Strumenti, Componenti aggiuntivi, Componente aggiuntivo Risolutore (in inglese Tools, Add-ins, Solver add-in). Vediamo come si utilizza.
Supponiamo di voler risolvere l'equazione (che non ammette una soluzione simbolica)

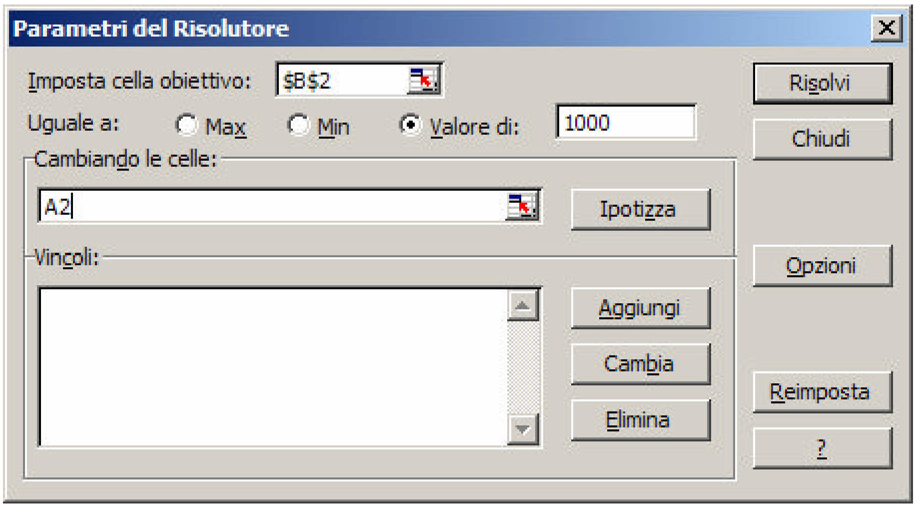
$$x^3+2^x = 1000.$$

In generale supponiamo di risolvere un'equazione del tipo $f(x)=k$ (qualunque equazione si può scrivere in tale forma, eventualmente con $k=0$).
Tutto ciò che dobbiamo fare è scrivere in una cella un valore iniziale per x, cioè un ipotetico e plausibile valore per la soluzione e in un'altra cella il corrispondente valore $f(x)$. Nel nostro esempio scriviamo 10 in A2 e calcoliamo $f(x)$ in B2 mediante la formula

=A2^3+2^A2

	A	B
1	x	f(x)
2	10	2024

Ovviamente 10 non è la soluzione dell'equazione, dato che $f(10)=2024$.
Chiamiamo ora il risolutore e impostiamo come cella obiettivo B2 al valore 1000, cambiando la cella A2.



Risolvere $x^3+2^x=1000$ è equivalente a calcolare la funzione inversa di f in 1000, cioè $f^{-1}(1000)$; questo è il modo con cui lavora il risolutore di Excel. Cliccando su "Risolvi" (e confermando con "Mantieni la soluzione del Risolutore") cambiano sia A2 (cioè x) sia B2 (cioè $f(x)$).

	A	B
1	x	f(x)
2	8.55035	1000

Il risolutore, partendo da 10, ha innescato un algoritmo (per i più curiosi si tratta di una variante dell'algoritmo di Newton) che conduce all'approssimazione della soluzione

$x=8.55035.$

Non vogliamo qui sollevare problemi spinosi, e cioè come cambia l'efficienza del risolutore quando si sceglie un valore iniziale x lontano dalla soluzione, oppure che cosa accade se l'equazione ammette più di una soluzione. Se la scelta del valore iniziale è ragionevole, e se $f(x)$ è una funzione sufficientemente regolare, l'algoritmo converge molto rapidamente ed è assai affidabile.

3.3 Il TIR di un'operazione finanziaria

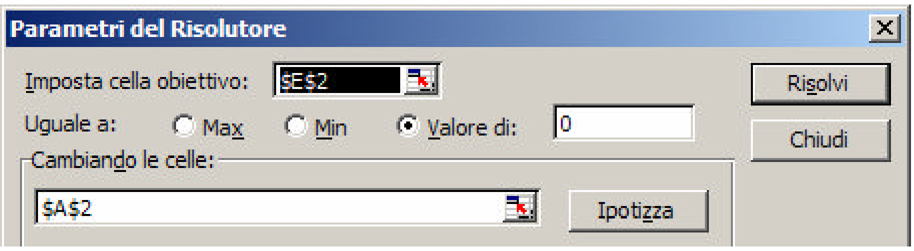
Abbiamo visto, in 1.6, la valutazione di un'operazione finanziaria mediante il VAN, che è un indice <u>soggettivo</u>: dipende dal *costo-opportunità dei mezzi propri* del valutatore, cioè dal tasso di investimento al quale il valutatore può accedere comunemente.
Un indice <u>oggettivo</u> di un'operazione finanziaria è invece il Tasso di Rendimento Interno (TIR): è il tasso implicito di rendimento dell'operazione. Il TIR è il tasso di valutazione per il quale l'operazione non risulta né conveniente né sconsigliabile; quindi il TIR è per definizione la soluzione dell'equazione $\mathrm{VAN}(x) = 0$; nel nostro esempio è la soluzione dell'equazione[2]

$$-1000 + \frac{200}{1+x} + \frac{200}{(1+x)^2} + \frac{200}{(1+x)^3} + \frac{300}{(1+x)^4} + \frac{400}{(1+x)^5} = 0.$$

È impensabile tentare di risolvere in modo simbolico una tale equazione (che, una volta semplificata, dà luogo ad un'equazione polinomiale di 5° grado, per la quale non esistono formule risolutive). L'unica via percorribile è quella di approssimare la soluzione. D'altra parte abbiamo visto che risulta $\mathrm{VAN}(0.03) > 0$ e $\mathrm{VAN}(0.065) < 0$: per la continuità della funzione $\mathrm{VAN}(x)$ la soluzione dell'equazione $\mathrm{VAN}(x) = 0$ è compresa tra 3% e 6.5%.
Riprendiamo il foglio che avevamo costruito per il calcolo del VAN.

	A	B	C	D	E
1	**tasso**	**t**	**a[t]**	**VAN[t]**	**VAN totale**
2	0.03	**0**	-1000	-1000	91.05
3		**1**	200	194.17	
4		**2**	200	188.52	
5		**3**	200	183.03	
6		**4**	300	266.55	
7		**5**	300	258.78	

Risolvere l'equazione $\mathrm{VAN}(x) = 0$ significa impostare la cella E2 al valore 0 cambiando la cella A2. Chiamiamo il risolutore.



Ecco il risultato.

	A	B	C	D	E
1	**tasso**	**t**	**a[t]**	**VAN[t]**	**VAN totale**
2	0.058783	**0**	-1000	-1000	0.00
3		**1**	200	188.90	
4		**2**	200	178.41	
5		**3**	200	168.50	
6		**4**	300	238.72	
7		**5**	300	225.47	

Il TIR dell'operazione finanziaria è dunque 5.878%.

[2] Si può dimostrare che un'operazione i cui flussi cambiano di segno una volta sola ammette un solo tasso interno finanziariamente accettabile, cioè maggiore di -1.

Se i flussi positivi sono tutti uguali, allora è possibile ottenere direttamente il TIR di un'operazione finanziaria con la funzione TASSO; se nell'esempio precedente i 5 flussi positivi fossero tutti uguali a 300, allora la formula
=TASSO(5;300;–1000)
fornirebbe il tasso interno dell'operazione. Ecco per esempio il confronto tra il Risolutore e la funzione TASSO (in A3).

	A	B	C	D	E
1	**tasso**	**t**	**a[t]**	**VAN[t]**	**VAN totale**
2	0.152382	**0**	-1000	-1000	0.00
3	15%	**1**	300	260.33	
4		**2**	300	225.91	
5		**3**	300	196.03	
6		**4**	300	170.11	
7		**5**	300	147.62	

Il formato cella della funzione TASSO è di default "Percentuale" con 0 cifre decimali; per modificarlo è sufficiente andare in Formato, Celle, e selezionare "Generale" oppure aumentare il numero di cifre decimali (a 3, per esempio).

	A	B	C	D	E
1	**tasso**	**t**	**a[t]**	**VAN[t]**	**VAN totale**
2	0.152382	**0**	-1000	-1000	0.00
3	15.238%	**1**	300	260.33	
4		**2**	300	225.91	
5		**3**	300	196.03	
6		**4**	300	170.11	
7		**5**	300	147.62	

3.4 Il TAN di un finanziamento

Trattiamo un caso reale, tratto dalla pubblicità di un quotidiano di questi giorni: un'agenzia di prestiti a medio termine offre un finanziamento di 2000 euro da ripagarsi con 60 rate mensili di 39.46 €. Qual è il tasso di finanziamento effettivamente applicato, il cosiddetto Tasso Annuo Nominale? Se la mia banca mi fa credito ad un tasso di finanziamento del 6%, che cosa mi conviene scegliere?
L'operazione finanziaria in questione è descritta dalla seguente tabella:

```
tempi   0     1/12     2/12   ...    60/12
flussi 2000   -39.46  -39.46  ...   -39.46
```

e il TAN è (in perfetta analogia con il TIR) la soluzione dell'equazione VAN(x) = 0:

$$2000-\sum_{t=1}^{60}\frac{39.46}{\left(1+x\right)^{t/12}}=0.$$

La risoluzione di tale equazione è del tutto simile a quella svolta per il calcolo del TIR. Nella colonna B inseriamo i tempi (in mesi) 0, 1, 2, …, 60; nella colonna C scriviamo i flussi: 2000 in C2, –39.46 da C3 fino a C62 (magari rendendo la rata un parametro del problema in A4, e scrivendo =$A$4 nelle celle da C3 a C62). In D2 la formula

=C2*(1+$A$4)^(B2/12)

copiata verso il basso fino a D62, fornisce i valori attuali dei flussi. In E2 scriviamo la formula per il VAN dell'intero finanziamento:

=SOMMA(D2:D62)

Ipotizziamo un valore iniziale del TAN del 3%. La figura seguente mostra solo le prime righe della tabella. Il VAN, valutato al tasso del 3%, è negativo, dunque il tasso interno sarà maggiore del 3% (ricordiamo che il VAN di un finanziamento è una funzione crescente del tasso di valutazione).

	A	B	C	D	E
1	**TAN**	**t**	**a[t]**	**VAN[t]**	**VAN totale**
2	0.03	0	2000	2000	-198.24
3	**RATA**	1	-39.46	-39.36	
4	-39.46	2	-39.46	-39.27	
5		3	-39.46	-39.17	
6		4	-39.46	-39.07	
7		5	-39.46	-38.98	
8		6	-39.46	-38.88	

Chiamiamo il risolutore, e impostiamo la cella E2 al valore 0 cambiando la cella A2.

	A	B	C	D	E
1	**TAN**	**t**	**a[t]**	**VAN[t]**	**VAN totale**
2	0.070681	0	2000	2000	0.00
3	**RATA**	1	-39.46	-39.24	
4	-39.46	2	-39.46	-39.01	
5		3	-39.46	-38.79	
6		4	-39.46	-38.57	
7		5	-39.46	-38.35	
8		6	-39.46	-38.14	

Il TAN del finanziamento è dunque del 7.068%: mi conviene rivolgermi alla banca, il cui tasso di finanziamento del 6% è minore del tasso interno del finanziamento proposto dall'agenzia.
Per questa operazione, in cui i flussi negativi sono tutti uguali, è possibile utilizzare direttamente la funzione TASSO. La formula

=TASSO(60;−39.46;2000)

immessa in una cella qualsiasi fornisce il tasso implicito i_{12} (è un tasso mensile, ovviamente) del finanziamento.

A1 | fx =TASSO(60;-39.46;2000)

	A	B	C	D	E
1	0.00570744				
2					

Per ottenere il tasso equivalente annuo i è sufficiente ricordare la relazione tra tassi in capitalizzazione composta:

$$\left(1+i_{12}\right)^{12} = 1+i,$$

da cui si ottiene il tasso annuo $i = 7.068\%$.

A2 | fx =(1+A1)^12-1

	A	B	C	D	E
1	0.00570744				
2	0.070680651				

3.5 Costituzione di capitale

Vogliamo disporre, tra 10 anni, di un capitale di 20 000 euro. Per costituire tale capitale investiamo oggi 5000 euro al tasso composto del 4% annuo, e versiamo ogni anno, a partire dall'anno prossimo, una quota C di integrazione del capitale; per tenere il passo con l'inflazione tale integrazione viene rivalutata ogni anno del 2.5%. Ci chiediamo quanto deve valere C affinché tra 10 anni si raggiunga il montante desiderato. Impostiamo la seguente tabella.

	A	B	C	D	E
1	**saldo iniziale**	5000	**t**	**integrazione**	**montante**
2	**tasso attivo**	0.04	0	0	5000
3	**tasso inflazione**	0.025	1	800	6000
4	**integrazione**	800	2	820.00	7060.00
5			3	840.50	8182.90
6			4	861.51	9371.73
7			5	883.05	10629.65
8			6	905.13	11959.96
9			7	927.75	13366.11
10			8	950.95	14851.71
11			9	974.72	16420.50
12			10	999.09	18076.41

Nelle celle B1:B4 ci sono i parametri del problema, che possono essere modificati in modo che il foglio si aggiorni. In B4 abbiamo inserito 800, un valore plausibile per l'integrazione iniziale.
La colonna D è costruita nel seguente modo: 0 in D2 (la primaintegrazione è al tempo 1), in D3 la formula

=$B$4,

in D4 la formula (che si copia fino a D12):

=D3*(1+$B$3)

che rivaluta del 2.5% l'integrazione dell'anno precedente.
La colonna E è costruita nel seguente modo: in E2 la formula

=$B$1,

in E3 la formula (che si copia fino a E12):

=E2*(1+$B$2)+D3

che somma al montante l'integrazione.
Il montante al termine dei 10 anni (18076.41) è evidenziato nella cella E12. Come si vede, l'integrazione iniziale di 800 € è troppo bassa.
Chiamiamo ora il risolutore, impostando a 20000 la cella E12, cambiando la cella B4. Il risultato è il seguente.

	A	B	C	D	E
1	**saldo iniziale**	5000	**t**	**integrazione**	**montante**
2	**tasso attivo**	0.04	0	0	5000
3	**tasso inflazione**	0.025	1	944.15	6144.15429
4	**integrazione**	944.15	2	967.76	7357.68
5			3	991.95	8643.94
6			4	1016.75	10006.45
7			5	1042.17	11448.87
8			6	1068.22	12975.05
9			7	1094.93	14588.98
10			8	1122.30	16294.85
11			9	1150.36	18097.00
12			10	1179.12	20000.00

L'integrazione iniziale che costruisce, ai tassi stabiliti, un capitale di 20000 € in 10 anni è dunque di 944.15 €.
Si osservi che l'equazione da risolvere è assai complessa, prima ancora che da risolvere, persino da scrivere in forma simbolica; detta x l'integrazione iniziale e a_t il montante al tempo t, risulta

$a_0 = 5000$
$a_1 = 1.04 \cdot 5000 + x$

$a_2 = 1.04a_1 + 1.025x$
$a_3 = 1.04a_2 + 1.025^2x$
…
$a_{10} = 1.04a_9 + 1.025^9x$

l'equazione da risolvere è $a_{10} = 20000$, cioè
$1.04a_9 + 1.025^9x = 20000$
e la sua esplicitazione in funzione della sola incognita x è quantomeno problematica.

