

Dispense per Modelli Matematici
A.A. 1999\00

F.Patrone e M. Margiocco

8 settembre 2001

1 Alberi ed evoluzione della specie

Siano date m specie viventi S_1, S_2, \dots, S_m , che si distinguono fra loro in base a n caratteri C_1, C_2, \dots, C_n ciascuno dei quali puo' presentarsi in 2 forme diverse, che indicheremo con 0 e 1 e hanno il significato di "primitiva" e "derivata" (antica-moderna, originaria-evoluta, o con termine tecnico, plesiomorfa-apomorfa). Semplificando molto le cose, possiamo pensare che C_1, C_2, \dots, C_n siano codificati da n geni del DNA che possono subire una mutazione che li fa passare dalla forma 0 alla forma 1. 0 e 1 possono anche significare assenza e presenza di un certo carattere, comparso con l'evoluzione.

Vogliamo ricostruire l' "albero filogenetico" (albero evolutivo), che da un progenitore comune $S_0 = (00\dots 0)$ (tutti i caratteri nella forma 0) ha condotto alle m specie S_1, S_2, \dots, S_m . Facciamo un esempio con $m = 4, n = 5$ e sia data la seguente "matrice dei caratteri" $A = (a_{ij})$, dove a_{ij} = valore del carattere j per la specie i :

	C_1	C_2	C_3	C_4	C_5
S_1	1	0	0	0	1
S_2	1	0	1	0	1
S_3	1	1	0	0	0
S_4	0	0	0	1	0

Definiamo nell'insieme delle specie 2 relazioni $<$ e $<<$, nel seguente modo :

$S < S'$ (S é antenato diretto di S') se e solo se S ed S' differiscono per 1 solo carattere, che vale 0 in S e 1 in S' .

$S << S'$ (S é antenato di S') se e solo se S é antenato diretto di S' , oppure esiste una catena di specie intermedie S_1, S_2, \dots, S_k che va da S a S' e tale che ogni specie sia antenata diretta della successiva (cioé $S < S_1 < S_2 < \dots < S_k < S'$).

La relazione $<$ si puó rappresentare con un grafo orientato G , i cui vertici sono le specie. In esso 2 vertici S, S' sono congiunti da un arco orientato (da S verso S') se e solo se $S < S'$. Nel nostro esempio il grafo G ha 4 vertici S_1, S_2, S_3, S_4 e un solo arco da S_1 a S_2 , corrispondente alla comparsa del carattere C_3 (vedi figura 1.1):

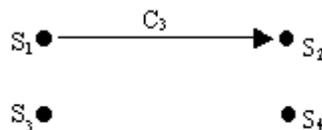


Figura 1.1

Vogliamo completare questo grafo (aggiungendo eventualmente altre specie) in modo da avere un albero evolutivo che congiunga tutti i vertici. Il procedimento é il seguente: eliminiamo S_2 , che é discendente di S_1 , e restano solo le 3 specie S_1, S_3, S_4 , con la seguente matrice dei caratteri:

	C_1	C_2	C_3	C_4	C_5
S_1	1	0	0	0	1
S_3	1	1	0	0	0
S_4	0	0	0	1	0

S_1 ed S_3 hanno un “ultimo predecessore comune”, che indichiamo con $S_1 \cap S_3 = (10000)$, in cui un carattere vale 1 se e solo se vale 1 sia in S_1 che in S_3 . Aggiungendo il vertice $S_1 \cap S_3$ al grafo G si ottiene la figura 1.2:

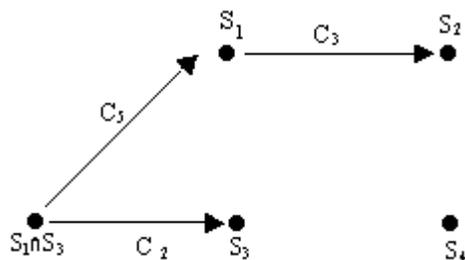


Figura 1.2

e resta fuori solo il vertice C_4 . Eliminiamo S_1, S_2, S_3 , che sono discendenti di $S_1 \cap S_3$, e restano solo le 2 specie $S_4, S_1 \cap S_3$, con la matrice dei caratteri:

	C_1	C_2	C_3	C_4	C_5
S_4	0	0	0	1	0
$S_1 \cap S_3$	1	0	0	0	0

queste 2 specie hanno un “ultimo predecessore comune” $S_1 \cap S_3 \cap S_4 = (00000) = S_0$, che é il capostipite. Aggiungendo ancora il vertice S_0 al grafo G , si ottiene l’albero filogenetico completo T (vedi Fig. 1.3):

questo é un albero orientato, con radice (il capostipite S_0) e rappresenta entrambe le relazioni $<$ e $<<$. Infatti, dati 2 vertici S, S' di T si ha che $S < S'$ se e solo se esiste un arco orientato di T che va da S a S' , si ha che $S << S'$ se e solo se esiste un cammino orientato (cioé una successione finita di archi orientati consecutivi) di T che va da S a S' . L’albero filogenetico T gode delle seguenti proprietà:

1. Ogni lato dell’albero corrisponde alla comparsa di un carattere e ogni carattere compare una volta sola in tutto l’albero.

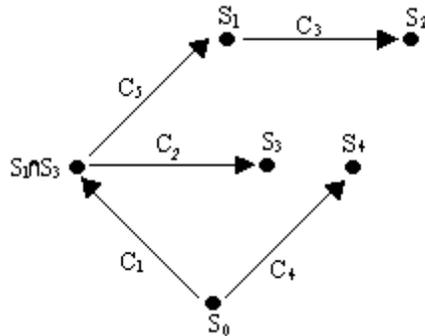


Figura 1.3

- Per ogni specie S , il cammino evolutivo che conduce dal capostipite S_0 a S è unico. In altre parole, se S' e S'' sono entrambi antenati di S , allora o S' è antenato di S'' oppure S'' è antenato di S' .
- Poiché in ogni albero il numero di vertici è uguale al numero di lati più 1, un albero filogenetico con n caratteri dá luogo a $n + 1$ specie. Detto m il numero iniziale di specie in esame, dev'essere allora $m \leq n + 1$.

Per avere queste proprietà bisogna però fare la seguente ipotesi (molto plausibile dal punto di vista biologico): ognuna delle mutazioni che provocano la comparsa di un nuovo carattere è unica ed irripetibile, quindi può comparire una sola volta in tutto l'albero filogenetico. Inoltre essa non può essere annullata da una mutazione esattamente contraria, cioè il passaggio inverso $1 \rightarrow 0$ non è possibile. Detto in parole povere: i discendenti degli elefanti non possono perdere la proboscide, e i discendenti dei coccodrilli non possono acquistare la proboscide, perché essa è già comparsa negli elefanti. Animali "ibridi" come l' "eledrillo" o il "coccofante" non possono esistere.

A questo punto è d'obbligo rivolgersi le seguenti 2 domande: l'albero filogenetico esiste sempre? Se esiste, è unico? Dal punto di vista strettamente matematico, la risposta a entrambe queste domande è NO, come dimostrano i seguenti 2 esempi:

Esempio 1.1 consideriamo

	C_1	C_2
S_1	1	0
S_2	0	1
S_3	1	1

Non esiste nessun albero filogenetico perché S_3 è discendente diretto sia di S_1 che di S_2 e questi sono discendenti diretti del capostipite $S_0 = (00)$.

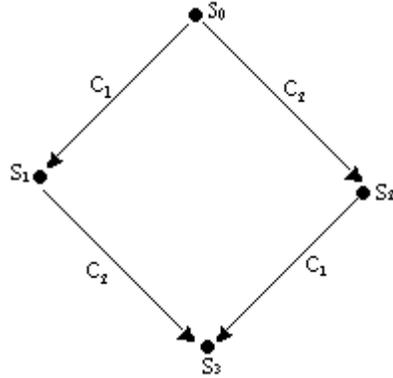


Figura 1.4

Ognuno dei 2 caratteri C_1 , C_2 compare 2 volte e ciò é impossibile. Il procedimento che abbiamo applicato prima per trovare l'albero filogenetico porta in effetti a ricostruire un "grafo filogenetico" orientato (vedi figura 1.4), però non si tratta di un albero in quanto da S_0 a S_2 esistono 2 cammini orientati diversi. ■

Esempio 1.2 consideriamo ora

	C_1	C_2	C_3	C_4
S_1	1	1	0	0
S_2	0	0	1	1

Esistono 4 diversi alberi filogenetici (vedi Fig. 1.5)

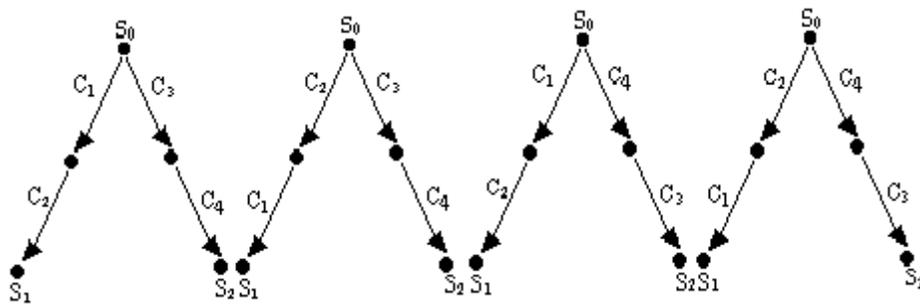


Figura 1.5

Questo perché ognuno di questi 4 alberi ha 2 "rami" (cammini orientati

senza biforcazioni intermedie) di lunghezza 2 (cioé formati da 2 archi consecutivi) e su ciascun ramo non si sa quale dei 2 caratteri corrispondenti ai 2 archi sia comparso prima e quale dopo (prima C_1 o C_2 , prima C_3 o C_4). I 4 alberi filogenetici sono sostanzialmente equivalenti, a parte l'ordine di comparsa dei caratteri su ciascun ramo. ■

Data una matrice dei caratteri relativa a n caratteri C_1, C_2, \dots, C_n per m specie S_1, S_2, \dots, S_m , se non esiste nessun albero filogenetico che la rappresenti si dice che i caratteri C_1, C_2, \dots, C_n sono *incompatibili* per le specie S_1, S_2, \dots, S_m . Si può dimostrare il seguente:

Teorema 1.1 *I caratteri C_1, C_2, \dots, C_n sono incompatibili per le specie S_1, S_2, \dots, S_m se e solo se esistono 2 caratteri diversi che compaiono in tutte e 3 le combinazioni 10–01–00 in 3 diverse specie (come nell'esempio 1 di cui sopra).*

Dimostrazione. Supponiamo che esistano 2 caratteri C_1, C_2 che compaiono in tutte e 3 le combinazioni 10–01–11 in 3 diverse specie S_1, S_2, S_3 (chiaramente i valori degli indici sono inessenziali). Siano S_4 ed S_5 le 2 specie in cui compaiono per la prima volta rispettivamente i 2 caratteri C_1 e C_2 . Poiché ogni carattere compare una volta sola e poi non scompare più, é chiaro che il carattere C_1 compare in tutte e sole le specie che sono discendenti di S_4 , mentre C_2 compare in tutte e sole le specie che sono discendenti di S_5 . Ma allora la specie S_3 é discendente sia di S_4 che di S_5 , mentre nessuna delle 2 specie S_4, S_5 é discendente dell'altra (se ad esempio S_4 fosse discendente da S_5 , dovrebbe avere i 2 caratteri nella forma 11, ma ciò é impossibile perché S_1 é discendente di S_4 e ha i 2 caratteri nella forma 10). Questo significa che nel grafo filogenetico relativo alla data matrice di caratteri ci sono 2 diversi cammini orientati tra il capostipite S_0 e la specie S_3 (uno che passa per S_4 , l'altro che passa per S_5) e quindi non esiste nessun albero filogenetico.

Viceversa, supponiamo che non esista nessun albero filogenetico per la data matrice di caratteri, cioè che nessun grafo filogenetico sia un albero. Allora deve esistere una specie S tale che dal capostipite S_0 ad S esistano 2 diversi cammini orientati. Sia S_1 la specie in cui i 2 cammini si separano per la prima volta, S_2 ed S_3 le due specie discendenti dirette di S_1 nei 2 cammini in questione, C_2 e C_3 i 2 caratteri che compaiono nei 2 passaggi $S_1 \rightarrow S_2$ e $S_1 \rightarrow S_3$. Allora é chiaro che nella matrice dei caratteri C_2 e C_3 compaiono in tutte e 3 le combinazioni 10–01–11, nelle tre specie S_2, S_3, S . Come volevasi dimostrare. ■

Esercizio 1.1 Per ciascuna delle seguenti 5 matrici di caratteri stabilire se i caratteri sono incompatibili e in caso contrario ricostruire un albero filogenetico e stabilire se é unico:

- CASO 1

$$\begin{matrix} 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 1 \end{matrix}$$

- CASO 2

$$\begin{matrix} 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 \end{matrix}$$

- CASO 3

$$\begin{matrix} 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 \end{matrix}$$

- CASO 4

$$\begin{matrix} 1 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 & 0 \end{matrix}$$

- CASO 5

$$\begin{matrix} 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \end{matrix}$$

■

Soluzione. Nel caso 1 ci sono 12 diversi alberi filogenetici.