3.6 Ottimizzazione vincolata

Il risolutore di Excel può essere utilizzato anche imponendo uno o più vincoli alle variabili e quindi si presta a risolvere problemi di ottimizzazione vincolata di funzioni in più variabili, cioè problemi che si presentano (per esempio con due variabili) nella forma

max $f(x,y)$
sub $g(x,y) = b$

f è la funzione da ottimizzare, g è la funzione di vincolo: vogliamo determinare, tra tutti i punti (x,y) che soddisfano il vincolo, cioè soddisfano l'equazione $g(x,y) = b$, quello che rende massimo il valore $f(x,y)$.
Per esempio, vogliamo risolvere il problema

max $5x+2y$
sub $x^2+y^2 = 20$

In A2 e in B2 inseriamo due valori ragionevoli per x e y, per esempio $x = 4$ e $y = 2$. In C2 calcoliamo il valore $f(x,y)$ mediante la formula

=5*A2+2*B2

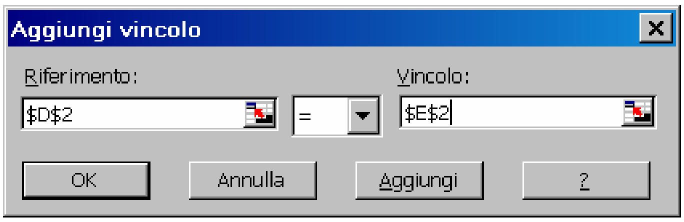
In D2 calcoliamo $g(x,y)$, mediante la formula

=A2^2+B2^2

In E2 scriviamo il valore della costante di vincolo b, cioè 20.

	A	B	C	D	E
1	**x**	**y**	**f(x,y)**	**g(x,y)**	**b**
2	4	3	26	25	20

Ora chiamiamo il risolutore: impostiamo la cella C2 al valore massimo cambiando le celle A2:B2 e aggiungiamo il vincolo, cliccando su "Aggiungi" e imponendo che sia D2=E2.



Cliccando su Risolvi la tabella cambia nel seguente modo.

	A	B	C	D	E
1	**x**	**y**	**f(x,y)**	**g(x,y)**	**b**
2	4,152201	1,661093	24,08319	20	20

Come si vede i valori delle celle D2 e E2 sono uguali, garantendo che la soluzione trovata rispetti effettivamente il vincolo. La soluzione è $x \approx 4.15$ e $y \approx 1.66$ e con tali valori la funzione f assume circa il valore 24.08: non è possibile far di meglio per un punto che soddisfi il vincolo.

4. Piani di ammortamento

Vogliamo costruire un foglio di lavoro che costruisca un piano di ammortamento a partire dai seguenti dati in ingresso:

- somma finanziata D_0
- tasso annuo di finanziamento i
- numero di rate annuali n

Modificando uno dei parametri deve aggiornarsi l'intero foglio.
Distinguiamo tra:

- ammortamenti a quota capitale C costante (dunque $C = D_0/n$), in cui il debito residuo D_t decresce linearmente: $D_{t+1} = D_t - C$;
- ammortamenti a rata costante R; in questo caso il debito residuo decresce in modo concavo; la legge ricorsiva è $D_{t+1} = D_t(1+i) - R$.

4.1 Ammortamenti a quota capitale costante

Supponiamo che la somma finanziata sia $D_0 = 20\,000$ €, $i = 8\%$, $n = 5$. Inseriamo questi parametri rispettivamente nelle celle B2, B3, B4. Nelle celle C3:C8 inseriamo i valori del tempo $t = 0, 1, \ldots, 5$.
Nella cella D3 scriviamo la formula
=$B$2
e scriviamo 0 nelle celle E2, F2, G2.

	A	B	C	D	E	F	G
1	**Ammortamento a quota capitale costante**						
2	**D[0] =**	20000	**t**	**D[t]**	**C[t]**	**I[t]**	**R[t]**
3	**i =**	0.08	0	20000	0	0	0
4	**n =**	5	1	16000	4000	1600	5600
5			2	12000	4000	1280	5280
6			3	8000	4000	960	4960
7			4	4000	4000	640	4640
8			5	0	4000	320	4320

Ora veniamo al nocciolo: compiliamo le celle D4:G4, che poi copieremo verso il basso.
Innanzitutto in E4 scriviamo la formula

=$B$2/$B$4

che calcola la quota capitale costante C.
In F4 calcoliamo gli interessi di periodo mediante la formula

=D3*$B$3

In G4 sommiamo quota capitale e quota interessi per ottenere la rata R

=E4+F4

e finalmente in D4 aggiorniamo il debito residuo:

=D3–E4.

Se ora selezioniamo le celle D4:G4 e copiamo verso il basso fino alla riga D8:G8, l'intero piano di ammortamento è pronto.
Un controllo è costituito dal fatto che l'ultimo debito residuo, cioè D_5 (la cella D8), sia uguale a 0.

4.2 Ammortamenti a rata costante

Con gli stessi parametri precedenti costruiamo ora un ammortamento a rata costante. Innanzitutto occorre ricavare, dai parametri D_0, i, n il valore della rata costante R.
Dall'equivalenza finanziaria

$$D_0 = \sum_{t=1}^{n} \frac{R}{(1+i)^t}$$

si ricava

$$R = \frac{D_0}{\sum_{t=1}^{n} \frac{1}{(1+i)^t}} = D_0 \frac{i}{1-(1+i)^{-n}}$$

Peraltro Excel fornisce una funzione =RATA che calcola in automatico R.
Prendiamo lo stesso schema del foglio precedente: il nocciolo è compilare le celle D4:G4.
Innanzitutto in G4 scriviamo la formula

=–RATA($B$3;$B$4;$B$2)

(Excel fornisce di default un valore negativo per la rata). In F4 calcoliamo gli interessi:

=D3*$B$3

In E4 calcoliamo per differenza la quota capitale

=G4–F4

e finalmente aggiorniamo il debito residuo in D4:

=D3–E4.

Selezioniamo le celle D4:G4 e copiamo verso il basso, finoa D8:G8.

	A	B	C	D	E	F	G
1	**Ammortamento a rata costante**						
2	**D[0] =**	20000	**t**	**D[t]**	**C[t]**	**I[t]**	**R[t]**
3	**i =**	0.08	0	20000	0	0	0
4	**n =**	5	1	16590.87	3409.13	1600	5 009.13
5			2	12909.01	3681.86	1327.27	5 009.13
6			3	8932.60	3976.41	1032.721	5 009.13
7			4	4638.08	4294.52	714.6083	5 009.13
8			5	0.00	4638.08	371.0466	5 009.13

Anche in questo caso possiamo controllare che l'ultimo debito residuo sia nullo, e verificare che il debito sia effettivamente estinto.

4.3 Parametrizzare il numero di righe di una tabella

Nei piani di ammortamento che abbiamo costruito nei paragrafi precedenti, modificando D_0 (la cella B2) oppure i (la cella B3) si aggiorna istantaneamente il piano di ammortamento con i nuovi parametri. Ma questo non accade, ovviamente, con il parametro n (la cella B4), perché in questo caso non solo si devono aggiornare le formule, ma si devono anche aggiungere nuove righe.
Per ottenere lo scopo di vedere apparire tante righe quante occorrono, semplicemente cambiando il contenuto di una cella utilizziamo la struttura di controllo SE di Excel.
Partiamo dallo schema appena utilizzato per l'ammortamento a rata costante. Riscriviamo la colonna C dei tempi nel seguente modo: scriviamo 0 in C3 e in C4 scriviamo la formula

=SE(C3<$B$4;C3+1;"")

e copiamola verso il basso fino ad arrivare a quello che riteniamo essere il limite superiore per i tempi, per esempio 40 anni, dunque fino alla cella C43.
Il significato è il seguente: fintantoché la cella precedente è minore di *n* allora la cella è incrementata di 1, altrimenti viene inserita la stringa vuota "", cioè la cella rimane vuota.
Si provi a modificare il numero di anni nella cella B4 e si vedrà in ogni caso riempire la colonna C da 0 fino al valore immesso in B4.
Ora nelle celle D4:G4 scriviamo le stesse formule di prima, ma le facciamo precedere dal controllo che il valore della colonna C (il tempo corrispondente) sulla stessa riga sia minore o uguale al numero di rate annue in B4. In caso contrario, immettiamo la riga vuota. Ecco in dettaglio la sintassi:

- in G4: =SE(C4<=$B$4;-rata($B$3;$B$4;$B$2);"")
- in F4: =SE(C4<=$B$4;D3*$B$3;"")
- in E4: =SE(C4<=$B$4;G4–F4;"")
- in D4: =SE(C4<=$B$4;D3–E4;"")

Come al solito selezioniamo ora le celle D4:G4 e copiamo verso il basso fino a D43:G43.
Se in B4 abbiamo lasciato 5, si riempiono solo le celle fino a $t = 5$; modificando il valore della cella B4 si aggiorna il piano, con il numero di righe opportune.
Ecco per esempio che cosa appare se si cambia la cella B4 in 10.

	A	B	C	D	E	F	G
1	**Ammortamento a rata costante**						
2	**D[0] =**	20000	**t**	**D[t]**	**C[t]**	**I[t]**	**R[t]**
3	**i =**	0.08	0	20000	0	0	0
4	**n =**	10	1	18619.41	1380.59	1600	2 980.59
5			2	17128.37	1491.04	1489.553	2 980.59
6			3	15518.05	1610.32	1370.27	2 980.59
7			4	13778.91	1739.15	1241.444	2 980.59
8			5	11900.63	1878.28	1102.313	2 980.59
9			6	9872.09	2028.54	952.0505	2 980.59
10			7	7681.27	2190.82	789.7673	2 980.59
11			8	5315.18	2366.09	614.5015	2 980.59
12			9	2759.81	2555.38	425.2145	2 980.59
13			10	0.00	2759.81	220.7844	2 980.59

Ora tutti i dati in ingresso sono parametrizzati.

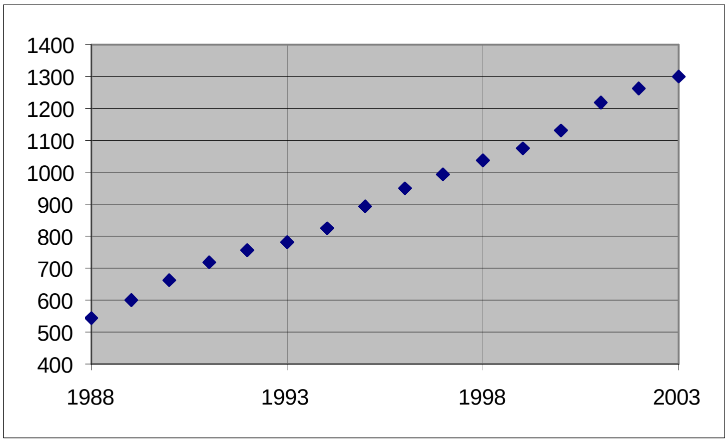
5. Regressione e fit di dati

Nota una tabella di dati relativi alle osservazioni di due grandezze X e Y, è naturale formulare ipotesi su quale possa essere una ragionevole funzione che rappresenti o che approssimi la relazione tra X e Y.
Il metodo dei *minimi quadrati* (detto anche di *regressione*) è una risposta largamente condivisa a tale problema.
Tratteremo i casi delle funzioni più semplici ma anche più interessanti e più utilizzati nella pratica scientifica:

- i modelli lineari: $f(x) = ax+b$
- i modelli potenza: $f(x) = ax^b$
- i modelli esponenziali: $f(x) = ab^x$

5.1 Modelli lineari: $f(x)=ax+b$

Il grafico seguente riporta i dati relativi all'andamento del PIL (Prodotto Interno Lordo, in miliardi di Euro) in Italia dal 1988 al 2003 (fonte ISTAT, www.istat.it).



Il grafico mostra un andamento sostanzialmente lineare. Sorge naturale la domanda: qual è la "miglior" funzione lineare che approssima i dati?
Come è noto la risposta teorica è la seguente: date le osservazioni

$X = x_1, x_2, \ldots, x_n$
$Y = y_1, y_2, \ldots, y_n$

la miglior funzione lineare $f(x) = ax+b$ che si descrive il comportamento di Y rispetto a X è quella i cui parametri a e b rendono minima la funzione

$$S(a,b) := \sum_{i=1}^{n} \left(ax_i + b - y_i \right)^2$$

che rappresenta la somma dei quadrati degli scarti dell'ordinata *teorica*

ax_i+b

da quella osservata

y_i.

Possiamo ottenere i valori di a e b in più modi.

I comandi PENDENZA e INTERCETTA

Supponiamo che i valori di X e di Y siano elencati rispettivamente in A2:A17 e in B2:B17.
In due celle qualsiasi scriviamo le formule

=PENDENZA(B2:B17;A2:A17)

=INTERCETTA(B2:B17;A2:A17)

(in inglese le funzioni sono rispettivamente SLOPE e INTERCEPT). Si osservi che per entrambe le funzioni il primo argomento è l'intervallo delle celle relative alla grandezza Y, il secondo argomento è l'intervallo delle celle relative alla grandezza X.

	A	B	C	D
1	**Anni**	**PIL (10^9 €)**	**pendenza**	**intercetta**
2	1988	546	49.682	-98 219
3	1989	598		
4	1990	660		
5	1991	720		
6	1992	759		
7	1993	782		
8	1994	827		
9	1995	894		
10	1996	951		
11	1997	994		
12	1998	1039		
13	1999	1072		
14	2000	1129		
15	2001	1218		
16	2002	1260		
17	2003	1301		

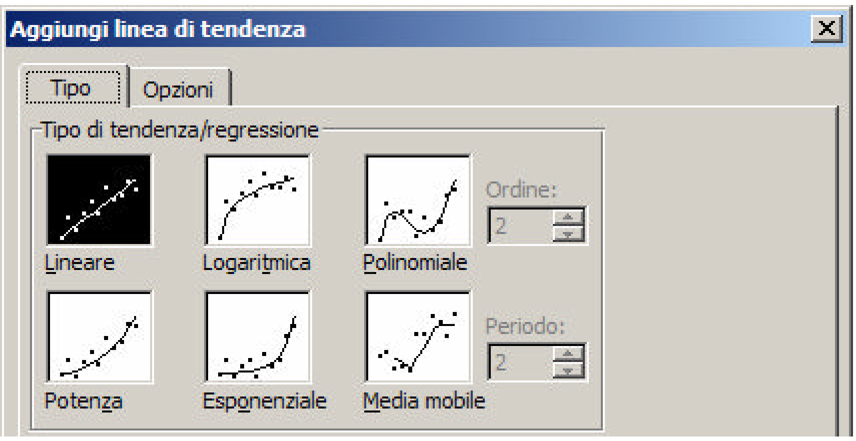
Dunque la funzione lineare che meglio si adatta ai dati (la cosiddetta *retta di regressione*) è

$f(x) = 49.682x - 98219.$

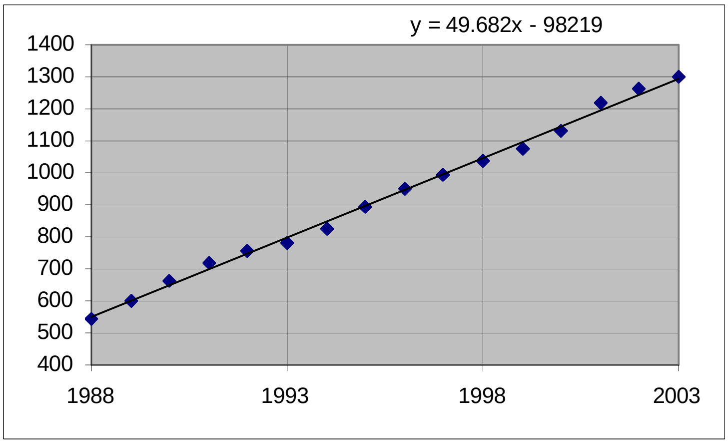
Il significato della pendenza è chiaro: in media, ogni anno il PIL è aumentato di circa 50 miliardi di euro.

Il menù "Aggiungi linea di tendenza"

Un metodo alternativo, che ci mostra anche il grafico della retta di regressione, è il seguente. Dopo aver costruito il grafico di Y rispetto a X, selezioniamolo: nella barra dei menù compare la voce "Grafico". Clicchiamo su Aggiungi linea di tendenza.



Nella scheda Tipo scegliamo "Lineare" e nella scheda "Opzioni" selezioniamo "Visualizza l'equazione sul grafico". Il risultato è il seguente: in alto leggiamo l'equazione della retta di regressione.