Nei casi 2 e 3 c'è un solo albero filogenetico.

Nel caso 4 i caratteri sono incompatibili.

Nel caso 5 ci sono 4 diversi alberi filogenetici.

■

Esempio 1.3 Consideriamo la seguente matrice dei caratteri:

	C_1	C_2	C_3	C_4	C_5	C_6	C_7
S_1	0	0	0	0	0	0	0
S_2	1	0	0	0	0	0	0
S_3	1	1	0	0	0	0	0
S_4	1	1	1	1	1	0	0
S_5	1	1	1	0	0	1	1

con 5 “specie” ($S_1 =$ pesci , $S_2 =$ anfibi , $S_3 =$ rettili , $S_4 =$ uccelli , $S_5 =$ mammiferi) e 7 “caratteri” ($C_1 =$ polmoni , $C_2 =$ uovo amniotico , $C_3 =$ sangue caldo , $C_4 =$ piume , $C_5 =$ ali , $C_6 =$ peli , $C_7 =$ mammelle). Sono possibili 4 diversi alberi filogenetici (vedi figura 1.6):

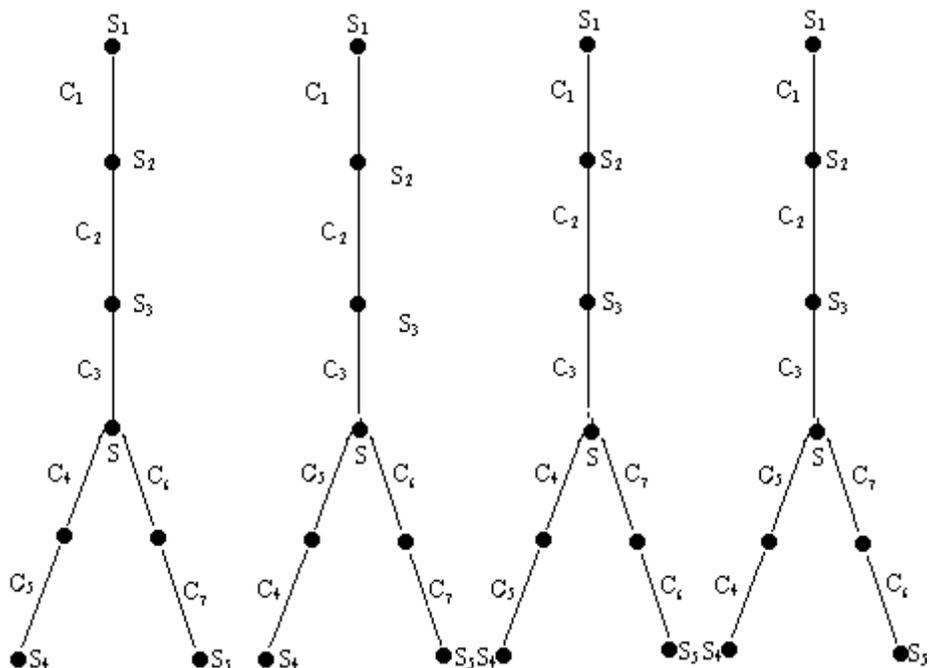


Figura 1.6

L’interpretazione é la seguente: tenendo conto solo dei 7 caratteri considerati, possiamo dire che i pesci sono antenati diretti degli anfibi (nel passaggio dai pesci agli anfibi compaiono i polmoni) e che gli anfibi sono antenati diretti dei rettili (compare l’uovo amniotico). Gli ultimi progenitori comuni degli uccelli e dei mammiferi erano una “specie” S che aveva i polmoni, l’uovo amniotico e il sangue caldo (quindi era discendente diretta dei rettili), ma non aveva piume né ali né peli né mammelle (erano i dinosauri). A questo punto l’albero evolutivo si biforca: sia gli uccelli che i mammiferi sono discendenti dei dinosauri, ma in ciascuno dei 2 rami compaiono 2 carat-

teri diversi (piume e ali per gli uccelli, peli e mammelle per i mammiferi). Senza ulteriori informazioni non possiamo sapere se sono comparse prima le piume o le ali, i peli o le mammelle e quindi sono possibili 4 diversi alberi filogenetici. ■

Osservazione 1.1 Nei casi pratici può capitare a volte di avere delle matrici di caratteri incompatibili. In tal caso, per ricostruire un albero filogenetico, si è costretti a rinunciare all'ipotesi che ogni carattere possa comparire una volta sola e poi non possa più scomparire. Rinunciando a questa ipotesi, la matrice dei caratteri dell'esempio 1.1 di cui sopra può essere spiegata da vari alberi filogenetici, per esempio (vedi Fig 1.7):

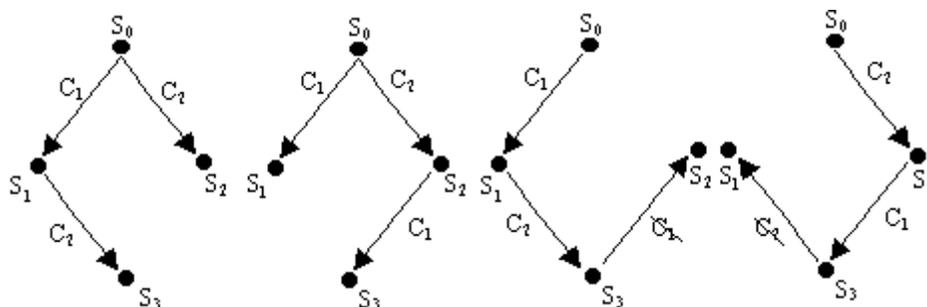


Figura 1.7

Cioè possiamo supporre che C_2 compaia 2 volte, oppure C_1 compaia 2 volte, oppure C_1 compaia e poi scompaia, oppure C_2 compaia e poi scompaia. È chiaro che si può scrivere un programma di computer in grado di calcolare tutti i possibili alberi filogenetici che spiegano una data matrice di caratteri e di scegliere fra essi quello (o quelli) che minimizzano il numero di volte in cui un carattere compare non per la prima volta, oppure il numero di volte in cui un carattere scompare, oppure tutt'e due le cose insieme, applicando i cosiddetti *criteri di parsimonia*. (Infatti in natura la scomparsa di un carattere, o la ricomparsa di un carattere già apparso, sono eventi quasi impossibili, con probabilità vicinissima a zero). ■

2 Distanza tra specie

Il metodo visto nel capitolo precedente per ricostruire l'albero evolutivo orientato che ha dato origine ad un certo numero di specie viventi, a partire dall'esame di certi caratteri qualitativi, ha senso solo se ciascun carattere può assumere solo 2 valori (primitivo - derivato, 0 - 1). Se qualche carattere può assumere un numero $n > 2$ di valori diversi, oppure infiniti valori variabili con continuità, il metodo non vale più. E lo stesso succede se ogni carattere assume solo 2 valori, ma non si sa quale dei 2 sia quello primitivo e quale sia quello derivato. In tal caso può convenire usare un altro metodo, che si basa sulla nozione di distanza e porta a costruire un albero evolutivo non orientato.

Un insieme X si chiama uno *spazio metrico* se è definita una *distanza* $d(x, y)$ per ogni coppia di elementi di X , cioè una funzione $d : X \times X \rightarrow \mathbb{R}$ che gode delle seguenti proprietà:

1. $d(x, y) = 0$ se $x = y$, $d(x, y) > 0$ se $x \neq y$ (positività)
2. $d(x, y) = d(y, x)$ (simmetria)
3. $d(x, y) \leq d(x, z) + d(z, y)$ (disuguaglianza triangolare)

In uno stesso insieme X si possono dare diverse definizioni di distanza. Ad esempio se $X = \mathbf{R}^n$ e se $A = (a_1, a_2, \dots, a_n)$, $B = (b_1, b_2, \dots, b_n)$ rappresentano 2 punti qualsiasi di \mathbf{R}^n , possiamo definire le seguenti 3 diverse distanze, che verificano tutte e tre le 3 proprietà precedenti:

$$d(A, B) = \left((a_1 - b_1)^2 + \dots + (a_n - b_n)^2 \right)^{1/2} \quad (\text{distanza euclidea})$$
$$d'(A, B) = \max(|a_1 - b_1|, \dots, |a_n - b_n|)$$
$$d''(A, B) = |a_1 - b_1| + \dots + |a_n - b_n|$$

Per il problema di costruire un albero evolutivo non orientato ci interessa la terza di queste distanze (d'') e ci interessa anche la definizione di distanza tra 2 vertici di un grafo.

Consideriamo un grafo non orientato G che sia *connesso*, cioè tale che si possa andare da un qualunque suo vertice ad un qualunque altro percorrendo un cammino formato da un numero finito di archi di G . Supponiamo inoltre che per ogni arco a di G sia definito un numero positivo $l(a)$ che ne rappresenta la *lunghezza*. Si può allora definire in modo naturale la *distanza* $d(A, B)$ di 2 vertici A, B del grafo come la lunghezza del cammino più corto che congiunge A con B (si può dimostrare che tale cammino esiste sempre). Questa distanza verifica le proprietà 1., 2., 3., di cui sopra e si potrebbe chiamare "distanza stradale" o "distanza topografica", interpretando i vertici come città e gli archi come strade.

A questo punto torniamo al problema di costruire un albero evolutivo per le 4 specie S_1, S_2, S_3, S_4 e i 5 caratteri C_1, C_2, C_3, C_4, C_5 esaminato nel paragrafo precedente :

	C_1	C_2	C_3	C_4	C_5
S_1	1	0	0	0	1
S_2	1	0	1	0	1
S_3	1	1	0	0	0
S_4	0	0	0	1	0

Usando la definizione $d''(A, B) = |a_1 - b_1| + \dots + |a_n - b_n|$ tra elementi di \mathbb{R}^n , possiamo costruire la seguente tabella delle distanze tra le 4 specie S_1, S_2, S_3, S_4 , viste come elementi di \mathbb{R}^5 :

	S_1	S_2	S_3	S_4
S_1	0	1	2	3
S_2	1	0	3	4
S_3	2	3	0	3
S_4	3	4	3	0

Osserviamo che queste distanze coincidono con le distanze del grafo dell'albero evolutivo costruito nel paragrafo precedente, se si suppone che ogni arco abbia lunghezza 1 : la distanza tra 2 specie coincide semplicemente col numero di caratteri per cui esse differiscono e si potrebbe chiamare *distanza evolutiva* (preciseró meglio questo concetto tra poco).

In ogni riga di questa matrice contrassegnamo il piú piccolo elemento diverso da 0 . Si vede subito che S_1 é la specie piú vicina a S_2 e S_2 é la piú vicina a S_1 , quindi cominciamo a raggruppare S_1 con S_2 , unendole con un segmento in \mathbb{R}^5 e sostituendole col punto medio M di questo segmento , cioè $M = (1, 0, 1/2, 0, 1)$. Abbiamo cosí i 3 punti M, S_3, S_4 , da cui si puó costruire la matrice delle distanze:

	M	S_3	S_4
M	0	5/2	7/2
S_3	5/2	0	3
S_4	7/2	3	0

Poi raggruppiamo M con S_3 e li sostituiamo con un unico punto N sul segmento MS_3 (N é il baricentro dei 3 punti S_1, S_2, S_3), ed infine uniamo N con S_4 . Otteniamo cosí il seguente schema di albero evolutivo: (vedi Fig. 2.1)

Restano da determinare le lunghezze dei lati x, y, z, u, v in modo che, in questo albero, le distanze tra le 4 specie S_1, S_2, S_3, S_4 siano proprio quelle originarie della matrice delle distanze, o per lo meno siano il piú vicino possibile ad esse (ottenere la perfetta uguaglianza non sempre é possibile). Si ha cioè il seguente sistema di 6 equazioni nelle 5 incognite x, y, z, u, v :

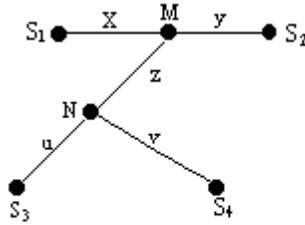


Figura 2.1

$$\begin{aligned}
 x + y &= 1 \\
 x + z + u &= 2 \\
 x + z + v &= 3 \\
 y + z + u &= 3 \\
 y + z + v &= 4 \\
 u + v &= 3
 \end{aligned}$$

in cui però la quinta equazione è conseguenza delle precedenti. Questo sistema ha l'unica soluzione $x = 0$, $y = z = u = 1$, $v = 2$, che dá luogo al seguente albero evolutivo (vedi Fig. 2.2).

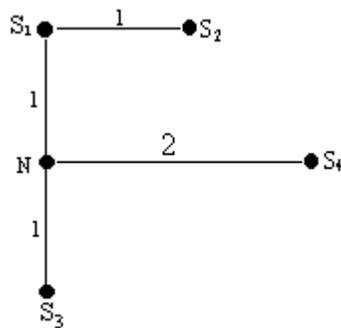


Figura 2.2

Questo albero coincide con quello trovato nel paragrafo precedente, ma non è orientato e non ha radice. Lo potremmo trasformare in un albero orientato ponendo la radice in un punto qualsiasi (in effetti nel paragrafo precedente avevamo trovato che il capostipite si trova nel punto medio del lato NS_4). Le lunghezze dei lati dell'albero rappresentano le distanze evolutive tra le

varie specie e quindi, ci possono dare un'idea (seppure vaga) del tempo che effettivamente é trascorso perché si verificasse l'evoluzione da una specie all'altra .