Utilizzare il risolutore

Un metodo più lento (ma un po' più trasparente) per ottenere i valori di a e b consiste nell'utilizzare il risolutore di Excel per minimizzare la funzione $S(a, b)$.
Innanzitutto operiamo una traslazione dei tempi, per semplificare i calcoli: anziché da 1988 a 2003, utilizziamo i valori da 0 a 15. In questo modo, anziché la retta $y=ax+b$, cerchiamo la retta $y=a(x-1988)+b$.
Scriviamo nelle celle D2 e E2 due possibili valori per a e b, per esempio 20 e 500.
Nella cella F2 scriviamo il quadrato dello scarto relativo al primo punto:

=($D$2*B2+$E$2-C2)^2

e copiamo verso il basso fino a F17. Ora in F18 sommiamo tutti i dati della colonna F.

	A	B	C	D	E	F
1	**Anni**	**Anni-1988**	**PIL (10^9 €)**	**a**	**b**	**S(a,b)**
2	1988	0	546	20	500	2116
3	1989	1	598			6084
4	1990	2	660			14400
5	1991	3	720			25600
6	1992	4	759			32041
7	1993	5	782			33124
8	1994	6	827			42849
9	1995	7	894			64516
10	1996	8	951			84681
11	1997	9	994			98596
12	1998	10	1039			114921
13	1999	11	1072			123904
14	2000	12	1129			151321
15	2001	13	1218			209764
16	2002	14	1260			230400
17	2003	15	1301			251001
18						1485318

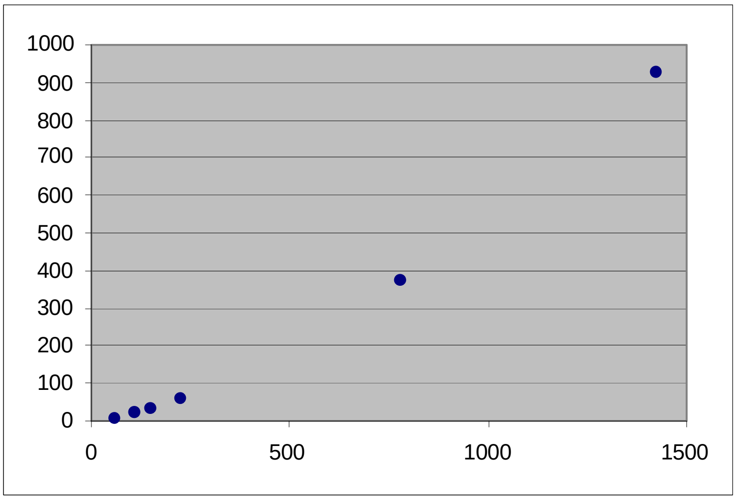
Ora chiamiamo il risolutore e impostiamo la cella F18 al valore minimo cambiando le celle D2:E2.
Risulta 49.68 in D2 (cioè a) e 549.26 in E2 (cioè b). Abbiamo così trovato a e b per l'equazione $y=a(x-1988)+b = ax-1988a+b$, la cui intercetta è proprio $-1988a+b = -98219$, cioè l'intercetta già trovata con gli altri metodi.

5.2 Regressione potenza

La tabella seguente mostra le distanze medie (in milioni di chilometri) e il tempo di rivoluzione intorno al Sole (in milioni di secondi, circa 11.5 giorni) di alcuni pianeti del Sistema Solare: quelli visibili a occhio nudo.

	A	B	C
1	**Pianeta**	**d (10^6 km)**	**t (10^6 s)**
2	Mercurio	58.2	7.7
3	Venere	108.3	19.4
4	Terra	149.6	31.5
5	Marte	227.8	59.2
6	Giove	777.9	373.6
7	Saturno	1422.7	928.0

Il grafico mostra i punti disposti su una curva regolare, crescente e convessa, che presumibilmente passa per l'origine.



Tutto lascia pensare ad una funzione potenza

$y = ax^b$,

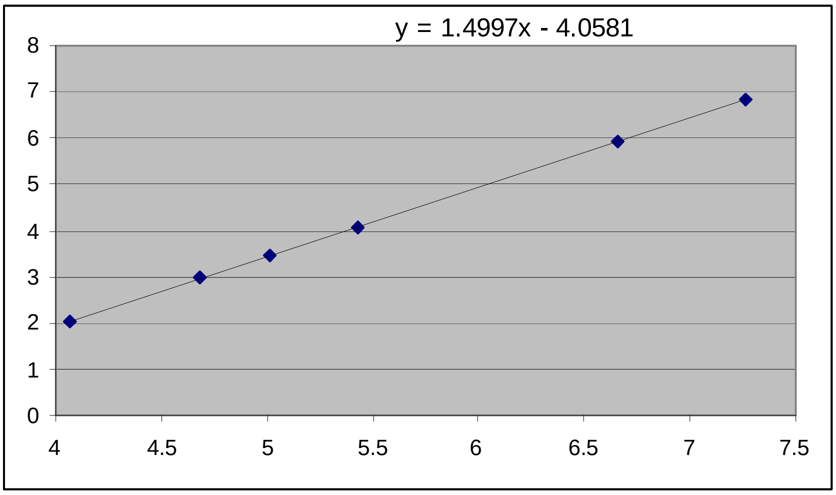
con $b > 1$. Se così fosse, passando ai logaritmi risulterebbe

$\ln(y) = \ln(a) + b\ln(x)$

e cioè, ponendo $Y = \ln(y)$ e $X = \ln(x)$, dovremmo ottenere una relazione lineare tra $\ln(y)$ e $\ln(x)$:

$Y = \ln(a) + bX$,

di pendenza b e intercetta $\ln(a)$.
Possiamo allora cercare la retta di regressione tra X e Y.



La retta di regressione tra Y e X ha circa equazione

$Y = 1.5X - 4$,

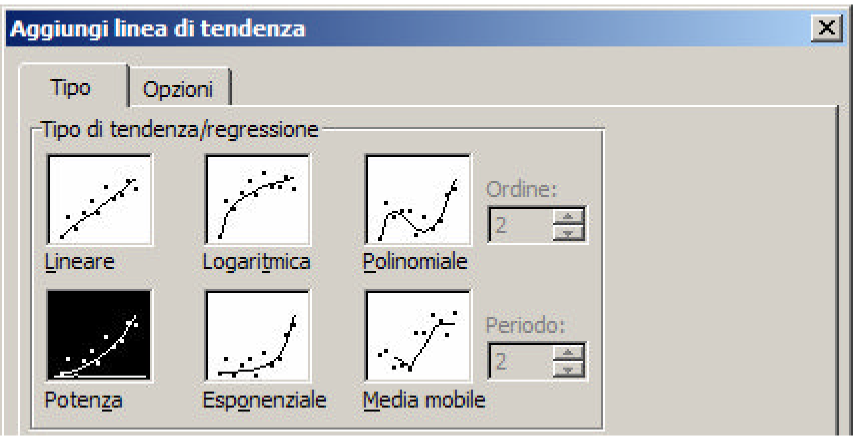
da cui

$b = 1.5$
$a = e^{-4} \approx 0.017$.

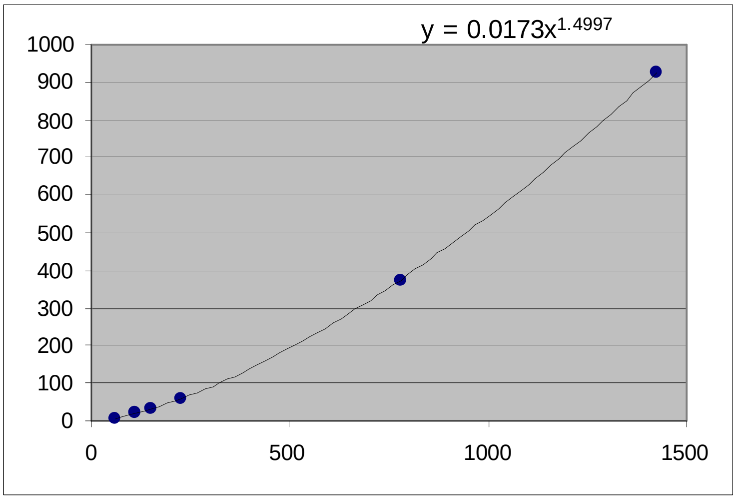
Dunque la funzione potenza che meglio approssima i dati iniziali è circa

$f(x) = 0.017 \cdot x^{1.5}$.

(si tratta della III legge di Keplero: il tempo di rivoluzione $f(x)$ di un pianeta intorno al Sole è una funzione potenza della sua distanza media x dal Sole, con esponente 1.5).
Allo stesso risultato saremmo arrivati chiedendo subito la regressione potenza con il comando "Aggiungi linea di tendenza" di Excel.



Ecco il grafico della funzione potenza che approssima i dati.



Vale la pena osservare che la retta di regressione di Y rispetto a X non fornisce la "migliore" funzione potenza di y rispetto a x, nel senso che la coppia (a, b) che minimizza la funzione

$$T(a,b) := \sum_{k=1}^{n} \left(\ln(a) + bX_k - Y_k\right)^2 .$$

(è questo il problema che abbiamo risolto, con $X_k = \ln(x_k)$ e $Y_k = \ln(y_k)$) non è uguale alla coppia (a, b) che minimizza la funzione:

$$S(a,b) := \sum_{k=1}^{n} \left(ax_k^b - y_k\right)^2 .$$

Tuttavia in generale ne è una buona approssimazione, e per questo motivo viene implementata nei sistemi di calcolo; il fatto è che minimizzare $T(a, b)$ è facile. Infatti, posto $A = \ln(a)$, $B = b$, $X = \ln(x)$, $Y = \ln(y)$ possiamo scrivere

$$T(A,B) := \sum_{k=1}^{n} \left(A + BX_k - Y_k\right)^2 .$$

Così T è quadratica nelle variabili A e B, e il sistema delle derivate parziali uguagliate a 0

$$\begin{cases} T_A'(A,B) = 0 \\ T_B'(A,B) = 0 \end{cases}$$

è un sistema lineare (e dunque risolubile in forma simbolica) nelle variabili A e B. Una volta calcolati A e B si risale facilmente ad a e b.
Invece minimizzare $S(a, b)$ è tutt'altro che facile e comunque non è possibile in forma simbolica. Utilizziamo ancora il risolutore, partendo dai dati ottenuti: $a = 0.017$ in D2, $b = 1.5$ in E2. In F2 scriviamo la formula

=($D$2*B2^$E$2-C2)^2

che calcola il quadrato dello scarto per il primo punto, copiamo fino a F7 e sommiamo in F8. Otteniamo che la somma dei quadrati degli scarti (cioè il numero da minimizzare) è 271.17.

	A	B	C	D	E	F
1	**Pianeta**	**d (10^6 km)**	**t (10^6 s)**	**a**	**b**	**(a*d^b-t)^2**
2	Mercurio	58.2	7.7	0.017	1.5	0.02
3	Venere	108.3	19.4			0.06
4	Terra	149.6	31.5			0.16
5	Marte	227.8	59.2			0.54
6	Giove	777.9	373.6			22.55
7	Saturno	1422.7	928.0			247.85
8						271.17

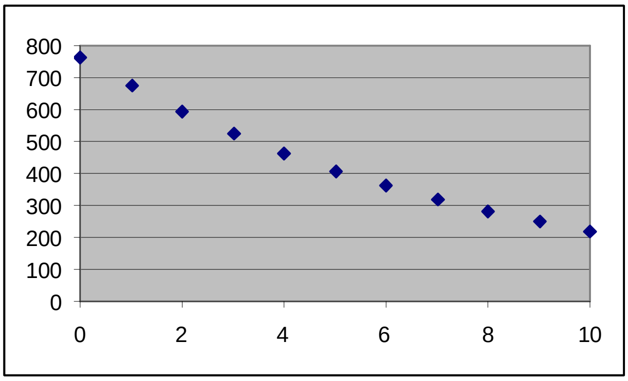
Ora chiamiamo il risolutore e minimizziamo F8 cambiando le celle D2:F2.
Otteniamo a = 0.0166 e $b = 1.5056$, e la somma dei quadrati degli scarti scende a 0.35: i nuovi valori di a e b sono cambiati di poco, ma il valore della somma degli scarti è migliorato di parecchio.

	A	B	C	D	E	F
1	**Pianeta**	**d (10^6 km)**	**t (10^6 s)**	**a**	**b**	**(a*d^b-t)^2**
2	Mercurio	58.2	7.7	0.016606	1.5055731	0.03
3	Venere	108.3	19.4			0.03
4	Terra	149.6	31.5			0.06
5	Marte	227.8	59.2			0.11
6	Giove	777.9	373.6			0.10
7	Saturno	1422.7	928.0			0.01
8						0.35

5.3 Regressione esponenziale

I dati seguenti riportano la pressione atmosferica media (in millimetri di mercurio) in funzione dell'altezza sul livello del mare (in km).

	A	B
1	**h (km)**	**p (mmHg)**
2	0	760
3	1	674
4	2	596
5	3	526
6	4	462
7	5	405
8	6	360
9	7	318
10	8	281
11	9	248
12	10	219



Il grafico e le informazioni sulla grandezza fisica in esame lasciano pensare ad un andamento esponenziale decrescente, cioè del tipo

$y = ab^x$

con $0 < b < 1$. La funzione esponenziale di regressione è, per definizione, quella che minimizza la funzione in a e b:

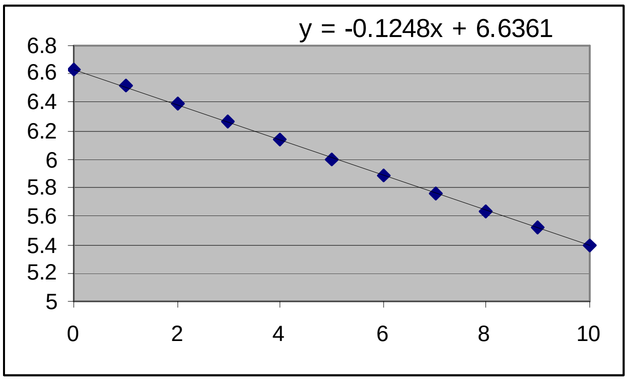
$$S(a,b) := \sum_{i=1}^{n} \left(ab^{x_i} - y_i \right)^2$$

Come prima, passiamo ai logaritmi. Se y fosse una funzione esponenziale di x, risulterebbe

$\ln(y) = \ln(a) + x\ln(b)$

cioè $\ln(y)$ sarebbe espresso da una funzione lineare di x, di pendenza $\ln(b)$ e intercetta $\ln(a)$. Proviamo allora a tracciare il grafico del logaritmo della pressione in funzione dell'altezza.

	A	B	C
1	h (km)	p (mmHg)	ln(p)
2	0	760	6.633318
3	1	674	6.51323
4	2	596	6.390241
5	3	526	6.265301
6	4	462	6.135565
7	5	405	6.003887
8	6	360	5.886104
9	7	318	5.762051
10	8	281	5.638355
11	9	248	5.513429
12	10	219	5.389072



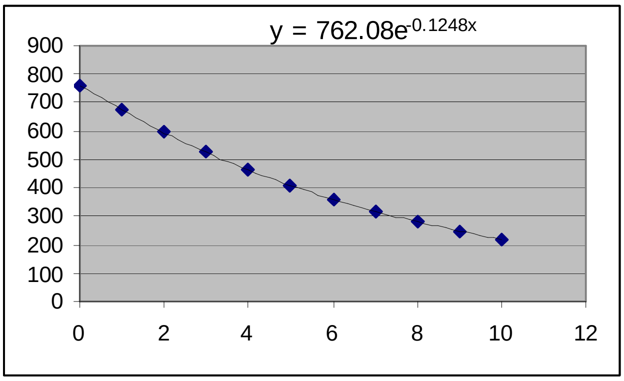
La retta di regressione tra x e $\ln(y)$ fornisce la relazione

$$\ln(y) = 6.64 - 0.125x$$

e dunque il modello esponenziale tra x e y è

$$y = \exp(6.64) \cdot \exp(-0.125)^x \approx 762 \cdot 0.88^x,$$

il che significa che la pressione decresce circa del 12% per ogni aumento di altezza di 1 km sul livello del mare.
Usando il comando "Aggiungi linea di tendenza" di tipo esponenziale otteniamo direttamente grafico ed equazione della funzione esponenziale cercata, espressa da Excel nella forma ae^{kx} anziché nella forma ab^x.



Anche per la regressione esponenziale Excel "passa ai logaritmi" e risolve poi il corrispondente problema di regressione lineare. Ricordiamo che anche in questo caso la coppia (a,b) che minimizza la funzione

$$T(a,b) := \sum_{i=1}^{n} \left(\ln(a) + x_i \ln(b) - y_i\right)^2$$

non è uguale alla coppia (a,b) che minimizza la funzione

$$S(a,b) := \sum_{i=1}^{n} \left(ab^{x_i} - y_i\right)^2 .$$

Tuttavia l'approssimazione è in generale assai buona, e quindi utilizzabile per qualsiasi scopo pratico.

5.4 Regressione logistica

Consideriamo ora una funzione per cui non sia possibile (come abbiamo fatto per le funzioni potenza e le funzioni esponenziali) la "linearizzazione". L'esempio seguente può essere utilizzato come modello per qualsiasi altra funzione. Supponiamo che le osservazioni di due grandezze X e Y siano le seguenti.

	A	B
1	**X**	**Y**
2	1	2
3	2	3
4	3	5
5	4	7
6	5	8

Vogliamo adattare a questi dati una *funzione logistica*, cioè una funzione a tre parametri a, b, c, del tipo

$$f(x) = \frac{a}{b + ce^{-ax}}.$$

La caratteristica della funzione logistica è di avere un asintoto orizzontale: se $x \to \infty$ allora $e^{-ax} \to 0$ e l'asintoto orizzontale è la retta $y = a/b$. Se $c > 0$, la funzione è crescente, dapprima in modo convesso e poi in modo concavo, con pendenza che tende a 0. È un modello molto importante per descrivere evoluzioni che presentano inizialmente una rapida crescita, per poi stabilizzarsi.
La somma dei quadrati degli scarti è la funzione

$$S(a, b, c) = \sum_{i=1}^{n} \left(\frac{a}{b + ce^{-ax_i}} - y_i \right)^2,$$

le cui derivate parziali non sono lineari rispetto ai parametri a, b, c. Partiamo dai valori iniziali $a = 1$, $b = 0.1$, $c = 1$. Calcoliamo come al solito i quadrati degli scarti, mediante la formula in E2

=($D$1/($D$2+$D$3*EXP(–$D$1*A2))–B2)^2

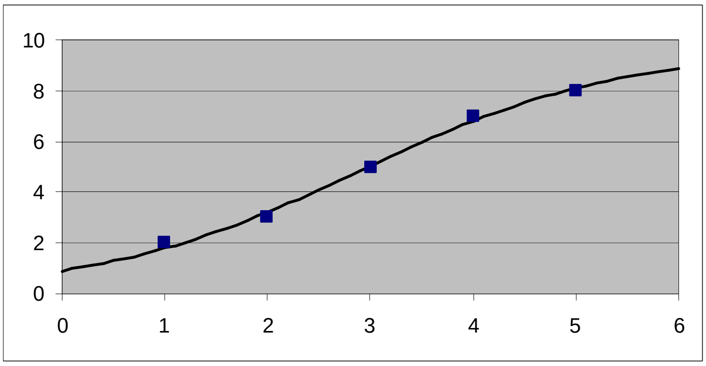
e la loro somma in E7. Con i valori iniziali impostati la somma dei quadrati degli scarti è 8.37.

	A	B	C	D	E
1	**X**	**Y**	**a =**	1	**(f(xi)-yi)^2**
2	1	2	**b =**	0,1	0,018852
3	2	3	**c =**	1	1,560642
4	3	5			2,8094579
5	4	7			2,1082112
6	5	8			1,8734473
7					8,3706104

Ora mediante il risolutore impostiamo la cella E7 al valore minimo, cambiando le celle D1:D3.
Otteniamo $a \approx 0.784$, $b \approx 0.081$, $c \approx 0.779$ e la somma dei quadrati degli scarti vale ora circa 0.14.

	A	B	C	D	E
1	**X**	**Y**	**a =**	0,7841832	**(f(xi)-yi)^2**
2	1	2	**b =**	0,0813957	0,0424866
3	2	3	**c =**	0,7793087	0,0469173
4	3	5			0,0017685
5	4	7			0,038036
6	5	8			0,009485
7					0,1386934

Il grafico seguente mostra il buon adattamento della curva ai dati.


10
8
6
4
2
0
0
1
2
3
4
5
6

6. Risolvere equazioni differenziali

Come si è visto, Excel si presta particolarmente a tradurre gli algoritmi di tipo ricorsivo, in cui ogni elemento è costruito a partire dai precedenti. Gli algoritmi per la risoluzione di equazioni differenziali sono di questo tipo.

6.1 Algoritmo di Eulero

Supponiamo di voler risolvere l'equazione differenziale nella funzione incognita $x(t)$

$$x' = \frac{x-t}{x+t}$$

con la condizione iniziale $x(0)=1$. Si tratta di un'equazione differenziale che non ammette una soluzione simbolica, e perciò si presta bene ad illustrare un metodo di calcolo che non è possibile effettuare se non mediante un'approssimazione.
Supponiamo di voler approssimare la funzione $x(t)$ nell'intervallo $t \in [0,4]$. Fissiamo un *passo* di approssimazione, per esempio $\Delta t = 0.1$; in questo modo l'algoritmo di Eulero fornisce le approssimazioni $x(0.1)$, $x(0.2)$, …, $x(2)$. Tanto più è piccolo Δt, tanto migliore sarà l'approssimazione. Conosciamo, della funzione incognita $x(t)$, il punto da cui parte: è il punto $(0,1)$; ma l'equazione differenziale ci fornisce anche, in tale punto, la derivata. Infatti sostituendo 0 a t e 1 a x nell'equazione differenziale, otteniamo $x'(0)=1$. Mediante il differenziale primo, possiamo approssimare $x(0.1)$:

$$x(0.1) \approx x(0) + x'(0) \cdot 0.1 = 1+1 \cdot 0.1 = 1.1.$$

Con $x(0.1)$ possiamo approssimare $x'(0.1)$ Sostituendo 0.1 a t e 1.1 a x nell'equazione differenziale otteniamo $x'(0.1) \approx \dfrac{1.1-0.1}{1.1+0.1} \approx 0.83$. Possiamo ora approssimare $x(0.2)$:

$$x(0.2) \approx x(0.1) + x'(0.1) \cdot 0.1 \approx 1.1 + 0.83 \cdot 0.1 \approx 1.183.$$

Possiamo così proseguire: mediante l'approssimazione di $x(t)$ otteniamo l'approssimazione di $x'(t)$ e quindi di $x(t+\Delta t)$:

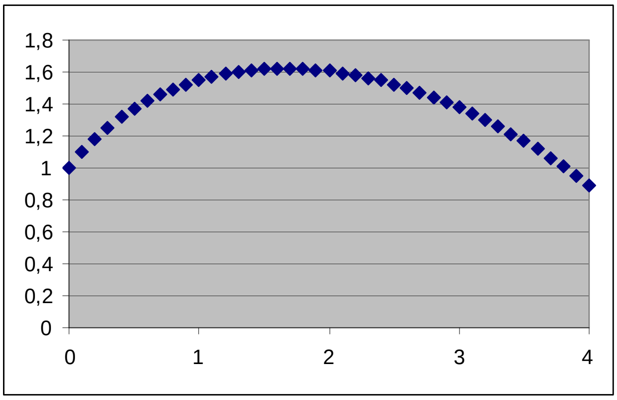
$$x(t+\Delta t) = x(t) + x'(t)\,\Delta t.$$

Vediamo come implementare questo algoritmo (noto come *algoritmo di Eulero*) con Excel. Scriviamo:

- in A2 il passo dell'approssimazione;
- in B2 il valore iniziale di t, cioè 0;
- in C2 il valore di $x(0)=1$
- in D2 scriviamo la formula =(C2−B2)/(C2+B2), che traduce l'equazione differenziale;
- in B3 la formula =B2+$A$2, che fa passare da t a $t+\Delta t$;
- in C3 l'algoritmo di Eulero: =C2+D2*$A$2;
- copiamo D2 in D3.

Ora selezioniamo le celle B3:D3 e copiamo verso il basso, fino al valore di t desiderato. Nella figura seguente sono mostrate le prime e le ultime celle della tabella, e il grafico nell'intervallo $[0,4]$.

	A	B	C	D
1	**dt**	**t**	**x(t)**	**x'(t)**
2	0.1	0	1	1
3		0.1	1.1	0.833333
4		0.2	1.183333	0.710843
5		0.3	1.254418	0.614003
6		0.4	1.315818	0.53375
7		0.5	1.369193	0.46501
8		0.6	1.415694	0.404672
9		0.7	1.456161	0.350698
10		0.8	1.491231	0.301685
11		0.9	1.521399	0.256628
12		1	1.547062	0.214782



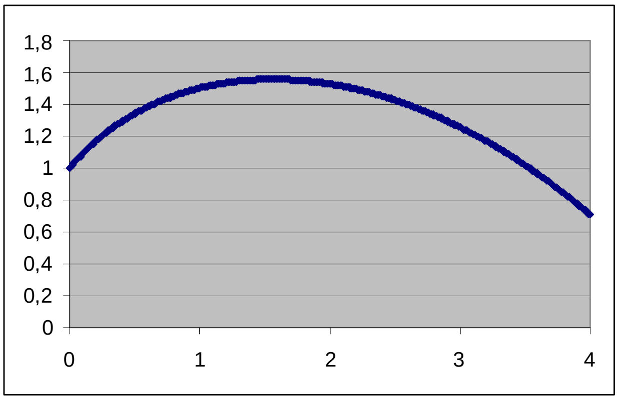
	A	B	C	D
39		3.7	1.063622	-0.55344
40		3.8	1.008278	-0.580608
41		3.9	0.950217	-0.608175
42		4	0.889399	-0.636193

Con $\Delta t = 0.1$ si ottiene l'approssimazione $x(1) = 1.547$ e $x(4) = 0.8894$. Un software professionale di matematica fornisce

$x(1)$ = 1.49828…
$x(4)$ = 0.686569…

L'errore relativo nella stima di $x(t)$ è circa il 3% in $t = 1$, e addirittura il 30% in $t = 4$. Come si vede, la propagazione dell'errore con l'algoritmo di Eulero può essere ragguardevole.
Diminuendo Δt, per esempio $\Delta t = 0.01$, l'approssimazione migliora. È sufficiente, allo scopo, modificare il parametro nella cella A2 e copiare le celle B3:D3 fino ad ottenere 4 nella colonna B (fino alla riga 402).

	A	B	C	D
1	**dt**	**t**	**x(t)**	**x'(t)**
2	0.01	0	1	1
3		0.01	1.01	0.980392
4		0.02	1.019804	0.961531
5		0.03	1.029419	0.943365
6		0.04	1.038853	0.925847



	A	B	C	D
101		0.99	1.501052	0.205155
102		1	1.503104	0.200992
103		1.01	1.505114	0.196855

	A	B	C	D
399		3.97	0.728314	-0.689968
400		3.98	0.721414	-0.693108
401		3.99	0.714483	-0.696254
402		4	0.70752	-0.699408

Risulta ora $x(1) = 1.503$ e $x(4) = 0.707$; l'approssimazione è nettamente migliorata: l'errore è 0.3% in $t = 1$ e 3% in $t = 4$.

6.2 Algoritmo di Runge-Kutta

Con l'algoritmo di Eulero l'approssimazione di $x(\Delta t)$ consiste nel sommare a $x(0)$ il differenziale primo $x'(0)\Delta t$:

$x(\Delta t) \approx x(0) + x'(0)\Delta t$

Il teorema di Lagrange ci consente di trasformare l'approssimazione in un'uguaglianza: infatti sotto opportune ipotesi di regolarità esiste un punto $c \in (0, \Delta t)$ tale che

$x(\Delta t) = x(0) + x'(c)\Delta t$

Non sappiamo dove sia c, ma ipotizziamo che scegliendo c come punto medio dell'intervallo $[0, \Delta t]$ l'approssimazione possa migliorare:

$x(\Delta t) \approx x(0) + x'(\Delta t/2)\Delta t$.

Tale metodo è noto come *algoritmo di Runge-Kutta*.
Consideriamo di nuovo l'equazione differenziale

$$x' = \frac{x-t}{x+t}$$

con la condizione iniziale $x(0) = 1$. Approssimando $x(0.1)$ con l'algoritmo di Eulero avevamo ottenuto:

$x(0.1) \approx x(0) + x'(0)\cdot 0.1 = 1 + 1\cdot 0.1 = 1.1$.

Con l'algoritmo di Runge-Kutta risulta:

$x(0.1) \approx x(0) + x'(0.05)\cdot 0.1$

Dobbiamo calcolare $x'(0.05)$; dall'equazione differenziale sappiamo che

$$x'(0.05) = \frac{x(0.05) - 0.05}{x(0.05) + 0.05}.$$

Dunque ci serve approssimare $x(0.05)$; per fare questo ricorriamo al differenziale primo:

$x(0.05) \approx x(0) + x'(0)\cdot 0.05 = 1+1\cdot 0.05 = 1.05$.

Sostituendo:

$$x'(0.05) \approx \frac{1.05 - 0.05}{1.05 + 0.05} \approx 0.909$$

e finalmente

$x(0.1) \approx x(0) + x'(0.05)\cdot 0.1 = 1+0.909\cdot 0.1 = 1.0909$.

Si prosegue poi nello stesso modo:

$x(0.2) \approx x(0.1) + x'(0.15)\cdot 0.1$

e così via, con passo Δt, fino a 4.
Vediamo ora l'implementazione dell'algoritmo di Runge-Kutta con Excel.
Le prime quattro colonne A, B, C, D sono le stesse dell'algoritmo di Eulero. Ci servono due nuove colonne E e F, in cui calcolare prima $x(t+\Delta t/2)$ con il differenziale primo, e poi $x'(t+\Delta t/2)$ mediante l'equazione differenziale.
In E2 immettiamo la formula

=C2+D2*$A$2/2

In F2 immettiamo la formula

=(E2-B2-$A$2/2)/(E2+B2+$A$2/2)

Passiamo alla riga 3: in B3 scriviamo la formula

=B2+$A$2

In C3 immettiamo la formula che rappresenta il cuore dell'algoritmo di Runge-Kutta:

=C2+F2*$A$2

Copiamo le celle D2:F2 in D3:F3, poi selezioniamo l'intervallo B2:F2 e copiamo verso il basso, fino alla riga 42 (cioè fino a $t = 4$).

	A	B	C	D	E	F
1	**dt**	**t**	**x(t)**	**x'(t)**	**x(t+dt/2)**	**x'(t+dt/2)**
2	0.1	0	1	1	1.05	0.909091
3		0.1	1.090909	0.832061	1.132512	0.766084
4		0.2	1.167517	0.707499	1.202892	0.655859
5		0.3	1.233103	0.608637	1.263535	0.56617
6		0.4	1.28972	0.526549	1.316048	0.490388
7		0.5	1.338759	0.456155	1.361567	0.424556
8		0.6	1.381215	0.394311	1.40093	0.366141
9		0.7	1.417829	0.338946	1.434776	0.313431
10		0.8	1.449172	0.288627	1.463603	0.265215
11		0.9	1.475693	0.242326	1.48781	0.220612
12		1	1.497755	0.199281	1.507719	0.178956

	A	B	C	D	E	F
39		3.7	0.883745	-0.6144	0.853025	-0.629363
40		3.8	0.820809	-0.644734	0.788572	-0.659994
41		3.9	0.754809	-0.675686	0.721025	-0.691278
42		4	0.685681	-0.707329	0.650315	-0.723289

Già con $\Delta t = 0.1$ l'algoritmo di Runge-Kutta fornisce una stima di $x(t)$ con un errore relativo pari a 0.03% in $t = 1$ e 0.1% in $t = 4$.
Con $\Delta t = 0.01$ l'errore è praticamente nullo.