Questo procedimento puó essere generalizzato a qualsiasi matrice di caratteri qualitativi, tenendo presente che:

1. Qualora nella matrice dei caratteri compaiano dei numeri qualsiasi (e non soltanto 0 e 1) conviene prima normalizzare i valori di ciascun carattere dividendoli per il loro scarto quadratico medio, per evitare di dare piú importanza ad alcuni caratteri che ad altri e per avere dei risultati indipendenti dalle unitá di misura .
2. Qualora non sia possibile trovare le lunghezze dei lati dell 'albero evolutivo in modo da riprodurre esattamente la matrice delle distanze originaria (desunta dai valori dei caratteri), conviene determinare tali lunghezze in modo da minimizzare la somma dei valori $(1 - \frac{T_{ij}}{S_{ij}})^2$, dove gli indici i, j variano da 1 a m (numero delle specie in esame) , S_{ij} rappresenta la distanza sperimentale tra le 2 specie i, j (desunta dalla matrice dei caratteri) e T_{ij} rappresenta la distanza teorica tra le stesse specie, misurata lungo l'albero evolutivo.

Illustriamo tutto il procedimento con il seguente.

Esempio 2.1 Sia data la seguente matrice dei caratteri:

	C_1	C_2	C_3	C_4	C_5
S_1	1	0	3	0	0
S_2	4	1	0	0	2
S_3	0	0	1	1	0
S_4	0	2	0	1	5

Per prima cosa normalizziamo i valori di ciascun carattere dividendoli per il loro scarto quadratico medio e otteniamo cosí la nuova matrice dei caratteri:

	C_1	C_2	C_3	C_4	C_5
S_1	0.72	0	2.45	0	0
S_2	2.87	1.21	0	0	0.97
S_3	0	0	0.82	2	0
S_4	0	2.41	0	2	2.43

dalla quale si costruisce subito la seguente matrice delle distanze (1):

	S_1	S_2	S_3	S_4
S_1	0	6.88	4.35	10.01
S_2	6.88	0	7.87	7.53
S_3	4.35	7.87	0	5.66
S_4	10.01	7.53	5.66	0

da questa matrice si vede che S_1 é la specie piú vicino a S_3 e viceversa; cominciamo allora, col raggruppare S_1 con S_3 , sostituendole col punto medio $M = (0.36, 0, 1.64, 1, 0)$ del segmento S_1S_3 in \mathbf{R}^5 . Si ottiene allora, la seguente nuova matrice dei caratteri:

	C_1	C_2	C_3	C_4	C_5
M	0.36	0	1.64	0	0
S_2	2.87	1.21	0	0	0.97
S_4	0	2.41	0	2	2.43

E la seguente nuova matrice delle distanze:

	M	S_2	S_4
M	0	6.33	8.84
S_2	6.33	0	7.53
S_4	8.84	7.53	0

dalla quale si vede che M é il punto piú vicino a S_2 e viceversa; raggruppiamo allora M con S_2 , sostituendoli con un nuovo punto N e ottenendo il seguente albero evolutivo:

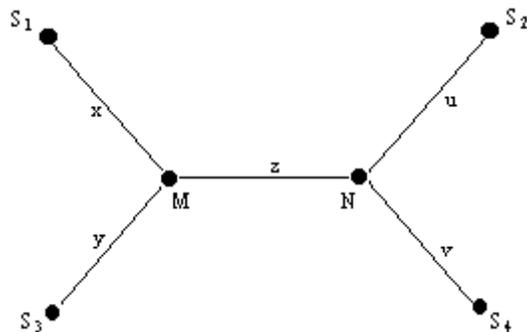


Figura 2.3

In questo albero rimangono da determinare le lunghezze dei 5 lati x, y, z, u, v in modo da riprodurre l'originaria matrice (1) delle distanze, o comunque da ottenere delle distanze il piú vicino possibile a quelle della matrice (1). Si vede subito che ottenere l'uguaglianza perfetta non é possibile, perché

equivale a risolvere il sistema di 6 equazioni in 5 incognite:

$$\begin{aligned}x + y &= 4.35 \\x + z + v &= 10.01 \\y + z + u &= 7.87 \\u + v &= 7.53 \\y + z + v &= 5.66\end{aligned}$$

e questo sistema non ha soluzioni, in quanto sottraendo a membro a membro la terza equazione dalla prima si ottiene $u - v = -3.13$, mentre sottraendo la sesta dalla quarta si ottiene $u - v = 2.21$. Cerchiamo allora, i valori di x, y, z, u, v che rendono minima la seguente funzione:

$$\begin{aligned}f(x, y, z, u, v) &= \left(1 - \frac{(x + z + u)}{6.88}\right)^2 + \left(1 - \frac{(x + y)}{4.35}\right)^2 + \left(1 - \frac{(x + z + v)}{10.01}\right)^2 \\&+ \left(1 - \frac{(y + z + u)}{7.87}\right)^2 + \left(1 - \frac{(u + v)}{7.53}\right)^2 + \left(1 - \frac{(y + z + v)}{5.66}\right)^2\end{aligned}$$

Questo si ottiene ponendo uguali a zero le derivate parziali f_x, f_y, f_z, f_u, f_v di f rispetto a x, y, z, u, v e ciò porta al seguente sistema lineare di 5 equazioni nelle 5 incognite x, y, z, u, v :

$$\begin{aligned}x + y &= 4.35 \\u + v &= 7.53 \\0,82x - 1.22y + 0.4z + 0.82u - 1.22v &= -1.22 \\1.45x + 2.14z + 1.65u + 0.69v &= 16.89 \\0.79x - 1.27y - 0.48z - 1.27u + 0.79v &= -2.14\end{aligned}$$

che ha l'unica soluzione $x=2.33, y=2.02, z=1.33, u=3.27, v=4.26$. Da questi valori si deduce che l'originale matrice delle distanze (1), viene sostituita dalla seguente matrice delle distanze (2), misurate lungo l'albero evolutivo:

	S_1	S_2	S_3	S_4
S_1	0	6.93	4.35	7.92
S_2	6.93	0	6.62	7.53
S_3	4.35	6.62	0	7.61
S_4	7.92	7.53	7.61	0

■

2.1 Precisioni sul concetto di distanza evolutiva tra specie (Lattes)

Il modo migliore di definire la distanza evolutiva tra 2 specie, naturalmente, è quello di usare tecniche di sistematica molecolare, cioè analizzare la sequenza

del DNA.

Consideriamo un certo gene G del DNA il quale in una certa popolazione di esseri viventi può comparire con un certo numero n di *alleli* diversi A_1, A_2, \dots, A_n (esempi: biondo - bruno, occhi azzurri - occhi neri, gruppo sanguigno A - B - O eccetera).

Consideriamo inoltre 2 diverse specie (o popolazioni) A,B di esseri viventi e supponiamo che gli alleli A_1, A_2, \dots, A_n compaiano in A con le frequenze (o probabilità p_1, \dots, p_n e in B con le frequenze q_1, \dots, q_n . Vogliamo definire la “distanza genetica” tra A e B per quanto riguarda il gene G .

Essa può essere stabilita sperimentalmente prelevando a caso il gene G da un individuo della specie A e da un individuo della specie B e vedendo qual é la probabilità p_{AB} di ottenere identità di sequenza, cioè lo stesso allele (con tecniche di ibridazione molecolare) . Questa probabilità chiaramente é data da:

$$p_{AB} = p_1q_1 + p_2q_2 + \dots + p_nq_n$$

p_{AB} é un numero compreso tra 0 e 1 ed é tanto piú alta quanto piú A e B sono simili. Però accade anche se A e B sono identiche (cioé $p_i=q_i$ per ogni i da 1 a n), p_{AB} non vale 1, bensí vale $p_1^2+p_2^2+\dots+p_n^2$. Per rimediare a questo inconveniente si può definire il *coefficiente di identità normalizzato* z (introdotto da Nei nel 1972):

$$z = p_{AB}/(p_{APB})^{1/2} = (p_1q_1 + \dots + p_nq_n)/\left((p_1^2 + \dots + p_n^2)(q_1^2 + \dots + q_n^2)\right)^{1/2}$$

Dove $p_A=p_1^2+\dots+p_n^2$ ' e la probabilità di ottenere identità di sequenza prelevando a caso il gene G da 2 individui della specie A e $p_B=q_1^2+\dots+q_n^2$ é l'analoga probabilità per la specie B. Interpretando geometricamente $\mathbf{P} = (p_1, p_2, \dots, p_n)$ e $\mathbf{Q} = (q_1, q_2, \dots, q_n)$ come vettori di \mathbb{R}^n , si ha che $z = \frac{\mathbf{P} \cdot \mathbf{Q}}{|\mathbf{P}||\mathbf{Q}|} = \cos \Theta$, dove Θ é l'angolo formato dai 2 vettori \mathbf{P} e \mathbf{Q} .

Quindi, z é un numero compreso tra 0 e 1(non può essere negativo perché i 2 vettori \mathbf{P} e \mathbf{Q} hanno tutte le componenti positive. z vale 1 se e solo se $\mathbf{P} = \mathbf{Q}$ (cioé le 2 specie presentano la stessa frequenza per tutti gli alleli), z vale 0 se e solo se i 2 vettori \mathbf{P} e \mathbf{Q} sono ortogonali, il che si verifica se e solo se le 2 specie A e B sono ” complementari“, cioè non hanno nessun allele in comune.

Questa misura di somiglianza (identità genetica) che varia tra 1 e 0, può essere trasformata in una misura di diversità, o distanza, che varia tra 0 e infinito, ponendo $D = -\log z$ (logaritmo naturale in base $e = 2.71828 \dots$). Il numero D si chiama *distanza genetica* (o distanza evolutiva)tra A e B (per quanto riguarda il gene G).

Piú in generale, si può definire la distanza genetica tra A e B tenendo conto di piú geni G_1, G_2, \dots, G_m (in teoria di tutti i geni del DNA). A questo scopo Nei ha introdotto il *coefficiente di identità normalizzato medio* z^* :

$$z^* = p^*_{AB} / (p^*_A p^*_B)^{1/2}$$

dove p^*_{AB} , p^*_A e p^*_b indicano rispettivamente i valori medi di p_{AB} , p_A e p_B per gli m geni G_1, G_2, \dots, G_m . Si definisce allora la *distanza genetica* $D(A, B)$ tra le specie A e B come $D(A, B) = -\log z^*$. Tale misura é valida quando il tasso di mutazione é lo stesso per tutti i geni.

Nei osserva che quando 2 popolazioni si separano, in esse possono accumularsi indipendentemente nuove mutazioni. Nell'ipotesi che queste mutazioni siano neutre (cioé non siano né favorite né sfavorite dalla selezione naturale) e avvengano con un tasso costante nel tempo, la distanza genetica D sarà proporzionale al tempo T trascorso dalla separazione fino ad oggi. Però le dimensioni delle popolazioni determinano fenomeni casuali di variazioni delle frequenze dei vari alleli (deriva genetica), perciò bisogna porre anche la condizione che le due popolazioni abbiano le stesse dimensioni. Se, infine, il *tasso di mutazione* K é lo stesso nelle 2 popolazioni, si può dimostrare che la distanza genetica D é data da :

$$D = 2KT$$

Nei ha stimato il tasso di mutazione K in 10^{-7} mutazioni / anno circa, per cui il tempo di divergenza T tra le 2 popolazioni può essere stimato da $T = D \times 5$ milioni di anni. Ciò significa che un valore di distanza genetica $D=1$ vuol dire che le 2 popolazioni si sono separate da circa 5 milioni di anni.

3 Grafi: definizioni e proprietà

Intuitivamente, un *grafo* é semplicemente un insieme di punti (o vertici) e di archi (o lati) che congiungono fra loro, a due a due, alcuni di questi vertici. Piú precisamente, un grafo consiste in un insieme finito V di vertici, in un insieme finito A di archi ed in una funzione f che ad ogni arco associa una coppia non ordinata di vertici (i 2 vertici congiunti da quell'arco). Consideriamo *multigrafi non orientati*: multigrafi significa che tra 2 vertici ci puó essere anche piú di un arco, non orientati significa che il verso di percorrenza degli archi non ha importanza. Vediamo adesso alcune definizioni e proprietà molto semplici riguardanti i grafi:

Definizione 3.1 *un cammino di un grafo é una successione finita di archi tale che ogni arco abbia in comune un vertice con l'arco precedente e l'altro vertice con l'arco successivo.*

Definizione 3.2 *due vertici di un grafo sono collegati se esiste un cammino che li unisce.*

Definizione 3.3 *un grafo si dice connesso se in esso due vertici qualsiasi sono collegati.*

Definizione 3.4 *dato un vertice A di un grafo G , la componente connessa di G contenente A é l'insieme dei vertici di G collegati con A (compreso A stesso) e degli archi di G che congiungono questi vertici.*

Teorema 3.1 *Due diverse componenti connesse di un grafo sono disgiunte. Ogni grafo si puó decomporre in un numero finito di componenti connesse. Un grafo é connesso se e solo se ha una sola componente connessa. La dimostrazione é immediata.*

Definizione 3.5 *Un cammino semplice é un cammino formato da archi tutti distinti.*

Definizione 3.6 *Un ciclo é un cammino semplice in cui il primo vertice coincide con l'ultimo.*

Definizione 3.7 *Un grafo aciclico é un grafo privo di cicli.*

Definizione 3.8 *L'indice ciclico di un grafo é il numero minimo di archi che bisogna togliergli per renderlo aciclico. L'indice ciclico, in un certo senso, é il numero di "cicli indipendenti" del grafo. Ad esempio, i 4 grafi della figura 3.1 hanno rispettivamente, 1,2,3,7 cicli, ma l'indice ciclico é rispettivamente 1, 2, 2, 3.*

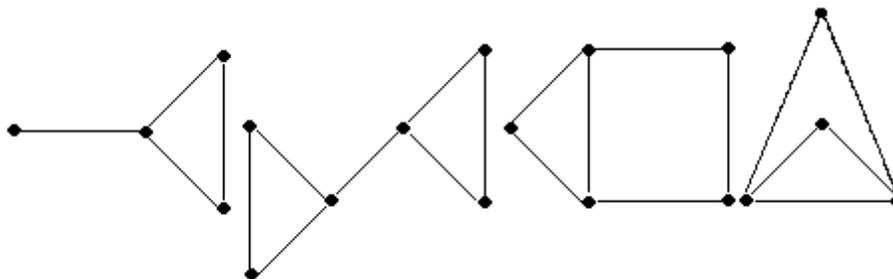


Figura 3.1

Definizione 3.9 *Un albero é un grafo aciclico connesso; una foresta 'e un grafo aciclico non connesso (le componenti connesse di una foresta sono alberi).*

Teorema 3.2 *L'indice ciclico di un grafo si puó calcolare semplicemente contando il numero di vertici, archi e componenti connesse del grafo. Per la precisione, detti v , a , p , c rispettivamente il numero di vertici, archi, componenti connesse e l'indice ciclico, si ha $v + c = a + p$. (Ad esempio, per i 2 grafi di figura 3.2 l'indice ciclico vale rispettivamente $10 + 1 - 5 = 6$ e $9 + 1 - 6 = 4$, il che non appare evidente a prima vista dal disegno).*

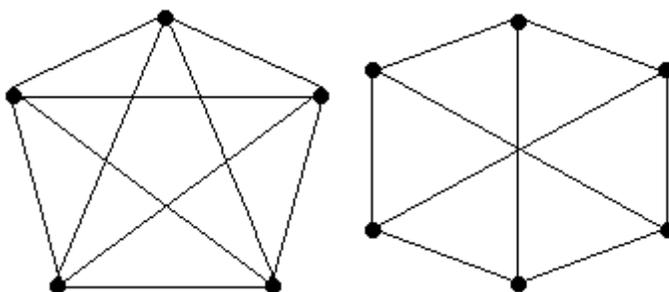


Figura 3.2

Dimostrazione. Sia G un grafo, v il numero dei suoi vertici e G' il grafo che si ottiene da G togliendogli tutti gli archi e lasciando solo i vertici isolati. Per il grafo G' si ha $v + c = a + p = v$, in quanto $c = 0, a = 0, p = v$. Aggiungiamo ora ad uno ad uno gli archi di G a G' , fino ad ottenere nuovamente G . Al generico passo di questo procedimento, possono verificarsi 2 casi : o il nuovo arco aggiunto congiunge 2 vertici che si trovano nella stessa componente connessa, oppure in 2 componenti connesse diverse. Nel

primo caso v, p rimangono uguali, a, c aumentano di 1, quindi, i 2 numeri $v + c, a + p$ aumentano entrambi di 1. Nel secondo caso v, c rimangono uguali, a aumenta di 1, p diminuisce di 1, quindi, i 2 numeri $v + c, a + p$ rimangono uguali. Siccome i 2 numeri $v + c, a + p$ erano uguali per G' , essi restano uguali durante tutto il procedimento e quindi, sono uguali anche per G , come volevasi dimostrare. ■

Dal teorema(3.2) seguono immediatamente questi altri 2 teoremi:

Teorema 3.3 *In un grafo connesso ($p = 1$) l'indice ciclico é dato da $c = a - v + 1$, cioé é uguale al numero di archi meno il numero di vertici piú 1.*

Teorema 3.4 *In un albero ($p = 1, c = 0$) si ha $v = a + 1$ (il numero di vertici é uguale al numero di archi piú 1)*

Definizione 3.10 *Il grado di un vertice di un grafo é il numero di archi aventi quel vertice come estremo.*

Teorema 3.5 *In ogni grafo il numero dei vertici di grado dispari é pari.*

Dimostrazione. Siano v, a i numeri di vertici e di archi del grafo. É chiaro che il grado massimo possibile di un vertice é a . Indichiamo con v_1, v_2, \dots, v_a il numero dei vertici di grado 1, 2, \dots, a rispettivamente. Facendo la somma dei gradi di tutti i vertici otteniamo $v_1 + 2v_2 + 3v_3 + \dots + av_a$ e questa somma chiaramente é uguale a $2a$ (perché in tal modo ogni arco viene contato 2 volte). Ma allora, il numero totale di vertici di grado dispari é dato da:

$$v_1 + v_3 + v_5 + \dots = 2a - 2v_2 - 2v_3 - 4v_4 - 4v_5 - \dots$$

e quindi, é pari, come volevasi dimostrare. ■

Esempio 3.1 Nelle molecole degli idrocarburi (composti chimici contenenti solo carbonio e idrogeno) il numero di atomi di idrogeno deve essere pari, ad esempio, non puó esistere un idrocarburo di formula C_2H_3 . Infatti, nel *grafo chimico* (grafo i cui vertici corrispondono agli atomi e i cui lati corrispondono ai legami fra gli atomi) gli atomi di carbonio sono vertici di grado 4 e gli atomi di idrogeno sono vertici di grado 1. ■

Definizione 3.11 *In un grafo un vertice di grado 3 si chiama anche ramificazione semplice, un vertice di grado 4 si chiama anche ramificazione doppia, un vertice di grado 5 si chiama anche ramificazione tripla e cosí via. Definiamo allora, il numero di ramificazioni r di un grafo con a archi come $r = v_3 + 2v_4 + 3v_5 + \dots + (a - 2)v_a$. Se il grafo ha solo vertici di grado 1 e 2 si ha $r = 0$.*

I seguenti 3 teoremi sono immediati:

Teorema 3.6 *Il numero di ramificazioni di un grafo é uguale al numero dei vertici di grado 1 piú il doppio della differenza tra numero di archi e numero di vertici. In simboli, $r = v_1 + 2a - 2v$ (infatti $v_1 + 2a - 2v = v_1 + (v_1 + 2v_2 + 3v_3 + 4v_4 + \dots) - 2(v_1 + v_2 + v_3 + v_4 + \dots) = v_3 + 2v_4 + 3v_5 + \dots = r$).*

Teorema 3.7 *In un grafo connesso il numero di ramificazioni é uguale al numero dei vertici di grado 1 piú il doppio dell'indice ciclico meno 2. In simboli, $r = v_1 + 2c - 2$ (infatti, per il teorema 3.3, si ha $c = a + 1 - v$ e quindi, $r = v_1 + 2(a - v) = v_1 + 2(c - 1) = v_1 + 2c - 2$).*

Teorema 3.8 *In un albero il numero di ramificazioni é uguale al numero di vertici di grado 1 (foglie) meno 2 (infatti, un albero é un grafo connesso con $c = 0$ e quindi, per il teorema 3.7, $r = v_1 - 2$).*

Definizione 3.12 *Se ad ogni lato di un grafo connesso G é associato un numero reale positivo che ne rappresenta la lunghezza, si chiama distanza tra 2 vertici A e B di G la lunghezza del piú breve cammino di G che li congiunge (si dimostra facilmente che tale cammino esiste sempre).*

3.1 Topologia molecolare (Da un articolo di “Le Scienze”, n. 219, novembre 1986)

I chimici finora hanno studiato le proprietà di oltre 7 milioni di molecole diverse, sia naturali che artificiali (cioé sintetizzate in laboratorio). É possibile sfruttare ancora piú a fondo queste conoscenze, per esempio servirsi di esse per prevedere le proprietà di sostanze chimiche prima ancora che esse siano sintetizzate, cioè prima che anche una sola delle loro molecole esista fisicamente? Questa interessante prospettiva sta diventando realtà grazie a una nuova tecnica che, sebbene sia ancora agli inizi, ha già raggiunto molti successi: *la topologia molecolare*.

Ai fini della topologia molecolare la forma tridimensionale di una molecola non ha importanza, e non hanno importanza neanche la natura e la lunghezza dei legami chimici tra gli atomi, gli angoli tra i legami e talvolta perfino i tipi di atomi che formano la molecola. Ciò che importa é quanti atomi contiene la molecola, come ogni atomo é legato agli atomi vicini e se gli atomi sono uniti a formare una singola catena aperta, una catena ramificata, una o piú catene chiuse (anelli o cicli), o una combinazione di anelli, catene aperte e ramificazioni. Conta cioè la “formula di struttura” della molecola, a prescindere però dalla lunghezza dei legami e dagli angoli fra essi (la perdita di informazione che si ha in tal modo potrebbe sembrare grave, ma in pratica spesso si rivela meno importante di quanto sembra).

Gli strumenti piú importanti per fare previsioni chimiche col metodo topologico sono i cosiddetti *indici topologici*, cioè dei numeri che si possono calcolare conoscendo soltanto la struttura topologica della molecola. 4 esempi particolarmente semplici e importanti di indici topologici sono: il numero di atomi, il numero di legami, il numero di cicli indipendenti, il numero di ramificazioni.

Quando si vuole fare una previsione chimica basata sulla topologia molecolare, il primo passo consiste nell'applicare un certo indice X opportunamente scelto a un numero relativamente piccolo di molecole ben conosciute. Si riportano in un grafico i valori di X per le varie molecole sull'asse delle ascisse, mentre sull'asse delle ordinate si riportano i valori di una certa proprietà fisico - chimica Y delle molecole, che interessa studiare (per esempio la temperatura di ebollizione). Ciascuna molecola é rappresentata da un punto nel grafico. Se esiste una curva che collega abbastanza bene i vari punti, questa può servire per prevedere i valori di Y per molecole ancora sconosciute. La chiave del metodo consiste nel trovare l'indice che abbia la migliore correlazione possibile con la proprietà Y (ottenere correlazioni perfette e' impossibile, perche' i risultati sperimentali sono sempre soggetti a errori di una certa entità). Questo metodo ha trovato applicazioni molto importanti, fra cui: lo studio della corrosione, dell'attività di molti farmaci, del potere cancerogeno delle molecole, della velocità di diffusione delle sostanze inquinanti nell'ambiente.

Per analizzare la topologia di una molecola si parte da un disegno in cui gli atomi della molecola sono rappresentati da punti e i legami che li uniscono da segmenti: un cosiddetto *grafo chimico*. Nel grafo di una molecola spesso gli atomi di idrogeno si omettono, perché di solito essi non hanno una grande importanza per determinare le proprietà della molecola: in tal modo si ottiene il cosiddetto *grafo ridotto* della molecola.

Ogni numero che possa essere calcolato solo in base alla conoscenza del grafo di una molecola è un indice topologico della molecola stessa. Uno dei primi indici topologici presi in considerazione dai chimici, per quanto riguarda le molecole degli idrocarburi (composti contenenti solo atomi di carbonio e idrogeno) è il numero di atomi di carbonio, che non è altro che il numero dei vertici del grafo ridotto. I chimici hanno trovato molte proprietà delle molecole che sono correlate col numero di atomi di carbonio, senza rendersi conto di usare in effetti uno dei più semplici indici topologici della molecola. Il numero di atomi di carbonio è un indice adatto per analizzare le molecole a catena aperta non ramificata, ma non è altrettanto adatto per le molecole contenenti anelli o ramificazioni, che possono assumere forme diversissime pur avendo lo stesso numero di atomi di carbonio. È quindi necessario inventare altri indici che tengano conto anche della forma della molecola, dei cicli e delle ramificazioni.

Il primo indice topologico capace di caratterizzare la ramificazione delle molecole è stato l'*indice di Wiener*, proposto nel 1947 e basato sulla nozione di *distanza in un grafo*: la distanza tra 2 vertici A e B di un grafo è uguale al numero di lati che si devono attraversare per andare da A a B percorrendo il cammino più breve possibile attraverso il grafo. L'indice di Wiener di una molecola è dato dalla somma delle distanze tra tutte le coppie di atomi della molecola.

L'indice di Wiener di una molecola di idrocarburo si comporta in modo simile al numero di atomi di carbonio, in quanto in genere è maggiore per le molecole più grandi, ma fornisce anche una certa misura della ramificazione della molecola: a parità di numero di atomi di carbonio, esso è maggiore per le molecole più estese e minore per le molecole più compatte e ramificate. Esso ha una correlazione molto buona con svariate proprietà: la temperatura di fusione e di ebollizione, la viscosità, la tensione superficiale, l'indice di rifrazione, la conducibilità elettrica, il potere ossidante o riducente, la lunghezza d'onda della luce assorbita, la stabilità chimica e la reattività. Molte di queste correlazioni si spiegano perché l'indice di Wiener ha un'ottima correlazione con l'energia totale degli elettroni di legame della molecola (almeno per certi tipi di idrocarburi complessi, come gli spirani e gli idrocarburi aromatici policiclici). Più alto è l'indice di Wiener, maggiore è l'energia totale degli elettroni di legame. Perciò le molecole con indice di Wiener basso (molecole compatte, molto ramificate) hanno un'energia minore delle molecole con indice di Wiener alto (molecole estese), quindi sono più stabili e tendono a formarsi più facilmente in molti processi chimici.