6.3 Il modello preda-predatore di Lotka-Volterra

Con poco sforzo possiamo adattare l'algoritmo di Runge-Kutta per approssimare la soluzione di un sistema di equazioni differenziali.
Supponiamo di voler descrivere la soluzione del seguente sistema di equazioni differenziali nelle funzioni incognite $x(t)$, $y(t)$.

$$\begin{cases} x' = x - 0.2xy \\ y' = -y + 0.2xy \end{cases}$$

con le condizioni iniziali $x(0) = 10$, $y(0) = 2$. Vogliamo analizzare le funzioni incognite nell'intervallo di tempo [0, 8].
Si tratta del celebre modello *preda-predatore* formulato da Vito Volterra e Alfred Lotka. In un ambiente circoscritto convivono due specie, una delle quali (i predatori) si ciba dell'altra (le prede).
Le funzioni incognite rappresentano:

- $x(t)$: il numero di prede al tempo t
- $y(t)$: il numero di predatori al tempo t

In assenza di predatori le prede crescerebbero con tasso di crescita costante a: $x' = ax$. La presenza dei predatori fa sì che il tasso di crescita a non sia costante, ma decrescente (per esempio linearmente) con il numero y di predatori:

$x' = (a - by)x$

I predatori, in assenza di prede, si estinguerebbero con tasso di mortalità costante c: $y' = -cy$. La presenza delle prede fa sì che il tasso di decrescita $-c$ dei predatori cresca (per esempio linearmente) con il numero delle prede:

$y' = (-c+dx)y$

Nell'esempio proposto i parametri sono così scelti: $a = c = 1$, $b = d = 0.2$; inizialmente le prede sono 5 volte più numerose dei predatori.
Scegliamo il passo $\Delta t = 0.1$. Per approssimare $x(0.1)$, $y(0.1)$ calcoliamo

$x(0.1) \approx x(0)+x'(0.05)\cdot 0.1$
$y(0.1) \approx y(0)+y'(0.05)\cdot 0.1$

Dunque ci serve calcolare $x(0.05)$ e $y(0.05)$, che approssimiamo mediante il differenziale primo:

$x(0.05) \approx x(0)+x'(0)\cdot 0.05$
$y(0.05) \approx y(0)+y'(0)\cdot 0.05$

Dopodiché si prosegue nello stesso modo per $t = 0.2, 0.3, \dots, 8$.
Vediamo ora l'implementazione in Excel.
Scriviamo:

- in A2 il passo utilizzato, per esempio 0.1;
- in B2 il valore iniziale di t, cioè 0;
- in C2 il valore iniziale $x(0)$, cioè 10;
- in D2 il valore iniziale $y(0)$, cioè 2;
- in E2 il valore di $x'(0)$, cioè la formula =C2–0.2*C2*D2;
- in F2 il valore di $y'(0)$, cioè la formula = –D2+0.2*C2*D2;
- in G2 l'approssimazione di $x(0.05)$, cioè la formula =C2+E2*$A$2/2;
- in H2 l'approssimazione di $y(0.05)$, cioè la formula =D2+F2*$A$2/2;
- in I2 l'approssimazione di $x'(0.05)$, cioè la formula =G2–0,2*G2*H2;
- in J2 l'approssimazione di $y'(0.05)$, cioè la formula = –H2+0,2*G2*H2;
- in B3 aggiorniamo la variabile tempo incrementandola di 0.1;

Finalmente in C3 e D3 il vero e proprio algoritmo:

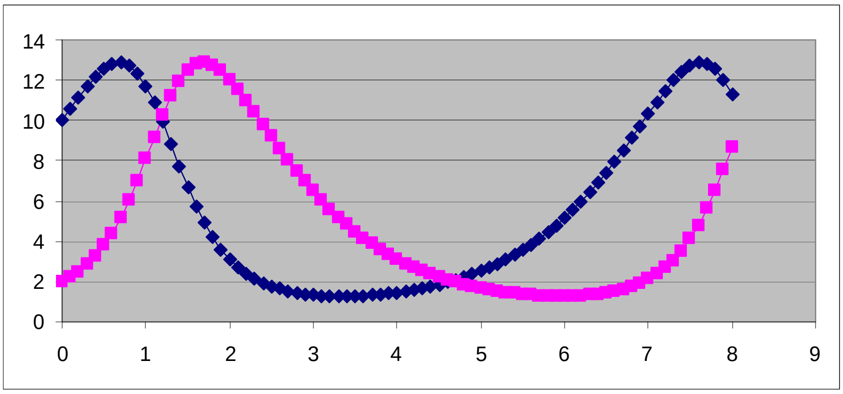
=C2+I2*$A$2
=D2+J2*$A$2

Si selezionano ora le celle E2:J2 e si copiano in E3:J3. Poi si seleziona tutta la riga B3:J3 e si copia verso il basso, fino alla riga 82.
La figura seguente mostra le prime e le ultime righe.

	A	B	C	D	E	F	G	H	I	J
1	**dt**	**t**	**x(t)**	**y(t)**	**x'(t)**	**y'(t)**	**x(t+dt/2)**	**y(t+dt/2)**	**x'(t+dt/2)**	**y'(t+dt/2)**
2	0,1	0	10	2	6	2	10,3	2,1	5,974	2,226
3		0,1	10,60	2,22	5,89	2,49	10,89	2,35	5,78	2,77
4		0,2	11,18	2,50	5,59	3,09	11,45	2,65	5,38	3,43
5		0,3	11,71	2,84	5,06	3,82	11,97	3,03	4,71	4,22
6		0,4	12,18	3,26	4,23	4,69	12,40	3,50	3,72	5,17
7		0,5	12,56	3,78	3,06	5,71	12,71	4,07	2,37	6,27

	A	B	C	D	E	F	G	H	I	J
78		7,6	12,85	4,80	0,52	7,53	12,88	5,17	-0,45	8,15
79		7,7	12,81	5,61	-1,57	8,76	12,73	6,05	-2,67	9,35
80		7,8	12,54	6,55	-3,88	9,87	12,35	7,04	-5,04	10,34
81		7,9	12,04	7,58	-6,21	10,67	11,72	8,11	-7,30	10,91
82		8	11,31	8,67	-8,30	10,94	10,89	9,22	-9,19	10,86

Il grafico seguente mostra l'evoluzione periodica delle due specie.



7. Numeri casuali

Ogni software di calcolo possiede un *generatore di numeri casuali*, che simula l'estrazione a caso (cioè con distribuzione uniforme) di un numero reale nell'intervallo [0, 1).
Con Excel il comando per ottenere un numero casuale è

=CASUALE()

(=RAND() nella versione inglese); si tratta di una funzione senza argomenti. Ecco una matrice di numeri casuali ottenuta copiando il comando =CASUALE() nelle celle A1:E4.

	A	B	C	D	E
1	0,272097	0,212512	0,741955	0,991819	0,029641
2	0,183859	0,845849	0,378348	0,842362	0,596357
3	0,018741	0,356621	0,568826	0,068987	0,724048
4	0,912218	0,074653	0,841576	0,122172	0,688525

È possibile ottenere facilmente numeri casuali uniformemente distribuiti in un intervallo $[a, b)$. Per esempio, il comando

=6+2*CASUALE()

genera un numero casuale nell'intervallo $[6, 8)$.

	A	B	C	D	E
1	6,607158	7,944122	7,156552	7,337392	6,518058
2	6,994437	7,177371	6,341496	6,545424	6,065796
3	7,05381	7,23367	6,611818	7,601418	6,57257
4	6,900835	7,811904	6,605657	7,169853	6,702358

In generale il comando è

=*a*+(*b*–*a*)*CASUALE()

È anche possibile generare numeri interi uniformemente distribuiti nell'intervallo n, n+1, …, n+m, utilizzando la funzione TRONCA(NUMERO). Per esempio il comando

=1+TRONCA(6*CASUALE())

genera un numero compreso tra 1 e 6 e simula così un dado.

	A	B	C	D	E
1	5	6	2	3	3
2	4	6	3	6	5
3	3	5	6	5	2
4	2	2	4	2	1

In generale il comando per generare un numero intero compreso nell'intervallo n, …, n+m è

=*n*+TRONCA(*m**CASUALE)).

È anche possibile simulare un numero aleatorio discreto (e finito)

$$X \sim \begin{cases} a_1 & a_2 & \dots & a_n \\ p_1 & p_2 & \dots & p_n \end{cases}$$

(con $\sum_{i=1}^{n} p_i = 1$), almeno nel caso in cui n è piccolo. Per esempio, per simulare il numero aleatorio

$$X \sim \begin{cases} 0 & 1 & 2 \\ 0.1 & 0.3 & 0.6 \end{cases}$$

compiliamo la tabella seguente, dove nella colonna C abbiamo calcolato le probabilità cumulate di *X*.

	A	B	C
1	0	0,1	0,1
2	1	0,3	0,4
3	2	0,6	1

Nella cella D1 generiamo un numero casuale tra 0 e 1 con il solito =CASUALE(). Nella cella E1 scriviamo la formula

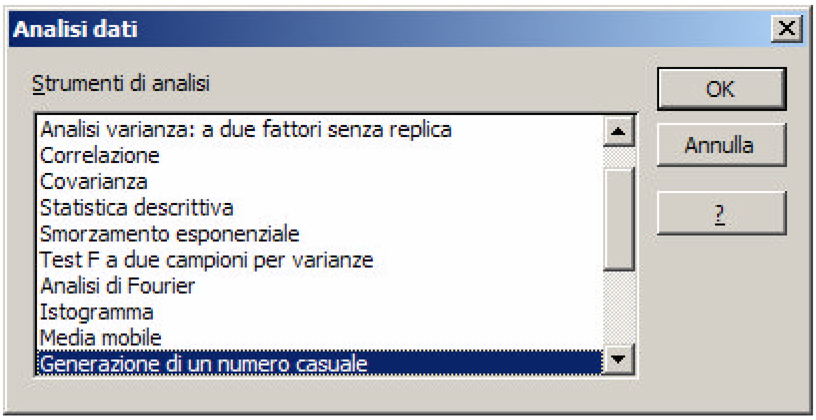
=SE(D1<$C$1;$A$1;SE(D1<$C$2;$A$2;$A$3))

che sfrutta il numero casuale in D1 per assegnare a *X* i numeri 0, 1, o 2 a seconda che D1 sia minore di 0.1, compreso tra 0.1 e 0.4, compreso tra 0.4 e 1, il che (se D1 ha distribuzione uniforme tra 0 e 1) accade proprio con probabilità rispettive 0.1, 0.3, 0.6..
Copiamo ora le celle D1:F1 in basso fino alla riga 1000 e abbiamo così costruito, nella colonna E, 1000 esperimenti sul numero aleatorio *X*. Nella figura seguente vengono mostrate le prime righe.

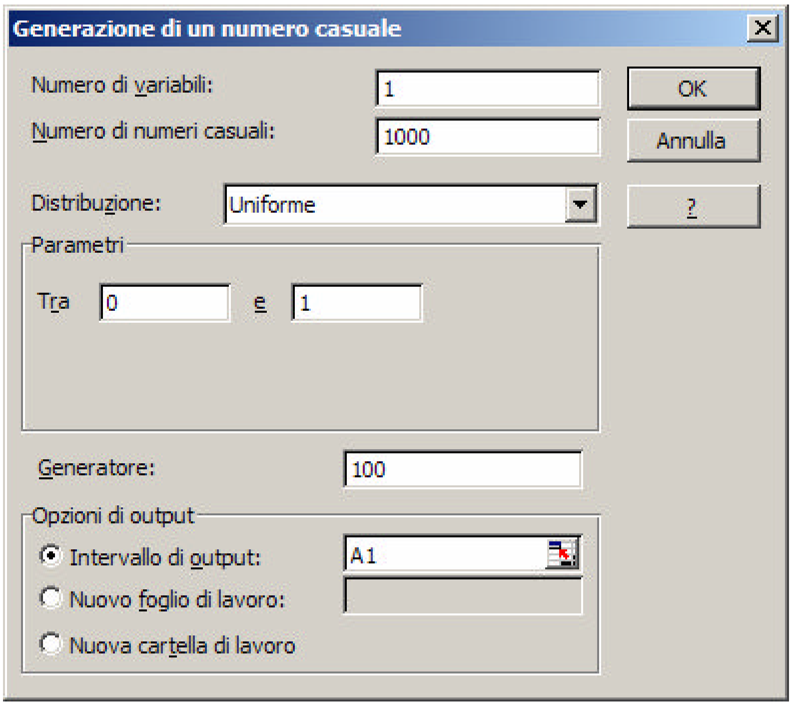
	A	B	C	D	E
1	0	0,1	0,1	0,502122	2
2	1	0,3	0,4	0,004932	0
3	2	0,6	1	0,051551	0
4				0,332013	1
5				0,384522	1
6				0,530114	2
7				0,945759	2
8				0,810762	2
9				0,694588	2
10				0,651902	2

7.1 Il generatore di numeri casuali di Excel

Per ottenere un gran numero di numeri casuali sfruttando la funzione CASUALE occorre copiare la formula; per esempio se volessimo simulare il lancio di 1000 dadi occorrerebbe scrivere in A1 la formula =1+TRONCA(6*CASUALE()) e poi copiarla fino alla cella A1000.
C'è un metodo più rapido, che sfrutta un componente aggiuntivo di Excel: *Analisi dati* dal menù Strumenti (che occorre caricare la prima volta cliccando su Strumenti, Componenti aggiuntivi, Strumenti di analisi). Nel ricco menù statistico di questo componente troviamo anche "Generazione di un numero casuale".



Se lo clicchiamo si apre un menù che offre diverse distribuzioni di numeri casuali. Possiamo generare rapidamente, ad esempio, 1000 numeri uniformemente distribuiti in [0, 1) con le seguenti impostazioni.



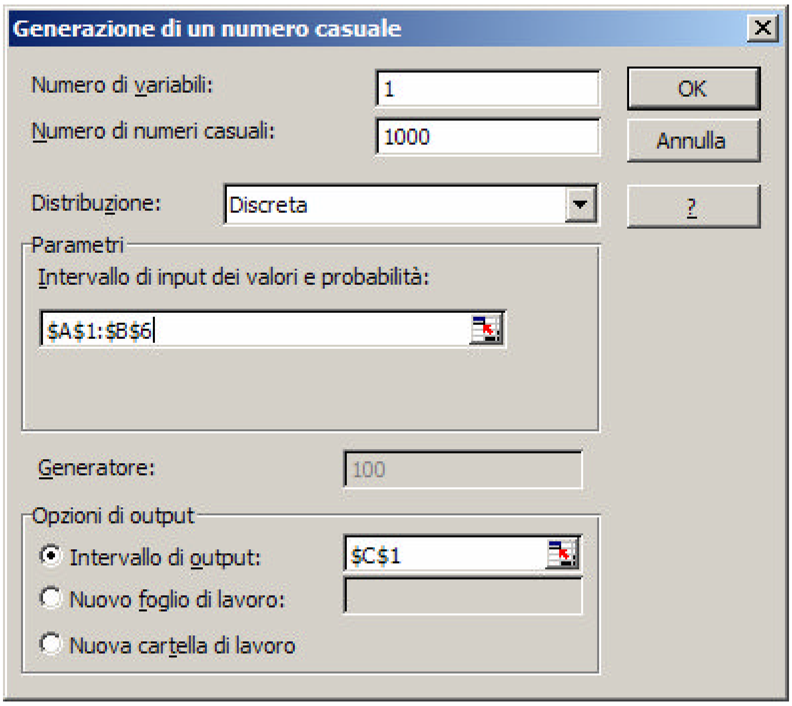
Il "Numero di variabili" ci consente di riempire più di una colonna. Il "Generatore" ci consente di scegliere un *seme* (un numero intero) per la sequenza di numeri casuali, in modo che impostando lo stesso seme si abbia sempre la stessa sequenza.
Se vogliamo invece simulare un dado dobbiamo dapprima definire la tabella della distribuzione del corrispondente numero aleatorio discreto

$$X \sim \begin{cases} 1 & 2 & 3 & 4 & 5 & 6 \\ 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \end{cases}$$

In A1:A6 scriviamo i valori assunti 1, 2, …, 6; in B1 scriviamo "=1/6" e copiamo fino a B6.

	A	B
1	1	0,166667
2	2	0,166667
3	3	0,166667
4	4	0,166667
5	5	0,166667
6	6	0,166667

Ora chiamiamo il generatore di numeri casuali e impostiamolo come nella figura seguente.



Otteniamo così 1000 lanci di un dado.