L'indice di Wiener fornisce anzitutto una misura del volume di una molecola, pur dando anche qualche indicazione sulla forma. Un altro indice piú sensibile alla forma della molecola é l'*indice di Randic*, o *indice di connettività molecolare*, che é correlato con molte proprietá chimiche importanti, ancor piú dell'indice di Wiener.

L'indice di Randic si basa sulla nozione topologica di *grado di un vertice di un grafo*, che é semplicemente il numero di lati che partono da quel vertice. Per calcolare l'indice di Randic di un grafo si attribuisce a ogni lato del grafo un valore, che é uguale al reciproco della radice quadrata del prodotto dei gradi dei 2 vertici che il lato congiunge. L'indice di Randic di un grafo é allora, la somma dei valori di tutti i lati del grafo.

Alcune proprietá chimiche presentano correlazioni migliori con l'indice di Randic relativo non a tutta la molecola, ma a certe parti della molecola aventi una struttura particolare (sempre deducibile dal grafo della molecola). Non si ha quindi un solo indice di Randic, ma vari indici di Randic relativi a sottostrutture molecolari diverse, ciascuno dei quali é applicabile in casi diversi. Questi indici presentano ottime correlazioni con un gran numero di proprietá fisiche, chimiche e biologiche delle molecole, fra cui: la densitá, la solubilitá in acqua, il calore latente di evaporazione, l'odore, il sapore, la tossicitá, la velocitá di diffusione nell'ambiente, la capacitá di fungere da disinfettanti, anestetici, narcotici, allucinogeni, la capacitá di provocare mutazioni genetiche.

Quasi tutte le proprietá chimiche delle molecole sono strettamente correlate con qualche indice topologico, ma naturalmente in ogni singolo caso occorre trovare l'indice adatto, che sia correlato nel modo migliore con la specifica proprietá in esame. Facciamo un esempio, che riguarda le molecole degli idrocarburi: quando un idrocarburo viene usato come combustibile, due proprietá importanti della molecola sono la quantitá di fuliggine prodotta e il cosiddetto "numero di ottano" (che misura l'efficienza con cui il combustibile brucia senza detonare). Queste 2 proprietá sono correlate molto bene con 2 indici topologici particolarmente semplici: la quantitá di fuliggine prodotta con l'indice ciclico (numero di cicli indipendenti) e il numero di ottano col numero di ramificazioni.

Un progetto molto ambizioso riguardante gli indici topologici é quello di usarli per prevedere il *potere cancerogeno* delle molecole. Alcuni risultati abbastanza buoni sono stati ottenuti da Herndon e Szentpaly, che con una combinazione di indici semplici sono riusciti a prevedere il potere cancerogeno degli idrocarburi aromatici policiclici.

Risulta che solo le molecole aventi dimensioni comprese entro certi limiti tendono ad essere cancerogene: quelle piú grandi o piú piccole non lo sono. Inoltre certe regioni di una molecola sono piú importanti di altre, per quanto riguarda le proprietá cancerogene. Queste regioni sono note come regioni a insenatura e regioni *K*, *L*, *M* (vedi fig.3.3) e hanno un ruolo di rilievo nelle fasi iniziali della cancerogenesi. Affinché la molecola sia cancerogena le

regioni a insenatura e le regioni K e M devono essere chimicamente attive e le regioni L chimicamente inattive. Herndon e Szentpaly hanno combinato insieme 4 indici topologici (2 dei quali riguardano le dimensioni della molecola, mentre gli altri 2 riguardano certe caratteristiche strutturali delle regioni a insenatura e delle regioni K , L e M) in una formula che fornisce un quinto indice, correlato molto bene con le misure sperimentali del potere cancerogeno.

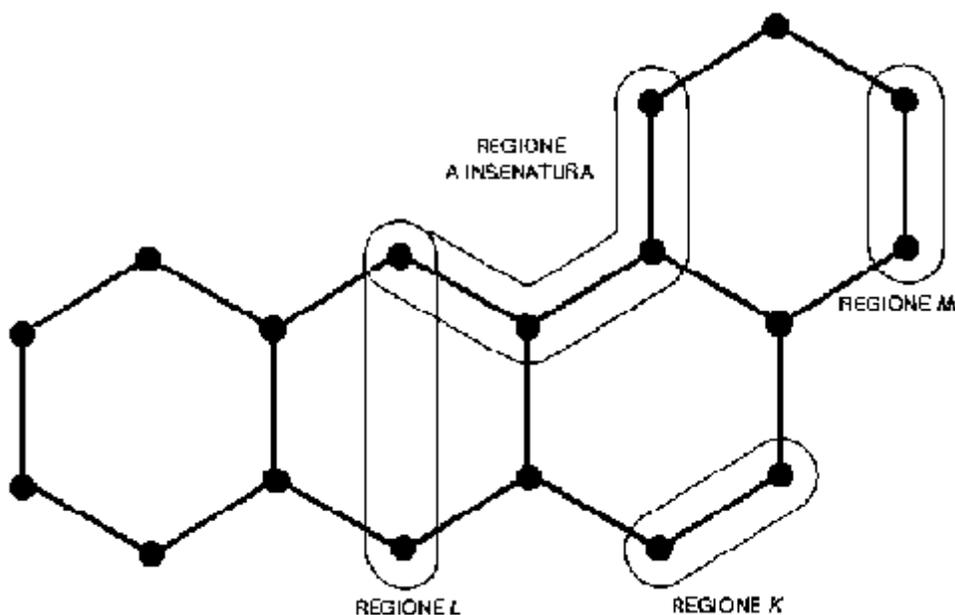


Figura 3.3

3.2 Grafi e molecole

La *formula di struttura* di una molecola può essere rappresentata con un grafo in cui i vertici rappresentano gli atomi e i lati rappresentano i legami. Si tratta di un grafo coi vertici etichettati, perché ad ogni vertice è associata un'etichetta (il simbolo dell'elemento chimico) cioè è definita una funzione $f : V \rightarrow X$, dove $V =$ insieme dei vertici, $X =$ insieme dei 92 elementi chimici.

Limitiamoci a considerare gli idrocarburi, cioè i composti contenenti solo carbonio (C) e idrogeno (H). La formula di struttura di un idrocarburo può essere rappresentata sia col *grafo esteso* che col *grafo ridotto*. Il grafo esteso comprende vertici di grado 4 (atomi di C) e di grado 1 (atomi di

H). Il grafo ridotto si ottiene dal grafo esteso “potando” i vertici di grado 1 e i lati che partono da essi e contiene vertici di grado 1 (gruppi $-CH_3$), 2 (gruppi $-CH_2-$), 3 (gruppi $-CH=$) e 4 (atomi di C).

Vedi gli esempi riportati di seguito

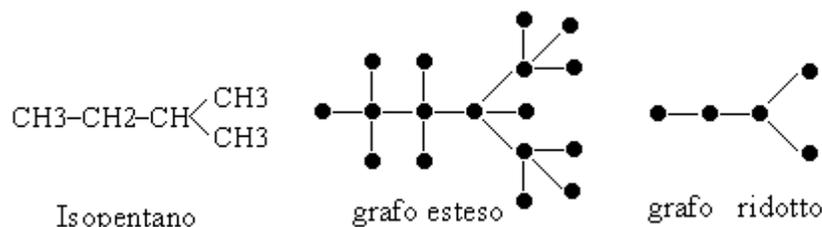


Figura 3.4

Nel grafo ridotto usiamo le seguenti notazioni:

- v = numero vertici = numero di atomi di C
- v_1, v_2, v_3, v_4 = numero vertici di grado 1,2,3,4
- a = numero lati = numero legami $C - C$ (i legami doppi o tripli contano 2 o 3)
- h = numero atomi di H
- r = numero di ramificazioni
- c = indice ciclico

r, c sono importanti per sapere se la molecola é lineare o ramificata, ciclica o aciclica. Da queste definizioni si ricava subito che :

$$\begin{aligned} v &= v_1 + v_2 + v_3 + v_4 \\ 2a &= v_1 + 2v_2 + 3v_3 + 4v_4 \\ h &= 3v_1 + 2v_2 + v_3 \\ r &= v_3 + 2v_4 \end{aligned}$$

Inoltre, dai teoremi visti nelle pagine precedenti si ricava subito che se un idrocarburo ha formula bruta $C_v H_h$, allora si ha $a = 2v - h/2$, $c = v + 1 - h/2$, $r = 2v - h + v_1$. Il numero di atomi di idrogeno deve essere per forza un numero pari compreso tra 0 e $2v + 2$. Conoscendo la formula bruta di un idrocarburo si possono calcolare subito il numero di legami $C - C$ e l'indice

ciclico, ma non il numero di ramificazioni (per calcolare quest'ultimo bisogna conoscere anche il numero v_1 di gruppi $-CH_3$).

Ad esempio, gli unici idrocarburi possibili con 5 atomi di carbonio (in base a considerazioni puramente topologiche) sono C_5H_{12} , C_5H_{10} , C_5H_8 , C_5H_6 , C_5H_4 , C_5H_2 . Queste molecole hanno rispettivamente indice ciclico 0, 1, 2, 3, 4, 5 (osserviamo che un legame doppio equivale ad un ciclo, un legame triplo a 2 cicli indipendenti). Una molecola di idrocarburo con v atomi di C deve avere indice ciclico compreso tra 0 e v . Gli idrocarburi aciclici hanno formula bruta C_vH_{2v+2} , alla quale corrispondono tante diverse formule di struttura quanti sono i diversi alberi non isomorfi con v vertici (per $v = 1, 2, 3, 4, 5, 6, 7, \dots$ il numero delle formule di struttura possibili é 1, 1, 1, 2, 3, 6, 12, \dots). Gli idrocarburi ciclici, con indice ciclico c , hanno formula bruta $C_vH_{2v+2-2c}$.

Mediante il grafo di una molecola si possono definire diversi indici topologici che spesso presentano buone correlazioni con le proprietà chimiche della molecola. Per stabilire queste correlazioni si può usare ad esempio il metodo della *correlazione lineare*.

In generale, supponiamo di avere N molecole diverse e di conoscere per esse i valori X_1, X_2, \dots, X_n di un certo indice topologico X e i valori Y_1, Y_2, \dots, Y_n di una certa proprietà chimica Y . Riportiamo in un grafico i valori di X sull'asse delle ascisse e i valori di Y sull'asse delle ordinate, in modo che ogni molecola sia rappresentata da un punto sul grafico (fig.3.5).

Data una generica retta $y = Ax + B$, poniamo $Z_i = Ax_i + B$ e $R_i = Y_i - Z_i$ (fig.3.5). Se supponiamo che la relazione lineare $y = Ax + B$ rappresenti abbastanza bene la dipendenza di Y da X , i *residui* $R_i = Y_i - Z_i = Y_i - AX_i - B$ devono essere abbastanza piccoli. Si definisce *retta di regressione* dei punti considerati la retta che rende minima la somma dei quadrati dei residui, cioè rende minima la funzione $S(A, B)$ definita da:

$$S(A, B) = \sum (Y_i - AX_i - B)^2 \quad (\text{la somma s'intende per } i \text{ che va da } 1 \text{ a } N)$$

Per trovare i valori di A, B , che rendono minima $S(A, B)$ conviene uguagliare a zero le derivate parziali $\delta S/\delta A, \delta S/\delta B$:

$$\begin{aligned} S(A, B) &= \sum Y_i^2 + A^2 \sum X_i^2 + NB^2 - 2A \sum X_i Y_i - 2B \sum Y_i + 2AB \sum X_i = \\ &= \sum (Y_i^2 + A^2 X_i^2 + B^2 - 2AX_i Y_i - 2BY_i + 2ABX_i) \end{aligned}$$

$$\frac{\delta S}{\delta A} = 2A \sum X_i^2 + 2B \sum X_i - 2 \sum X_i Y_i = 0$$

$$\frac{\delta S}{\delta B} = 2A \sum X_i + 2NB - 2 \sum Y_i = 0$$

Questo é un sistema lineare di di 2 equazioni nelle 2 incognite A, B , che ha l'unica soluzione :

$$A = (N \sum X_i Y_i - \sum X_i \sum Y_i) / (N \sum X_i^2 - (\sum X_i)^2)$$

$$B = (\sum Y_i - A \sum X_i) / N$$

Quindi, la funzione $S(A, B)$ ha un unico punto in cui potrebbe trovarsi un massimo relativo o un minimo relativo ; siccome però si tratta di una funzione continua, definita su tutto R^2 e che tende a piú infinito quando la distanza del punto (A, B) dall'origine tende a piú infinito (cioé quando $|A|$ oppure $|B|$ diventano grandi a piacere), per il teorema di Weierstrass sulle funzioni di 2 variabili il punto trovato deve essere per forza un punto di minimo assoluto. (Vedi Fig. 3.5).

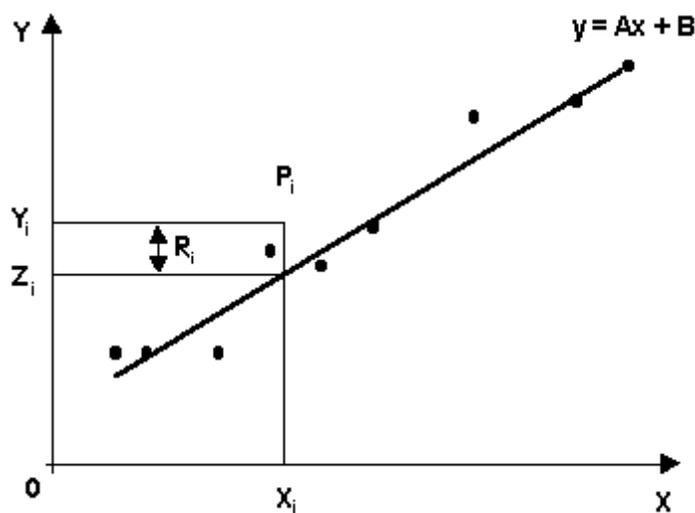


Figure 3.5

I valori trovati di A, B per la retta di regressione possono essere espressi in modo piú comodo introducendo le medie $M(X)$ e $M(Y)$ e le varianze $V(X)$ e $V(Y)$ delle 2 variabili X e Y nonché la covarianza $C(X, Y)$ tra X e Y . Le

definizioni di $M(X)$, $M(Y)$, $V(X)$, $V(Y)$, $C(X, Y)$ sono le seguenti:

$$\begin{aligned} M(X) &= (\sum X_j)/N & M(Y) &= (\sum Y_j)/N \\ V(X) &= M((X - M(X))^2) = (\sum X_i - M(X))^2/N \\ V(Y) &= M((Y - M(Y))^2) = (\sum Y_i - M(Y))^2/N \\ V(Y) &= M((Y - M(Y))^2) = (\sum Y_i - M(Y))^2/N \\ C(X, Y) &= M((X - M(X))(Y - M(Y))) = (\sum (X_i - M(X))(Y_i - M(Y)))/N \end{aligned}$$

si dimostra facilmente che $V(X)$, $V(Y)$, $C(X, Y)$ si possono esprimere anche così:

$$\begin{aligned} V(X) &= M(X^2) - (M(X))^2 & V(Y) &= M(Y^2) - (M(Y))^2 \\ C(X, Y) &= M(XY) - M(X)M(Y) \end{aligned}$$

e che valgono le seguenti proprietà (dove con k indichiamo un numero reale qualsiasi):

$$\begin{aligned} M(X + Y) &= M(X) + M(Y) \\ M(kX) &= kM(X) \\ V(X + Y) &= V(X) + V(Y) + 2C(X, Y) \\ V(X - Y) &= V(X) + V(Y) - 2C(X, Y) \\ V(kX) &= k^2V(X) \\ C(X, X) &= V(X) \\ C(X, Y) &= C(Y, X) \\ C(X, kY) &= kC(X, Y) \\ C(X, Y + Z) &= C(Y, X) + C(X, Z) \\ C(X, k) &= 0 \end{aligned}$$

Mediante queste proprietà è facile vedere che i coefficienti A , B della retta di regressione si possono esprimere così:

$$A = C(X, Y)/V(X) \quad B = M(Y) - AM(X)$$

Dal fatto che $B = M(Y) - AM(X)$ si vede che la retta di regressione passa per il punto $G = (M(X), M(Y))$, che rappresenta il baricentro della “nuvola” degli N punti $P_i = (X_i, Y_i)$.

Data una generica retta $Y = Ax + B$ e posto $Z_i = AX_i + B$, $R_i = Y_i - Z_i$, si ha :

$$V(Y) = V(Z + R) = V(Z) + V(R) + 2C(Z, R)$$

vogliamo vedere in quali casi succede che la varianza dei dati $V(Y)$ é uguale alla somma della varianza dei residui $V(R)$ e della varianza $V(Z)$ delle ordinate Z_i dei punti della retta di ascissa X_i . Questo succede se e solo se $C(Z, R) = 0$, cioè:

$$\begin{aligned} C(Z, Y - Z) &= 0 & C(Z, Y) - C(Z, Z) &= 0 \\ C(AX + B, Y) &= V(Z) = V(AX + B) = V(AX) = A^2V(X) \\ AC(X, Y) + C(B, Y) &= AC(X, Y) = A^2V(X) \\ A(C(X, Y)) - AV(X) &= 0 \end{aligned}$$

L'ultima uguaglianza é verificata per 2 valori del coefficiente angolare A : $A = 0$ e $A = C(X, Y)/V(X)$, che é proprio il coefficiente angolare della retta di regressione. Il coefficiente B non ha importanza perché né $V(Z)$ né $V(R)$ dipendono da B .

Quindi per la retta di regressione (e per tutte le rette ad essa parallele) possiamo dire che la varianza dei dati $V(Y)$ é “spiegata” in parte dalla “varianza della retta di regressione” $V(Z)$ e in parte dalla varianza dei residui $V(R)$. Se la correlazione lineare tra X e Y é molto buona, la varianza dei residui é molto piccola. La percentuale di varianza spiegata dalla retta di regressione é data da:

$$\begin{aligned} V(Z)/V(Y) &= A^2V(X)/V(Y) = \\ &= C^2(X, Y)V(X)/V^2(X)V(Y) = C^2(X, Y)/V(X)V(Y) \end{aligned}$$

Indicando con $S(X)$ e $S(Y)$ gli scarti quadratici medi di X e di Y , cioè le radici quadrate delle corrispondenti varianze $V(X)$ e $V(Y)$, si definisce *coefficiente di correlazione lineare* tra X e Y la quantità $R = C(X, Y)/S(X)S(Y)$. Poiché dalla formula precedente risulta che la percentuale di varianza spiegata dalla retta di regressione é uguale a R^2 , risulta che R^2 può variare tra 0 e 1, e quindi R può variare tra -1 e 1.

Osserviamo che R ha lo stesso segno di $C(X, Y)$ e quindi di A . Se $R > 0$, la retta di regressione é crescente e si dice che tra X e Y c'è una correlazione positiva. Se $R = 1$, i punti P_i sono esattamente allineati sulla retta di regressione. Se $R = 0$, tra X e Y non c'è nessuna correlazione lineare (questo succede quando i punti P_i sono distribuiti a caso, ma anche in altre situazioni). Se $R < 0$, la retta di regressione é decrescente e si dice che tra X e Y c'è una correlazione negativa (se $R = -1$ i punti sono esattamente allineati sulla retta di regressione).

Ad esempio con $R = 0.9$ si ha $R^2 = 0.81$, il che significa che la retta di regressione spiega l'81% della varianza dei dati (il restante 19% rappresenta la varianza dei residui) e la correlazione é abbastanza buona. In genere si considera buona la correlazione quando $R^2 > 0.8$, discreta quando R^2 é compreso tra 0.5 e 0.8, scarsa o nulla quando $R^2 < 0.5$ (peró per fare un test sulla significatività statistica del valore di R occorrerebbe conoscere la

distribuzione di probabilita' di X e Y come variabili aleatorie).

Osserviamo che se i punti P_i tendono a disporsi con grande precisione su una curva $Y = f(X)$ che non é una retta, facendo la correlazione lineare tra X e Y si trova un valore di R^2 molto piccolo, il che significa correlazione lineare scarsa o nulla. Però ciò non vuol dire che X e Y siano indipendenti, anzi é chiaro che tra esse c'è una buona *correlazione non lineare*. In tal caso si può dapprima ipotizzare la forma della funzione $f(X)$ (ad esempio una parabola $f(X) = AX^2 + BX + C$), poi determinare i valori dei parametri (ad esempio A, B, C) in modo da rendere minima la somma dei quadrati dei residui e infine riportare in un grafico, anziché i valori di X e di Y , i valori di $Z = f(X)$ e di Y . É chiaro che in tal modo i punti tendono a disporsi sulla retta $Y = Z$ e si ottiene un'ottima correlazione lineare tra Z e Y , con un valore di R^2 molto vicino a 1.

Facciamo un esempio: consideriamo i seguenti 8 indici topologici delle molecole di idrocarburi:

v = numero di atomi di C

h = numero di atomi di H

a = numero di legami C - C

r = numero di ramificazioni

c = indice ciclico

d = diametro della molecola (distanza massima tra 2 atomi di C, espressa come numero di legami del cammino piu' corto che li unisce)

W = indice di Wiener (somma delle distanze fra tutte le coppie di vertici)

R = indice di Randic (somma delle quantità $(gg')^{-1/2}$ per ogni lato del grafo, dove g, g' sono i gradi dei 2 estremi del lato)

Vogliamo vedere la correlazione di questi indici col punto di ebollizione Y (espresso in gradi Kelvin) per 10 diverse molecole di idrocarburi. (riportiamo i valori dell'indice topologico sull'asse delle ascisse e i valori di Y sull'asse delle ordinate). I dati sono riassunti nella seguente tabella (NB: ho ricavato i valori del punto di ebollizione Y da un grafico, quindi essi non sono del tutto attendibili). Vedi Tab.3.1

Le figure alle pagine seguenti illustrano graficamente la correlazione tra gli 8 indici topologici e il punto di ebollizione Y . Per ogni grafico sono riportati i valori di A, B (coefficienti della retta di regressione), R^2 (percentuale di varianza spiegata dalla retta di regressione).

Risulta che il punto di ebollizione Y ha una correlazione scarsa o nulla col numero di atomi di H, col numero di legami C - C, col numero di ramificazioni e con l'indice ciclico della molecola. Invece c'è una correlazione

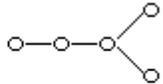
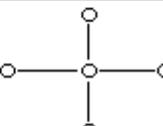
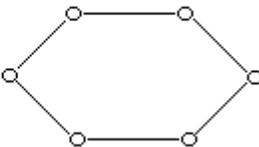
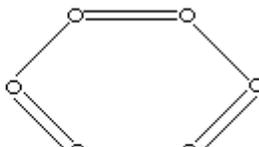
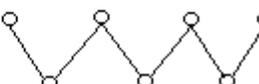
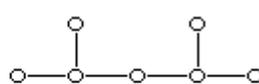
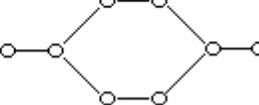
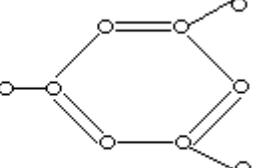
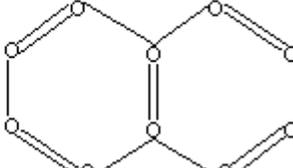
Formula bruta	Grafo ridotto	a	r	c	d	W	R	Y
C_5H_{12}		4	1	0	3	18	2.27	263
C_5H_{12}		4	2	0	2	16	2	260
C_6H_{12}		6	0	1	3	27	3	285
C_6H_6		9	6	4	3	27	3	288
C_7H_{16}		6	0	0	6	56	3.414	318
C_7H_{16}		6	2	0	4	48	3.125	310
C_8H_{16}		8	2	1	5	62	3.787	324
C_9H_{12}		12	9	4	4	84	4.1	349
$C_{10}H_{22}$		9	0	0	9	165	4.914	385
$C_{10}H_8$		16	12	7	5	111	4.121	361

Tabella 3.1

discreta o buona coi seguenti indici (in ordine di R^2 crescente): diametro della molecola ($R^2 = 0.703$), indice di Wiener ($R^2 = 0.921$), indice di Randic ($R^2 = 0.961$), numero di atomi di C ($R^2 = 0.971$). Questo é piuttosto sorprendente, perché il numero di atomi di carbonio v non discrimina tra molecole lineari e molecole ramificate e si sa che a parità di v , le molecole ramificate, piú compatte, hanno punto di ebollizione minore.

Invece l'indice di Wiener W , che tiene conto in un certo senso della forma della molecola, dovrebbe presentare una correlazione migliore col punto di ebollizione Y . Però osservando la figura si vede che nel grafico di Y in funzione di W i punti sono quasi perfettamente disposti lungo una curva, mentre nel grafico di Y in funzione di v i punti tendono a disporsi lungo una retta, ma con precisione minore (nel grafico di Y in funzione dell'indice di Randic R non si nota una tendenza a disporsi lungo una retta né lungo una curva). Questo é il motivo per cui la correlazione lineare tra v e Y é migliore di quella tra W e Y .