7.2 Il calcolo delle frequenze

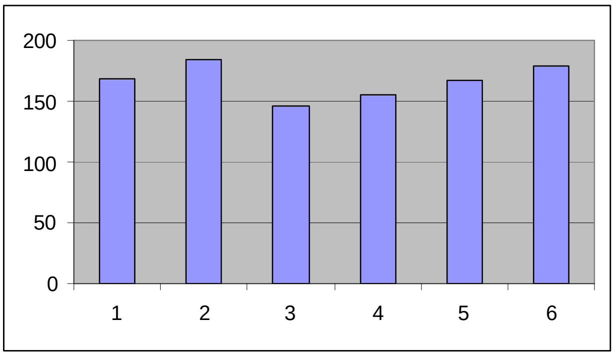
Mediante il generatore di numeri casuali simuliamo il lancio di 1000 dadi in C1:C1000.
Possiamo contare le frequenze delle diverse uscite utilizzando la funzione FREQUENZA. Selezioniamo le celle D1:D6 e scriviamo la formula

=FREQUENZA(C1:C1000;A1:A6)

e confermiamo con CTRL+SHIFT+INVIO. Otteniamo in D1:D6 le frequenze dei valori A1:A6 che compaiono nell'intervallo C1:C1000. Controlliamo con =SOMMA(D1:D6) in D7 che la somma sia 1000. Ecco le prime righe.

	A	B	C	D
1	1	0,166667	5	169
2	2	0,166667	4	184
3	3	0,166667	3	146
4	4	0,166667	6	155
5	5	0,166667	3	167
6	6	0,166667	2	179
7			5	1000
8			6	
9			5	
10			5	

Selezioniamo le celle D1:D6 e costruiamo l'istogramma delle frequenze, che appaiono ragionevolmente uniformi.



7.3 La distribuzione binomiale: lanci di *n* monete truccate

Presentiamo un esempio semplice di simulazione, di una distribuzione (la distribuzione *binomiale*) che non offre particolari difficoltà dal punto di vista teorico e che possiamo facilmente controllare. Ci serve per analizzare, nel paragrafo successivo, un esempio "impossibile" da trattare simbolicamente.

Come abbiamo visto per il dado, il generatore di numeri casuali consente di simulare agevolmente un numero aleatorio discreto (finito)

$$X \sim \begin{cases} a_1 & a_2 & \dots & a_n \\ p_1 & p_2 & \dots & p_n \end{cases}$$

con $\sum_{i=1}^{n} p_i = 1$. È sufficiente compilare una tabella dei valori $a_1, \dots, a_n$ e delle corrispondenti probabilità $p_1, \dots, p_n$ e chiamare il generatore di numeri casuali con distribuzione discreta; la somma delle probabilità naturalmente deve dare 1 (Excel esegue un controllo di coerenza).

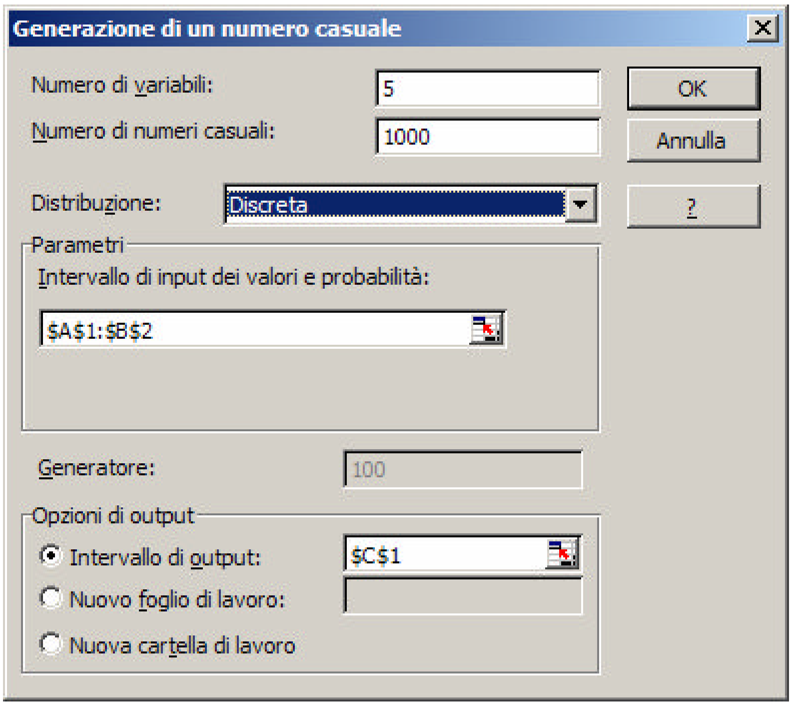
Vogliamo ora simulare il lancio di 5 monete truccate: per ciascuna moneta esce TESTA con probabilità $p = 0.6$ e CROCE con probabilità $1-p = 0.4$. Vogliamo sapere come si distribuisce la probabilità di avere, su 5 lanci, un numero di TESTE pari a 0, 1, …, 5.

Compiliamo le celle A1:B2 nel seguente modo:

	A	B
1	0	0.4
2	1	0.6

indicando con 0 l'uscita CROCE e con 1 l'uscita TESTA.

Chiamiamo il generatore, con le seguenti impostazioni.



In H1 calcoliamo la somma

=SOMMA(C1:G1)

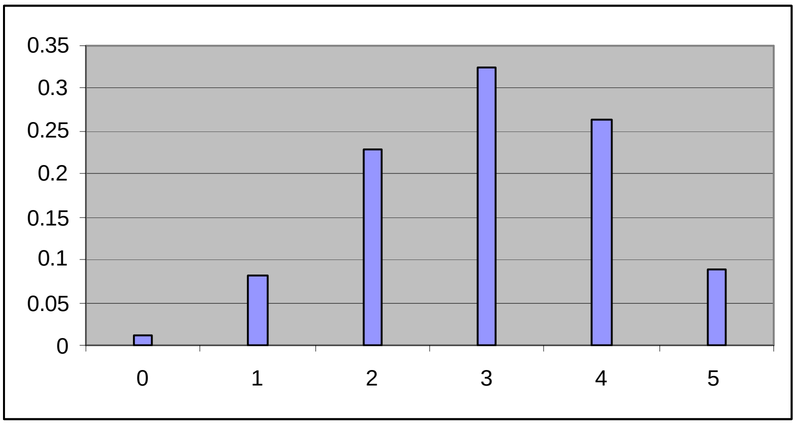
in modo da avere il numero di TESTE su 5 lanci; copiano H1 fino a H1000. Ora in I1:I6 scriviamo il numero di TESTE possibili: 0, 1, …, 5. Selezioniamo le celle J1:J6, scriviamo la formula

=FREQUENZA(H1:H1000;I1:I6)/1000

per ottenere le frequenze percentuali del numero di TESTE.
Ecco le prime dieci righe della tabella.

	A	B	C	D	E	F	G	H	I	J
1	0	0.4	0	0	0	1	1	2	0	0.011
2	1	0.6	0	1	0	1	0	2	1	0.082
3			1	1	1	1	0	4	2	0.229
4			0	0	0	1	0	1	3	0.325
5			1	0	1	0	0	2	4	0.264
6			1	1	0	1	0	3	5	0.089
7			0	1	1	0	1	3		1
8			1	1	0	0	1	3		
9			0	1	0	1	1	3		
10			0	0	1	0	0	1		

Ed ecco il grafico della distribuzione, che, come si vede, non è simmetrica, ma (poiché la moneta è truccata) è sbilanciata verso destra: è più probabile che escano quattro TESTE che una, è più probabile avere cinque TESTE che zero.



Confrontiamo infine le frequenze relative con le probabilità teoriche, che sono come è noto date da

$$p(k \text{ TESTE su 5 lanci}) = \binom{5}{k} 0.6^k \, 0.4^{5-k},$$

con $k = 0, 1, \ldots, 5$.
Excel possiede il comando DISTRIB.BINOM per la distribuzione binomiale; in K1 scriviamo
=DISTRIB.BINOM(I1;5;$B$2;FALSO)
e copiamo fino a K6. In K7 calcoliamo la SOMMA(K1:K6) per verificare che risulti 1.

	I	J	K
1	0	0.011	0.010
2	1	0.082	0.077
3	2	0.229	0.230
4	3	0.325	0.346
5	4	0.264	0.259
6	5	0.089	0.078
7		1	1

Ancora una volta le frequenze relative sono ragionevolmente vicine alle probabilità.

7.4 Somma nel lancio di *n* dadi truccati

L'esempio seguente è tutt'altro che banale, e talmente complesso dal punto di vista simbolico da scoraggiarne la trattazione: qui la strada della simulazione è senz'altro preferibile a quella teorica.
Supponiamo di avere un dado truccato modo che la probabilità sia proporzionale al numero uscito:

$$X \sim \begin{cases} 1 & 2 & 3 & 4 & 5 & 6 \\ p & 2p & 3p & 4p & 5p & 6p \end{cases}$$

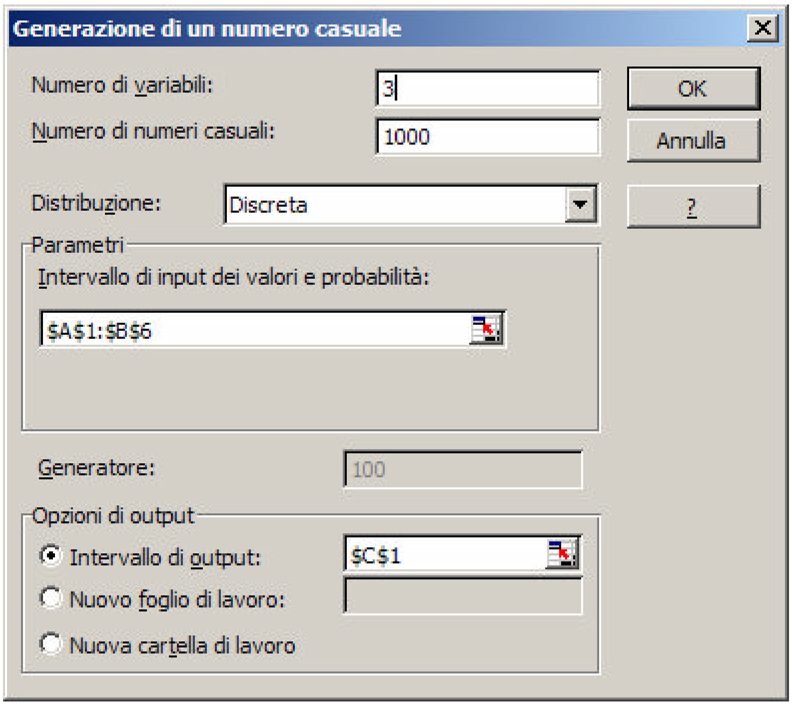
e poiché la somma delle probabilità deve dare 1, risulta $p = 1/21$.

$$X \sim \begin{cases} 1 & 2 & 3 & 4 & 5 & 6 \\ 1/21 & 2/21 & 3/21 & 4/21 & 5/21 & 6/21 \end{cases}$$

Costruiamo la corrispondente tabella in A1:B6.

	A	B
1	1	0.047619
2	2	0.095238
3	3	0.142857
4	4	0.190476
5	5	0.238095
6	6	0.285714
7		1

Proviamo ora a "lanciare" tre dadi ugualmente truccati e chiediamoci come si distribuisce la loro somma. Mediante il generatore costruiamo in C1:E1000 tre sequenze di 1000 lanci di un dado, impostandolo nel seguente modo.



In F1 calcoliamo la somma dei dadi con la formula

=SOMMA(C1:E1)

Ecco le prime righe.

	A	B	C	D	E	F
1	1	0.047619	1	1	3	5
2	2	0.095238	5	6	4	15
3	3	0.142857	6	2	5	13
4	4	0.190476	3	6	6	15
5	5	0.238095	5	5	4	14
6	6	0.285714	4	2	1	7
7		1	5	3	5	13
8			3	5	3	11
9			2	6	6	14
10			4	6	3	13

La colonna F contiene la simulazione di 1000 esperimenti sulla somma nel lancio di tre dadi truccati. Vogliamo sapere, ad esempio, qual è la probabilità di ottenere una somma minore di 10.
Per calcolare la frequenze relative dei valori ottenuti nella colonna F scriviamo in G1:G16 le possibili somme di due dadi, e cioè i numeri 3, 4, …, 18. Selezioniamo le celle H1:H16, scriviamo la formula

=FREQUENZA(F1:F1000;G1:G16)/1000

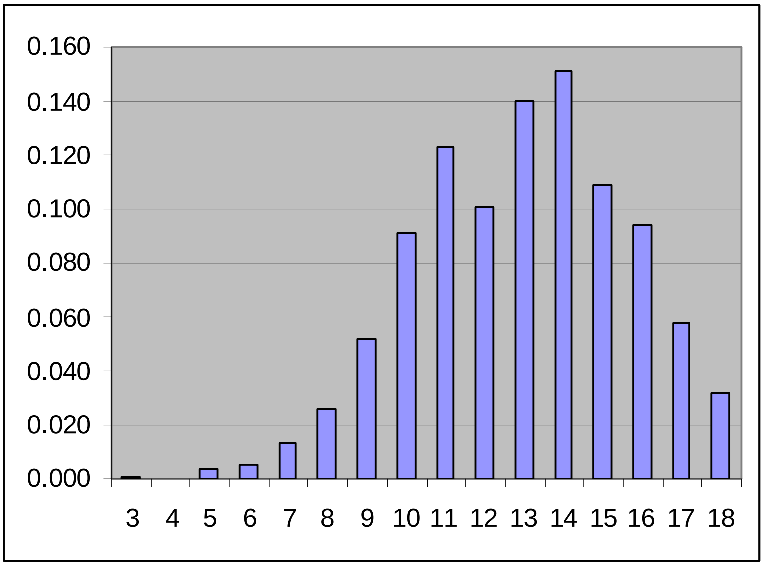
e confermiamo con CTRL+SHIFT+INVIO. Otteniamo la tabella seguente (in H17 abbiamo controllato che il totale delle frequenze fosse 1).

	G	H
1	3	0.001
2	4	0.000
3	5	0.004
4	6	0.005
5	7	0.013
6	8	0.026
7	9	0.052
8	10	0.091
9	11	0.123
10	12	0.101
11	13	0.140
12	14	0.151
13	15	0.109
14	16	0.094
15	17	0.058
16	18	0.032
17		1

Possiamo stimare la probabilità di ottenere una somma minore di 10 sommando le frequenze relative delle uscite 3, 4, …, 9: si ottiene 10.1%.
Costruiamo l'istogramma delle celle H1:H16. Si noterà che Excel, di default, assegna i valori 1, …11 sull'asse x. Per forzare sull'asse x i valori delle celle G1:G16 è sufficiente, durante la creazione del grafico, riempire, nella scheda "Serie" il campo "Etichette asse categorie (X)" con la formula

=FOGLIO1!G1:G16,

oppure selezionare con il mouse le celle G1:G16.