Peró operando un opportuno cambiamento di scala (sull'asse x , sull'asse y o su entrambi) si puó rendere lineare la correlazione tra W e Y . Se ipotizziamo una relazione del tipo $Y = AW^B$, da cui $\log(Y) = \log(A) + B \log(W)$, ci dovrebbe essere una relazione lineare tra $\log(W)$ e $\log(Y)$, per cui riportando in ascisse $\log(W)$ e in ordinate $\log Y()$ (scala bilogarithmica), i punti dovrebbero disporsi con grande precisione lungo una retta. Questo é proprio quel che succede nel nostro esempio, dove facendo la correlazione lineare tra $\log(W)$ e $\log(Y)$ si ottiene un valore di $R^2 = 0.995$ (Figura 3.15). Tutto questo si vede molto bene anche nei 2 grafici successivi. (Fig 3.6)

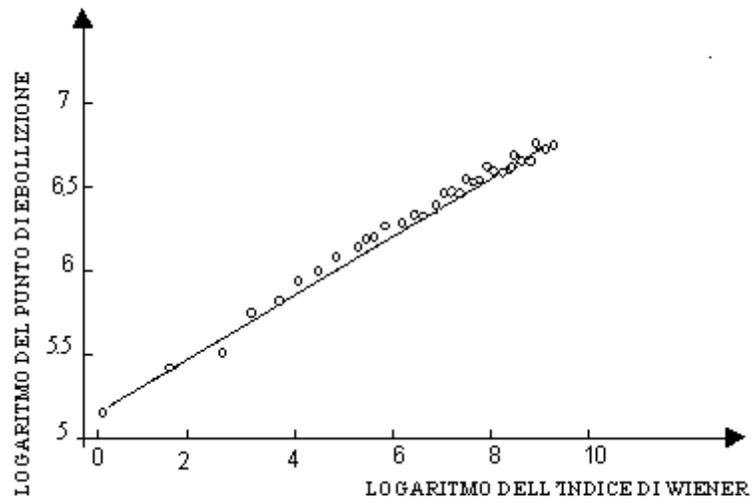
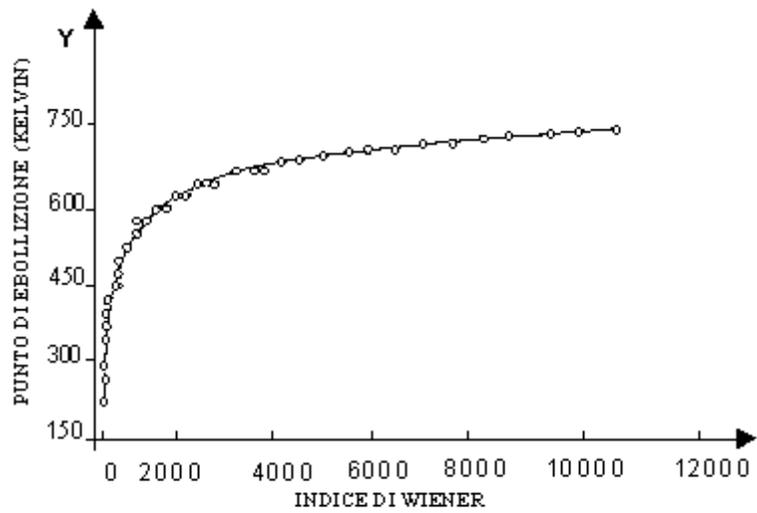


Figura 3.6 (grafici tratti dal n 219 di “Le Scienze”, novembre 1986)

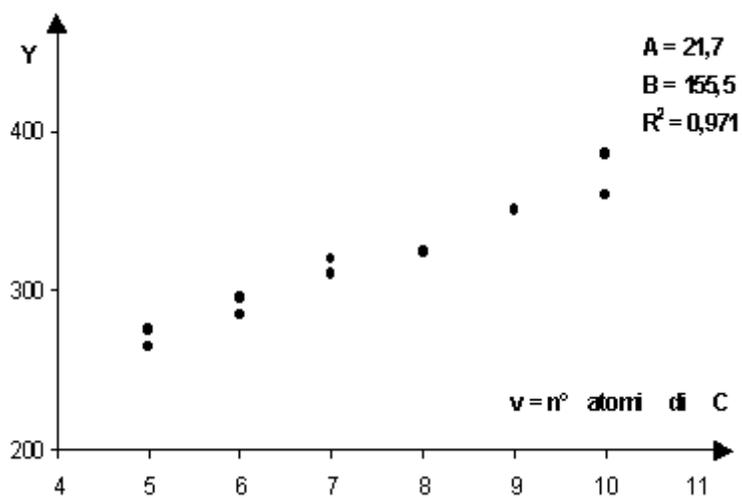


Figura 3.7 (tratta dal n 219 di "Le Scienze", novembre 1986)

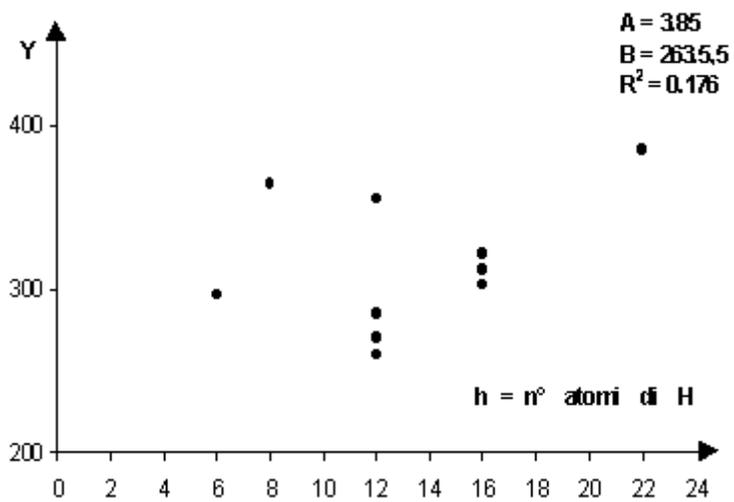


Figura 3.8

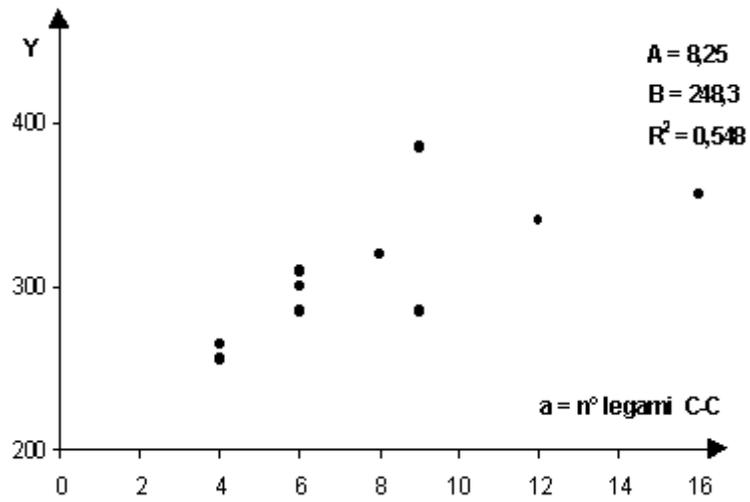


Figura 3.9

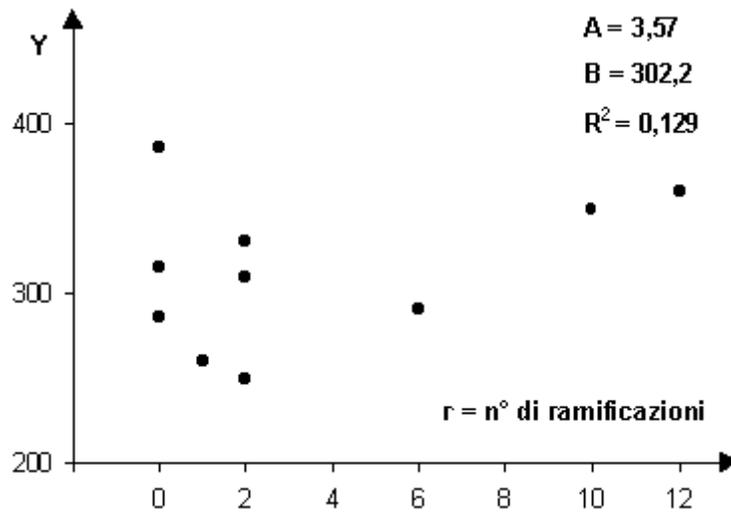


Figura 3.10

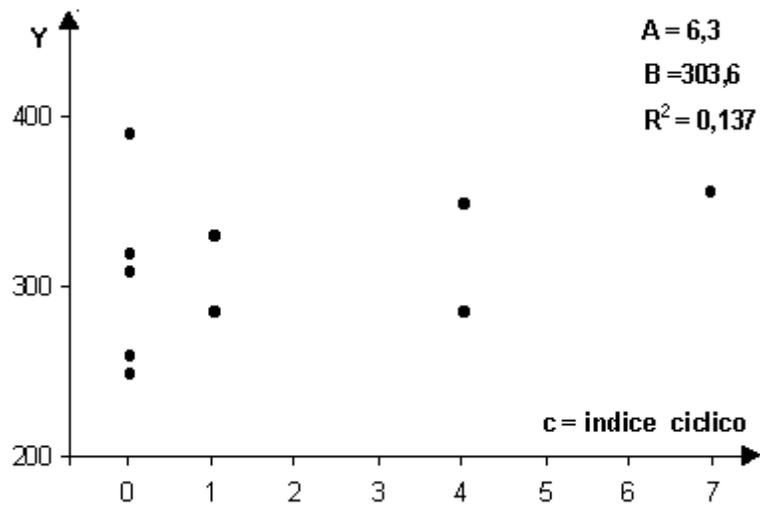


Figura 3.11

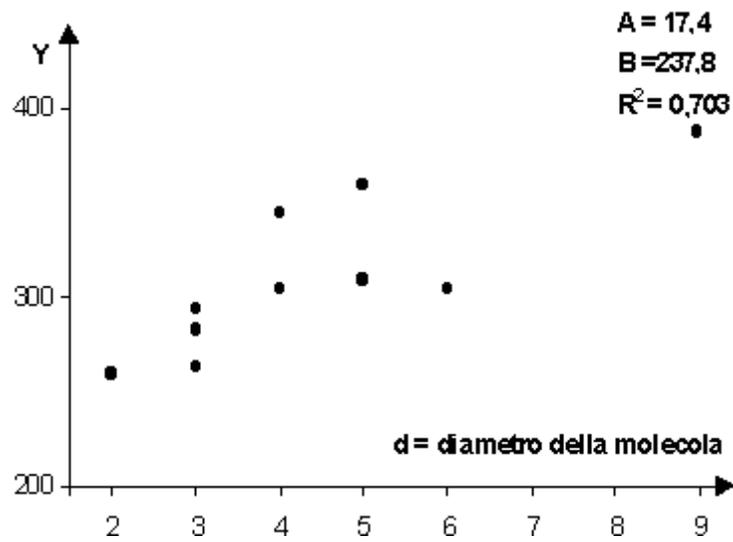


Figura 3.12

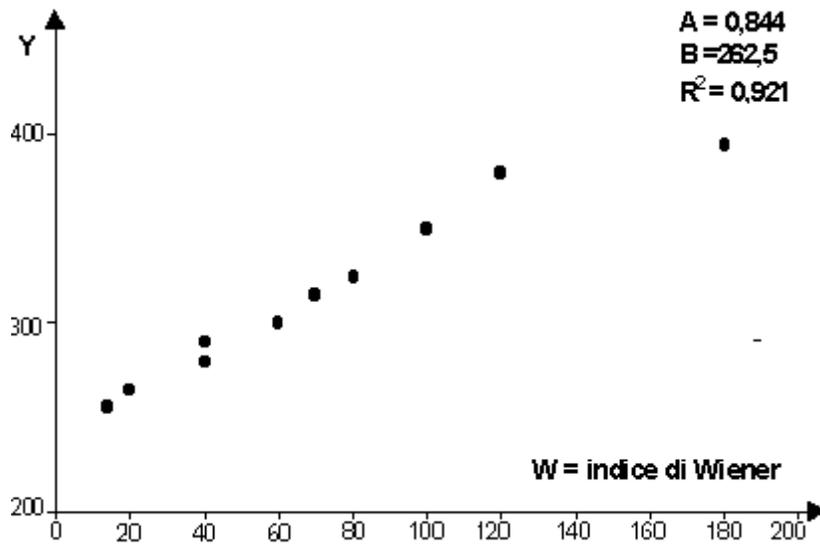


Figura 3.13

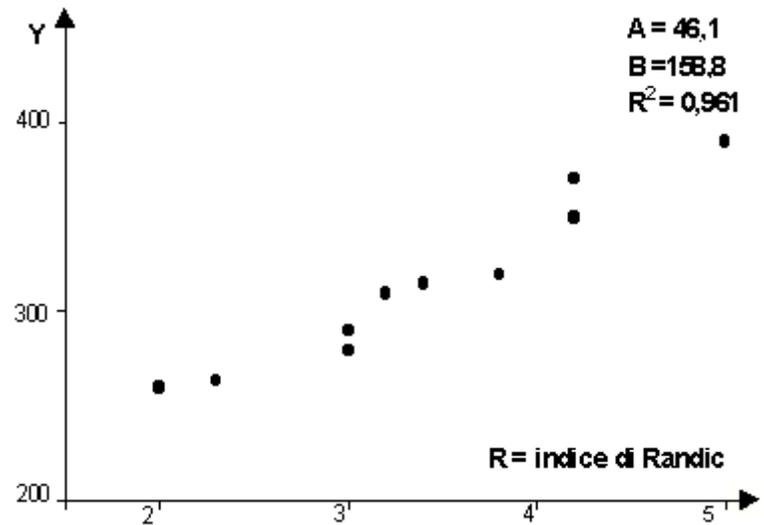


Figura 3.14

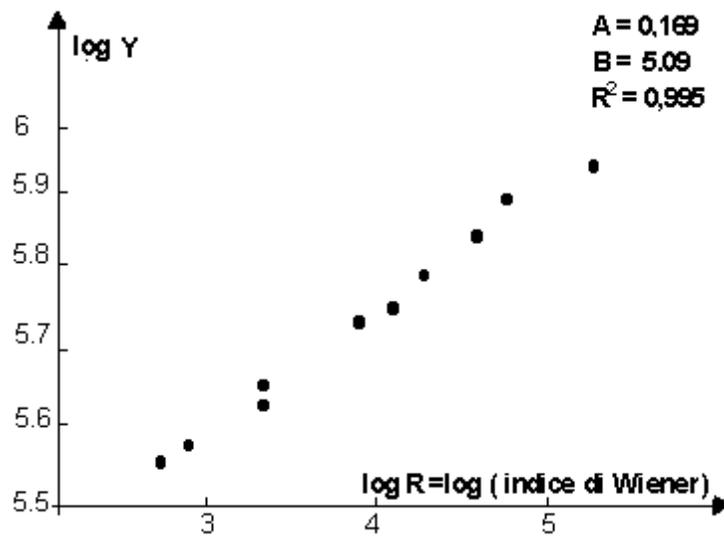


Figura 3.15

4 Il problema dei territori comuni (the tragedy of the commons)

Fin dal settecento (Hume, 1739) gli economisti hanno capito che se ciascuno pensa soltanto ai propri interessi, la ricchezza totale diminuisce e le risorse dell'ambiente si esauriscono in breve tempo. Ecco un esempio che illustra questo fatto.