Come si vede la distribuzione è nettamente sbilanciata verso destra, come è giusto che sia, dato che i numeri più alti hanno probabilità maggiore di uscire.
Per ottenere la funzione di ripartizione compiliamo la colonna I nel seguente modo: in I1 scriviamo

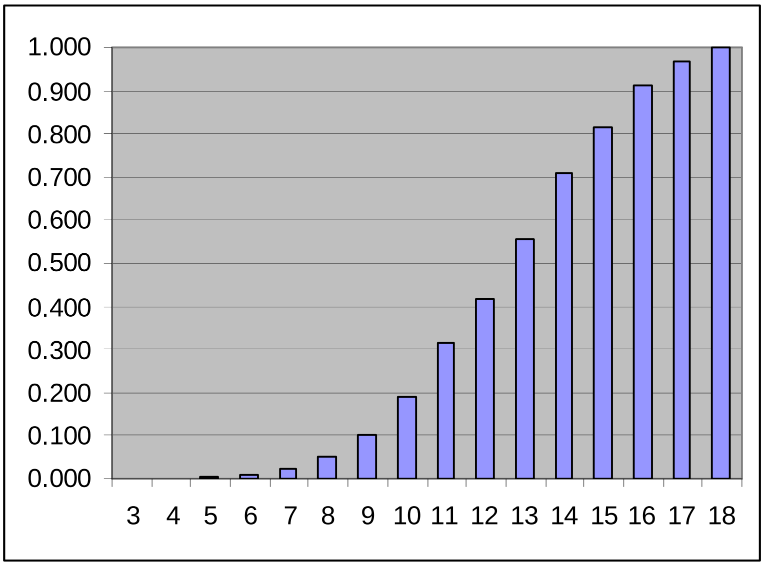
=H1

In I2 scriviamo la formula

=I1+H2

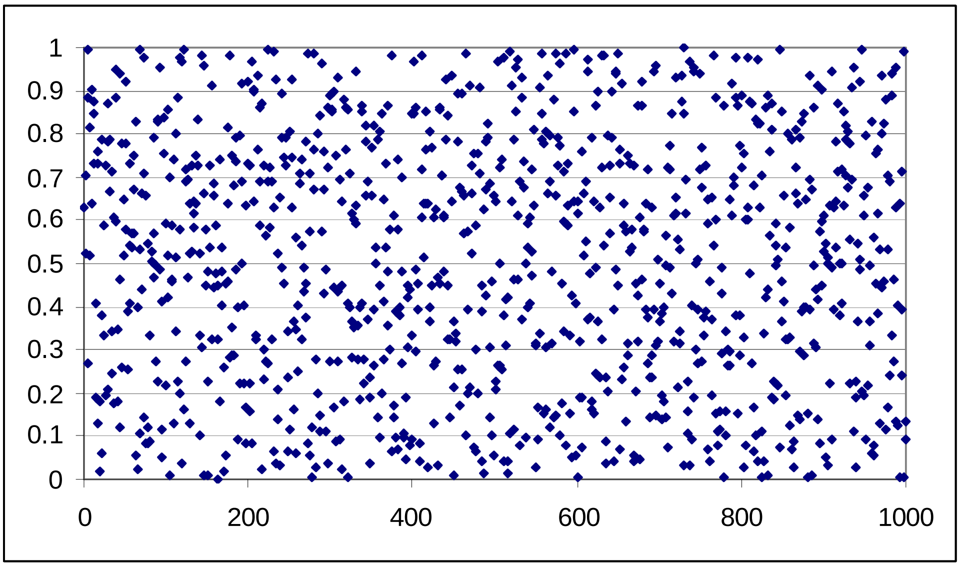
e copiamo fino a I16.

	G	H	I
1	3	0.001	0.001
2	4	0.000	0.001
3	5	0.004	0.005
4	6	0.005	0.010
5	7	0.013	0.023
6	8	0.026	0.049
7	9	0.052	0.101
8	10	0.091	0.192
9	11	0.123	0.315
10	12	0.101	0.416
11	13	0.140	0.556
12	14	0.151	0.707
13	15	0.109	0.816
14	16	0.094	0.910
15	17	0.058	0.968
16	18	0.032	1.000



7.5 Simulazione di un numero aleatorio continuo

Sappiamo già come simulare un numero aleatorio X uniformemente distribuito tra 0 e 1. Utilizziamo il generatore, con 1000 numeri casuali in A1:A1000.
Selezioniamo le celle A1:A1000 e tracciamo un grafico a dispersione.

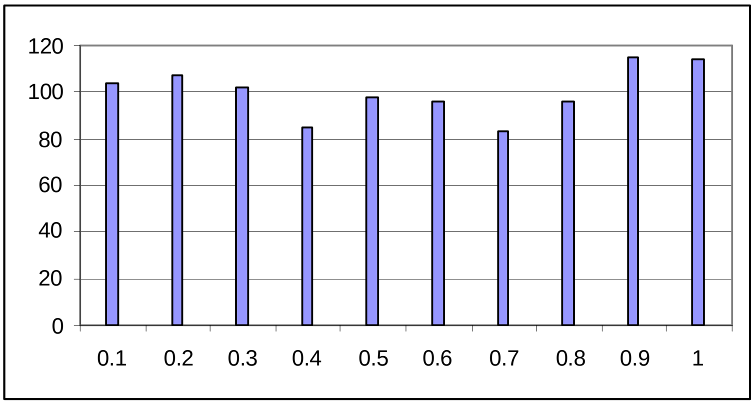


Come si vede, la distribuzione appare effettivamente uniforme: scorrendo il rettangolo $[0, 1000]\times[0, 1]$ sia orizzontalmente che verticalmente, non si notano regioni con maggiore densità di punti rispetto ad altre.
L'uniformità orizzontale conferma empiricamente il fatto che ogni numero estratto è **indipendente** dai precedenti; in altri termini l'algoritmo non degenera, continua a conservare "nel tempo" le proprie caratteristiche di casualità. L'uniformità verticale costituisce un test empirico del fatto che la distribuzione è **uniforme** tra 0 e 1: il generatore di numeri casuali "funziona bene".
Possiamo controllare l'uniformità anche calcolando la densità di frequenza; dividiamo l'intervallo [0,1] in 10 sottointervalli di ampiezza 0.1 e contiamo quale percentuale dei 1000 numeri estratti capita in ciascuno dei sottointervalli. Per fare questo nelle celle B1:B10 scriviamo gli estremi dei sottointervalli 0.1, 0.2, …, 1. Selezioniamo le celle C1:C10 e scriviamo la formula

=FREQUENZA(A1:A1000;B1:B10)

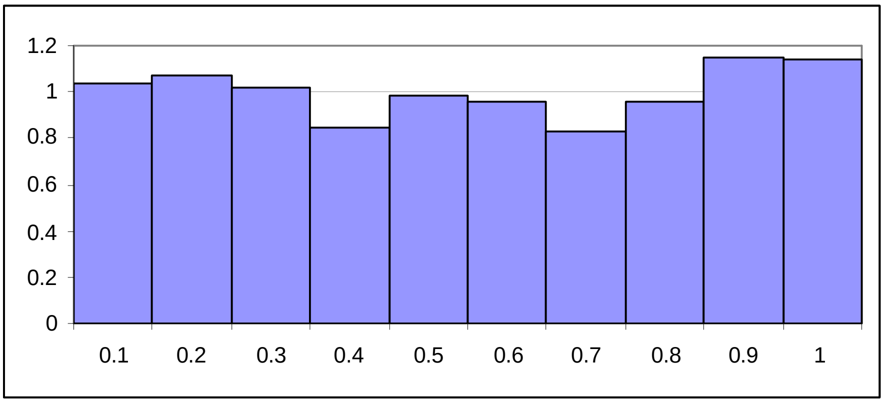
e confermiamo con CTRL+SHIFT+INVIO.
Otteniamo la tabella e il grafico seguenti; in C11 controlliamo la somma delle frequenze relative.

	B	C
1	0.1	104
2	0.2	107
3	0.3	102
4	0.4	85
5	0.5	98
6	0.6	96
7	0.7	83
8	0.8	96
9	0.9	115
10	1	114
11		1000

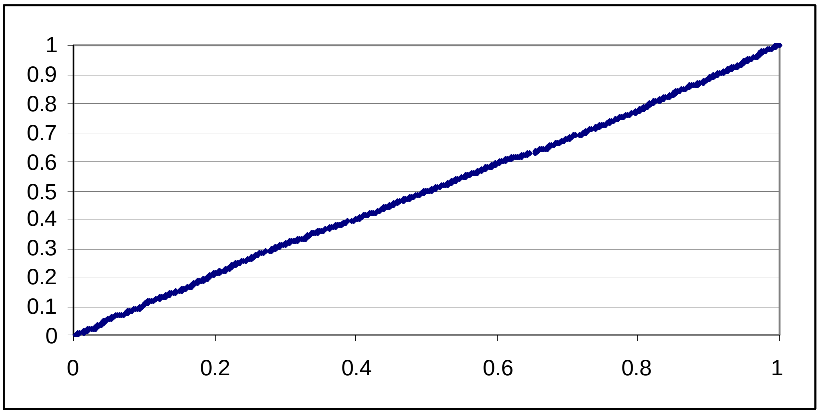


Le frequenze per ciascun intervallo sono tutte vicine a 100, cioè al 10% dei numeri casuali estratti.

Se "normalizziamo" le frequenze assolute, dividendo ciascuna di esse per il numero di numeri estratti (1000) e per l'ampiezza di ogni sottointervallo (0.1) otteniamo il vero e proprio *istogramma* della distribuzione, cioè un'approssimazione della funzione *densità di probabilità* $f(x)$.



È possibile approssimare anche la *funzione di ripartizione*? La risposta è sì: vediamo come. Innanzitutto **ordiniamo** in ordine crescente i 1000 numeri estratti: selezioniamo la colonna A, clicchiamo su Dati, Ordina, Continuare con la selezione corrente, Ordina per colonna A, Crescente.
Ora abbiamo gli stessi 1000 numeri ordinati dal minore al maggiore.
Inseriamo una colonna tra la colonna A e la colonna B; in questa colonna scriviamo i numeri 0.001, 0.002, …, 1. Ora selezioniamo le colonne A e B e costruiamo il grafico a dispersione: i numeri casuali della colonna A formeranno l'asse delle x, i numeri della colonna B l'asse y.



Sorpresa! Si ottiene il grafico di una retta, cioè di una crescita lineare, che è ciò che ci aspettiamo: il grafico mostra che i numeri casuali crescono *uniformemente* da 0 a 1 (cioè che il generatore funziona bene).
Interpretiamo il significato di un punto sul grafico di questa funzione $F(x)$: l'ascissa corrisponde ad un punto x^* scelto a caso nell'intervallo [0, 1), l'ordinata $F(x^*)$ ci dice quale percentuale dei 1000 numeri estratti è minore o uguale a x^*. È esattamente la definizione della funzione di ripartizione!
Otteniamo quindi, senza particolare sforzo, una stima dell'andamento della funzione di ripartizione semplicemente ordinandoli, e osservando come il grafico di tale funzione cresce da 0 a 1. Non ci deve dunque stupire il fatto che il grafico sia lineare. Come sappiamo tra la densità di probabilità $f(x)$ e la funzione di ripartizione $F(x)$ esistono le relazioni

$$f(x) = F'(x), \qquad F(x) = \int_0^x f(t)\,\mathrm{d}t,$$

che è esattamente ciò che abbiamo verificato: F è lineare se e solo se f è costante (cioè se e solo se la distribuzione è uniforme).
Nell'esempio del numero casuale X uniformemente distribuito tra 0 e 1 non abbiamo scoperto nulla di nuovo: ma la possibilità di approssimare sia $f(x)$ sia $F(x)$ può risultare indispensabile quando il numero aleatorio è più complesso; per esempio, quando è definito come funzione di uno o più numeri aleatori.

7.6 Funzioni di numeri aleatori

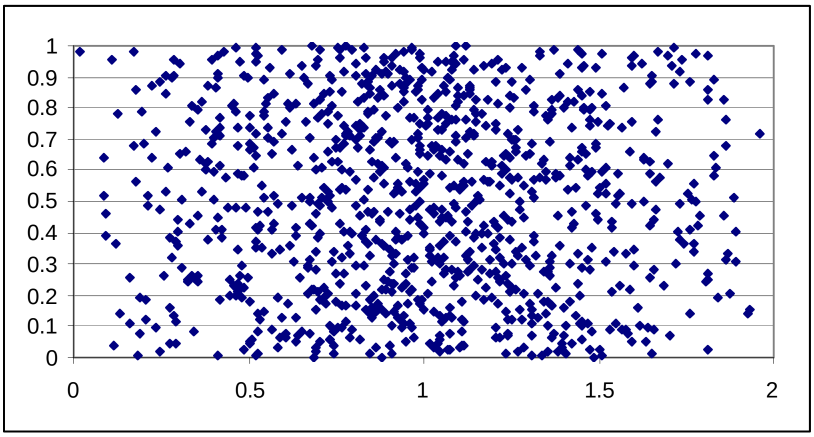
Proviamo ora a simulare la **somma** di due numeri casuali X e Y uniformemente distribuiti tra 0 e 1:

$Z = X+Y$.

Z assumerà dunque valori compresi tra 0 e 2.
Chiamiamo il generatore ponendo il "Numero di variabili" uguale a 2, in modo da avere 1000 coppie di numeri casuali nelle celle A1:B1000.
In C1 calcoliamo

=A1+B1

e copiamo fino a C1000. In D1:D1000 scriviamo i valori 0.001, 0.002, …, 1.
Tracciamo ora il grafico a dispersione della colonna D rispetto alla colonna C.

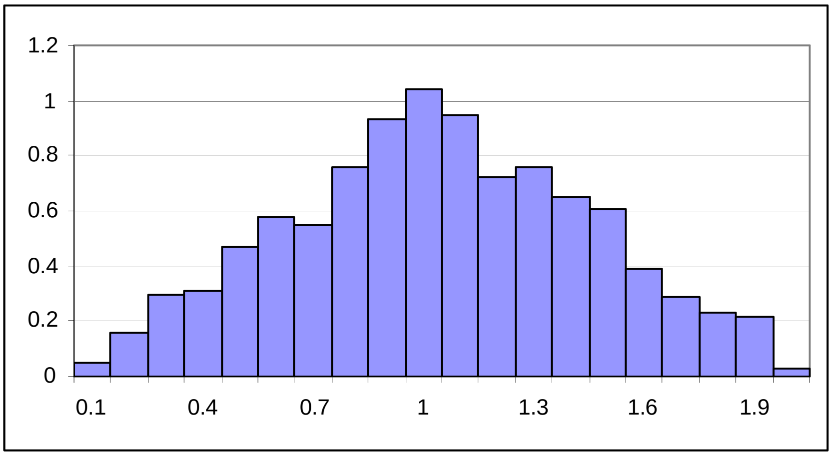


La distribuzione non appare per nulla uniforme tra 0 e 2: la densità dei punti sembra aumentare da 0 a 1 e poi diminuire da 1 a 2: appare più probabile osservare la somma di due numeri casuali più vicina a 1 piuttosto che a 0 o a 2.
Per confermare e quantificare questa impressione empirica, calcoliamo le frequenze. Dopo aver scritto i numeri 0.1, 0.2, …, 1.9, 2 nelle celle E1:E20, selezioniamo F1:F20 e scriviamo la formula

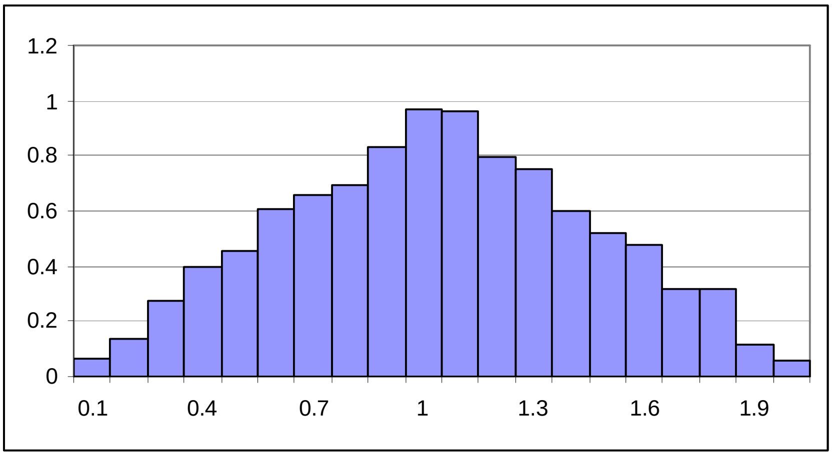
=FREQUENZA(C1:C1000;D1:D20)/(1000*0.1),

in modo da ottenere l'istogramma della distribuzione. Otteniamo la tabella e il grafico seguenti.

	E	F
1	0.1	0.05
2	0.2	0.16
3	0.3	0.3
4	0.4	0.31
5	0.5	0.47
6	0.6	0.58
7	0.7	0.55
8	0.8	0.76
9	0.9	0.93
10	1	1.04
11	1.1	0.95
12	1.2	0.72
13	1.3	0.76
14	1.4	0.65
15	1.5	0.61
16	1.6	0.39
17	1.7	0.29
18	1.8	0.23
19	1.9	0.22
20	2	0.03



Il grafico mostra un'approssimazione della *densità di probabilità* del numero aleatorio $Z = X+Y$. Una naturale congettura è che la densità di probabilità cresca linearmente da 0 a 1 (con pendenza 1) e decresca linearmente da 1 a 2 (con pendenza –1).
Se ripetiamo la stessa costruzione per 5000 numeri, otteniamo il seguente grafico, che avvalora la nostra congettura.

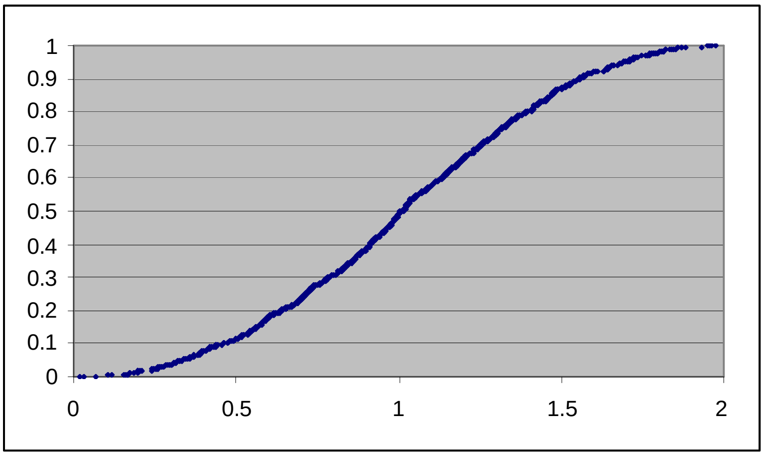


In effetti le cose stanno proprio così: se X e Y sono due numeri aleatori uniformemente distribuiti tra 0 e 1, si dimostra che il numero aleatorio $X+Y$ ha come funzione di densità di probabilità

$$f(x) = \begin{cases} x & 0 \le x < 1 \\ 2 - x & 1 \le x \le 2 \end{cases}$$

(ed è nulla fuori dall'intervallo $[0, 2]$)
Approssimiamo ora la funzione di ripartizione. Dobbiamo ordinare i numeri casuali della colonna C; questa volta non è possibile ordinarli direttamente, perché in C c'è una formula (la somma A+B). Dobbiamo prima sostituire alla formula i valori; selezioniamo le celle C1:C1000, copiamole con un CTRL+C, apriamo il menù Modifica, Incolla speciale e clicchiamo su "Valori". Ora le celle non

contengono più la formula, ma i risultati della formula. Possiamo così ordinare la colonna C con Dati, Ordina. Ecco che cosa otteniamo dal grafico a dispersione della colonna D rispetto alla colonna C.



Si tratta di un'approssimazione della funzione di ripartizione. Si può verificare che tale curva si sovrappone molto bene alla funzione

$$F(x) = \int_0^x f(t)\,\mathrm{dt} = \begin{cases} \frac{1}{2}x^2 & 0 \le x < 1 \\ 2x - \frac{1}{2}x^2 - 1 & 1 \le x \le 2 \end{cases}$$

cioè alla funzione integrale, tra 0 e 2, di $f(x)$.

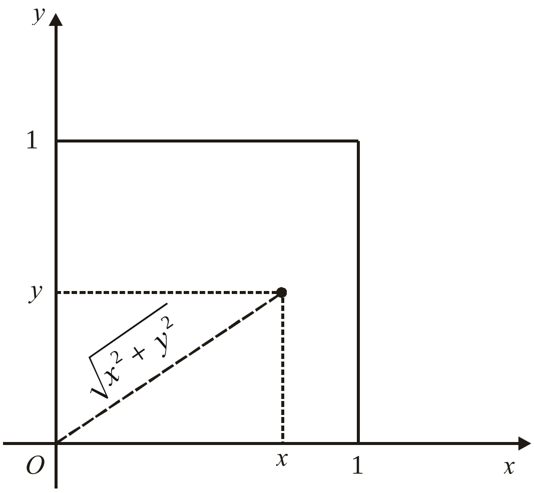
7.7 Un esempio conclusivo

Abbiamo già osservato che sfruttare un sistema di calcolo numerico introduce una sorta di *democratizzazione* nella matematica: non è necessario conoscere la complessa teoria dei sistemi dinamici per osservare empiricamente il comportamento di una successione ricorsiva; non è necessario conoscere il teorema di Laplace per calcolare il determinante di una matrice quadrata, e così via.

Mentre dal punto di vista teorico esiste una frattura tra problemi lineari (che ammettono risoluzione simbolica) e problemi non lineari (che non ammettono in generale soluzione simbolica), dal punto di vista numerico non c'è alcuna differenza: quando si implementa un problema in termini numerici la complessità svanisce. È ovvio che da una simulazione numerica non è possibile trarre conclusioni generali, ma in molti casi suggerisce congetture convincenti.

Si osservi che il problema del calcolo simbolico della funzione di ripartizione, o della funzione di densità, è intrinsecamente molto complesso, anche per numeri aleatori relativamente semplici; la soluzione, come vogliamo mostrare nell'esempio seguente, è spesso assai laboriosa da calcolare, se non impossibile, nel senso che non è esprimibile simbolicamente come funzione elementare. Di conseguenza la simulazione diventa, oltre che uno strumento di approssimazione, l'unico approccio possibile.

Supponiamo di scegliere un punto a caso nel quadrato unitario $[0, 1]\times[0, 1]$. Vogliamo sapere come si distribuisce la sua distanza dall'origine.



Traduciamo il problema in termini probabilistici: dati due numeri aleatori X, Y, entrambi uniformemente distribuiti tra 0 e 1, vogliamo determinare la funzione di densità e la funzione di ripartizione, del numero aleatorio

$$Z = \sqrt{X^2 + Y^2}\ .$$

Poiché $X \in [0, 1)$ e $Y \in [0, 1)$, allora $Z \in \left[0, \sqrt{2}\right]$.

Chiamiamo il generatore e costruiamo nelle colonne A e B 2 sequenze di 5000 numeri casuali ciascuna uniformemente distribuiti tra 0 e 1.
In C1 scriviamo la formula

=RADQ(A1^2+B1^2)

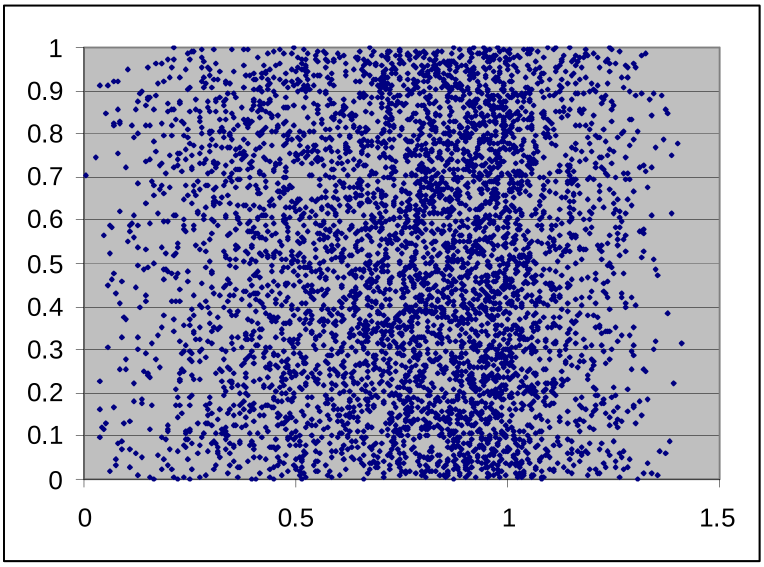
e copiamo fino a C5000 con un doppio click.
In D1 scriviamo

=1/5000

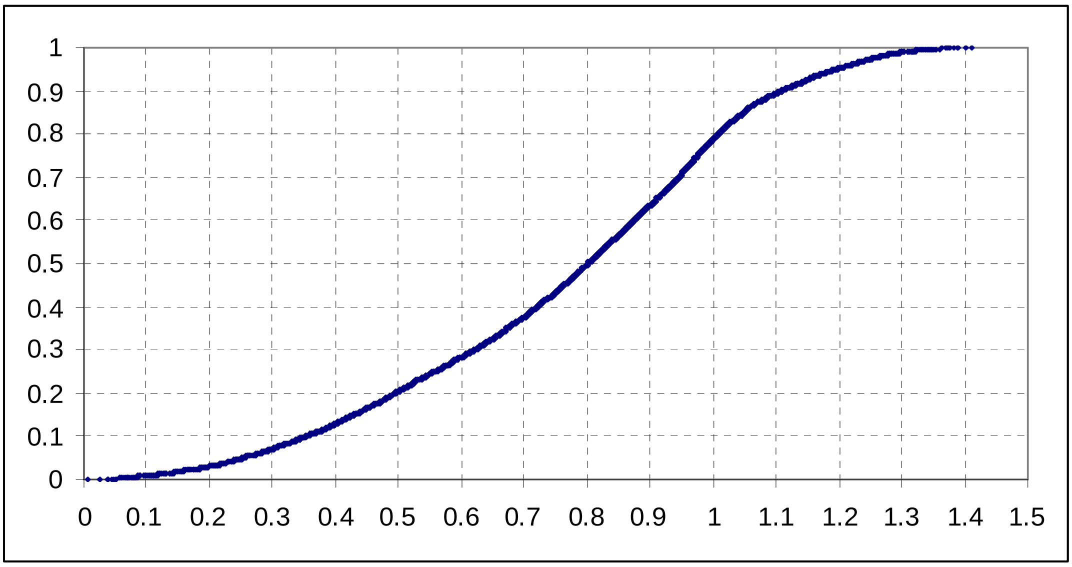
e in D2

=D1+1/5000

e con un doppio click copiamo D2 fino a D5000. Il grafico a dispersione è il seguente. Si osserva che la densità dei punti cresce da 0 fino a 1, e poi decresce rapidamente da 1 a $\sqrt{2}$.



Simuliamo la funzione di ripartizione ordinando i numeri casuali della colonna C. Si ottengono gli stessi punti del grafico precedente, ma ordinati in senso crescente.



Leggiamo qualche informazione su quest'ultimo grafico:

- la probabilità che Z sia minore di 1 è quasi l'80%;
- la probabilità che Z sia minore di 0.5 è solo del 20%;
- la probabilità che Z sia compreso tra 0.8 e 1è circa $0.8 - 0.5 = 30\%$.

Sono tutte informazioni che ricaviamo senza alcuna difficoltà.

Vediamo ora la simulazione della densità di probabilità. Dividiamo l'intervallo $\left[0, \sqrt{2}\right]$ dei valori assunti da Z in sottointervalli, per esempio di ampiezza $\Delta Z = 0.04$, contiamo per ciascun intervallo le frequenze dei 5000 numeri casuali, e normalizziamo dividendo per 5000 e per ΔZ.

Nella cella E1 scriviamo 0.04 e in E2 la formula

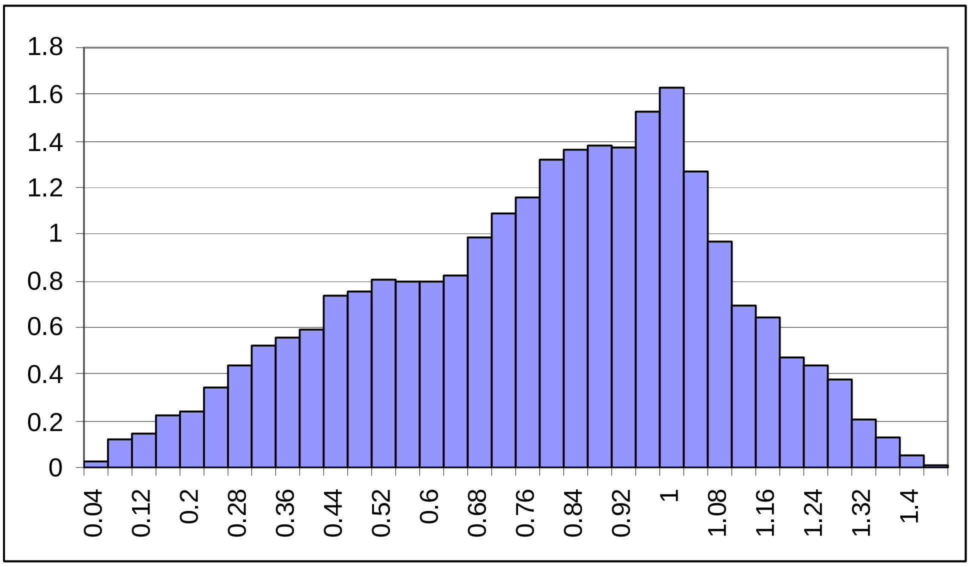
=E1+0.04

Copiamo E2 verso il basso fino al primo numero che supera $\sqrt{2}$, cioè 1.44 (quindi fino a E36).

Selezioniamo le celle F1:F36 e scriviamo la formula

=FREQUENZA(C1:C5000;E1;E36)/(5000*0.04)

e confermiamo con CTRL+SHIFT+INVIO.

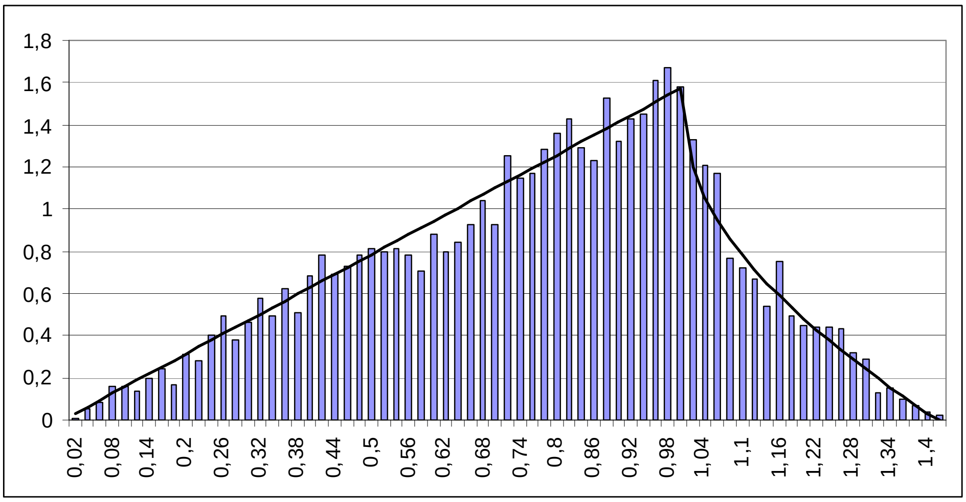


Se costruiamo l'istogramma osserviamo in effetti una crescita circa lineare tra 0 e 1 e una decrescita convessa tra 1 e $\sqrt{2}$.
Se avessimo cercato una soluzione simbolica per la funzione di ripartizione $F(z)$ e per la densità $f(z)$ avremmo trovato, dopo assai laboriosi calcoli, quanto segue.

$$F_Z(z) = \begin{cases} 0 & z < 0 \\ \dfrac{\pi}{4} z^2 & 0 \le z \le 1 \\ \sqrt{z^2 - 1} + z^2 \left(\sin^{-1}\left(\dfrac{1}{z}\right) - \dfrac{\pi}{4} \right) & 1 < z \le \sqrt{2} \\ 1 & \sqrt{2} < z \end{cases}$$

$$f_Z(z) = \begin{cases} 0 & z < 0 \\ \dfrac{\pi}{2} z & 0 \le z \le 1 \\ z\left(2\sin^{-1}\left(\dfrac{1}{z}\right) - \dfrac{\pi}{2} \right) & 1 < z \le \sqrt{2} \\ 0 & \sqrt{2} < z \end{cases}$$

Il grafico seguente mostra che la simulazione della densità si adatta molto bene al grafico di $f_Z(z)$. Non riportiamo invece il confronto tra la nostra simulazione e la la funzione di ripartizione, perché i grafici sono praticamente indistinguibili.


1,8
1,6
1,4
1,2
1
0,8
0,6
0,4
0,2
0
0,02
0,08
0,14
0,2
0,26
0,32
0,38
0,44
0,5
0,56
0,62
0,68
0,74
0,8
0,86
0,92
0,98
1,04
1,1
1,16
1,22
1,28
1,34
1,4