Consideriamo n pastori in un villaggio, i quali tutte le estati portano le pecore a pascolare in un pascolo comune vicino al villaggio. Sia x_i il numero di pecore possedute dall' i -esimo pastore e sia $s = x_1 + x_2 + \dots + x_n$ il numero totale di pecore. Sia C il costo di allevare una pecora e supponiamo che esso sia indipendente dal numero di pecore possedute da un pastore. Indichiamo con $f(s)$ il ricavo che un pastore ottiene portando a pascolare una pecora; esso dipende dal numero totale s di pecore al pascolo ed è una funzione decrescente di s , perché all'aumentare di s ogni pecora ha meno erba a disposizione. Inoltre, quando ci sono poche pecore al pascolo, aggiungere una pecora danneggia poco le rimanenti (quindi $f(s)$ diminuisce poco), mentre quando il pascolo è vicino al limite di saturazione, aggiungere una pecora danneggia molto le rimanenti e quindi $f(s)$ diminuisce molto. Indicando con M il numero massimo di pecore che il pascolo può sopportare, la funzione f è definita soltanto per i numeri interi compresi tra 1 e M , ma per comodità noi supponiamo di poterla prolungare ad una funzione di variabile reale $f : [0, M] \rightarrow \mathbb{R}$, definita in tutto l'intervallo $[0, M]$, decrescente e concava (cioè $f' \leq 0$, $f'' \leq 0$: questo traduce le osservazioni fatte prima). Supponiamo anche che $f(M) = 0$ e $f(0) = A > C$ (altrimenti non ci sarebbe nessun guadagno a portare le pecore al pascolo). Da queste ipotesi segue che $f(s)$ potrebbe essere costante ($= A$) in un intorno destro $[0, B]$ di 0, ma non appena $f(s)$ diventa $< A$, la sua derivata $f'(s)$ diventa strettamente negativa, quindi $f(s)$ diventa strettamente decrescente.

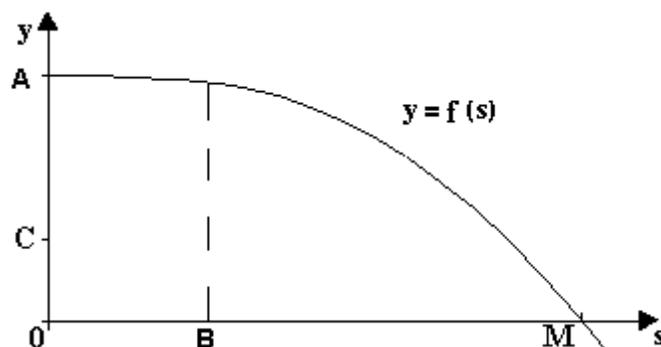


Figura 4.1

Il grafico di $f(s)$ é del tipo illustrato in figura 4.1

In primavera gli n pastori devono decidere contemporaneamente quante pecore portare al pascolo. Se l' i -esimo pastore decide di portare al pascolo x_i pecore, la sua funzione di utilitá u_i (ricavo - spesa) sará:

$$u_i(x_1, x_2 + \dots, x_n) = x_i f(x_1 + \dots + x_n) - C x_i \quad (1)$$

La situazione si puó quindi, schematizzare con un *gioco non cooperativo a n giocatori*, definito come una $(2n+1)$ -upla $G = (N, X_1, X_2 \dots, X_n, u_1, u_2 \dots, u_n)$, dove N é l'insieme degli n giocatori, X_1, \dots, X_n sono gli insiemi delle strategie degli n giocatori e $u_i : X_1 \times X_2 \times \dots \times X_n \rightarrow R$ é la funzione di utilitá del giocatore i -esimo. Ció significa che, fissate la strategie $x_1, x_2 \dots, x_n$ degli n giocatori (x_i appartiene all'insieme X_i), $u_i(x_1, x_2 \dots, x_n)$ rappresenta il guadagno del giocatore i -esimo. Nel nostro caso gli insiemi delle strategie X_1, \dots, X_n coincidono tutti con l'intervallo $[0, M]$. La strategia x_i dell' i -esimo giocatore é semplicemente il numero di pecore che egli decide di portare al pascolo e la sua funzione di utilitá u_i é data dalla formula (1) precedente.

Ricordiamo che per un gioco non cooperativo a n giocatori, un *equilibrio di Nash* é una n -pla di strategie $X^* = (x_1^*, x_2^* \dots, x_n^*)$ tale che nessun giocatore abbia interesse, partendo da X^* , a cambiare unilateralmente la sua strategia, ció tale che sia :

$$u_i(x_1^*, \dots, x_{i-1}^*, x_i, x_{i+1}^*, \dots, x_n^*) \leq u_i(x_1^*, \dots, x_{i-1}^*, x_i^*, x_{i+1}^*, \dots, x_n^*)$$

per ogni strategia x_i appartenete a X_i e per ogni i da 1 a n .

Teorema 4.1 *Nelle ipotesi fatte, il gioco dei territori comuni $G = (N, X_1, X_2 \dots, X_n, u_1, u_2 \dots, u_n)$ ha uno e un solo equilibrio di Nash.*

Dimostrazione. Se esiste un equilibrio di Nash, $(x_1^*, x_2^* \dots, x_n^*)$, in esso si devono annullare le derivate parziali $\delta u_i / \delta x_i$. Infatti, fissate le strategie degli altri giocatori, la funzione di utilitá u_i del giocatore i -esimo deve avere un massimo nel punto di ascissa x_i^* . Dev'essere quindi:

$$\begin{aligned} \delta u_i / \delta x_i(x_1^*, \dots, x_n^*) &= f(x_1^* + \dots + x_n^*) + x_i^* f'(x_1^* + \dots + x_n^*) - C = \\ &= f(s^*) + x_i^* f'(s^*) - C = 0 \end{aligned}$$

Da cui si ricava:

$$f(s^*) + x_i^* f'(s^*) = C \quad (2)$$

Ponendo in questa uguaglianza prima $i = 1$ e poi $i = 2$ e sottraendo membro a membro si ha :

$$(x_1^* - x_2^*) f'(s^*) = 0$$

e questo significa che $x_1^* = x_2^*$ oppure $f'(s^*) = 0$. Ma $f'(s^*)$ non può essere 0 perché altrimenti dovrebbe essere $f(s^*) = A$ e questo contrasta con la (2) perché $A > C$. Dev'essere quindi, $x_1^* = x_2^*$ e analogamente $x_2^* = x_3^* = \dots = x_n^*$. Indicando questo valore comune con x^* , si ha $s^* = nx^*$ e quindi, la (2) diventa:

$$f(nx^*) + x^* f'(nx^*) = C \quad (3)$$

Indichiamo ora con $g(x)$ il primo membro della (3) (con x al posto di x^*) e facciamone la derivata:

$$g(x) = f(nx) + x f'(nx)$$

$$g'(x) = n f'(nx) + f'(nx) + n x f''(nx) < 0 \quad \forall x \in [B/n, M/n]$$

Quindi, $g(x)$ è strettamente decrescente in $[B/n, M/n]$. Si ha $g(B/n) = A$ e $g(M/n) = M f'(M)/n < 0$ e poiché $A > C$, nell'intervallo $[0, M/n]$ esiste un unico punto x^* in cui $g(x^*) = C$, cioè vale la (3). Con ciò abbiamo dimostrato che l'equilibrio di Nash, se esiste, è unico perché deve necessariamente coincidere col punto $(x^*, x^* \dots, x^*)$.

Resta da vedere che un equilibrio di Nash esiste davvero, cioè che il punto $X^* = (x^*, x^* \dots, x^*)$ è effettivamente un equilibrio di Nash. Poiché X^* annulla tutte le derivate parziali $\delta u_i / \delta x_i$, basta far vedere che le derivate parziali seconde $\delta^2 u_i / \delta x_i^2$ sono ≤ 0 in tutto $[0, M]$ (perché allora, facendo variare la variabile x_i e tenendo fisse le altre al valore x^* , la funzione di utilità u_i presenta un massimo per $x_i = x^*$). E infatti,

$$\delta^2 u_i / \delta x_i^2 = 2 f'(x_1, \dots, x_n) + x_i f''(x_1, \dots, x_n) \leq 0$$

come volevasi dimostrare. ■

Poiché il gioco dei territori comuni ha uno e un solo equilibrio di Nash, se gli n pastori vogliono mettersi d'accordo prima su quante pecore far pascolare, sembrerebbe che la soluzione più razionale sia quella di scegliere proprio questo equilibrio di Nash, cioè scegliere $x_1 = x_2 = \dots = x_n = x^*$. Solo così infatti, nessuno dei pastori ha interesse a tradire unilateralmente l'accordo. Però in questo caso l'equilibrio di Nash ha un grave difetto: non è efficiente, nel senso che esiste un'altra scelta di strategie che permette a tutti i giocatori di guadagnare di più. Se gli n pastori, invece di mettersi d'accordo sull'equilibrio di Nash, si mettessero d'accordo per massimizzare l'utilità sociale $u = u_1 + u_2 + \dots + u_n$ e poi dividessero in parti uguali il guadagno, guadagnerebbero tutti di più.

Infatti, l'utilità sociale è data da :

$$\begin{aligned} u(x_1, \dots, x_n) &= (x_1 + \dots + x_n) f(x_1 + \dots + x_n) - C(x_1 + \dots + x_n) \\ &= s f(s) - C(s) = u(s) \end{aligned}$$

cioé u dipende solo dal numero totale $s = x_1 + \dots + x_n$ di pecore al pascolo. Derivando $u(s)$ rispetto a s otteniamo:

$$\begin{aligned} u'(s) &= f(s) + sf'(s) - C \\ u'(B) &= A - C > 0 \\ u'(M) &= Mf'(M) - C < 0 \\ u''(s) &= 2f'(s) + sf''(s) < 0 \quad \forall s \in [B, M] \\ u''(s) &\leq 0 \quad \forall s \in [0, M] \end{aligned}$$

Quindi, $u'(s)$ é decrescente in $[0, M]$, anzi, é strettamente decrescente in $[B, M]$. Allora, esiste un unico punto s' in $[0, M]$ tale che $u'(s') = 0$. Ciò dimostra che $u(s)$ ha un unico punto di massimo in $[0, M]$ per $s = s'$.

Si può provare che $s' < s^*$ dove s^* é il valore di s nell'equilibrio di Nash. Infatti, $s^* = nx^*$ e dalla (3) si deduce che $f(s^*) + f'(s^*)s^*/n = C$, ma allora $u'(s^*) = f(s^*) + s^* f'(s^*) - C = s^* f'(s^*)(n - 1)/n < 0$. Poiché $u'(s') = 0$ e u' é decrescente, dev'essere $s' < s^*$.

Posto $x' = s'/n$, se tutti i giocatori adottano la strategia x' ($< x^*$) e quindi, portano al pascolo meno pecore di quel che stabilisce l'equilibrio di Nash, guadagnano tutti di piú (guadagnano $u(s')/n$ anziché $u(s^*)/n$ e $u(s') > u(s^*)$).

A questo punto però, uno dei pastori potrebbe essere tentato di “fare il furbo” e tradire unilateralmente l'accordo, perché cosí facendo potrebbe guadagnare ancora di piú; infatti, il punto $(x', x' \dots, x')$ non é un equilibrio di Nash. Però naturalmente, gli altri giocatori reagirebbero cambiando a loro volta la loro strategia e si avrebbe una successione di strategie che tende all'equilibrio di Nash, col risultato che tutti i giocatori guadagnerebbero meno del “massimo sociale” $u(s')/n$.

Per illustrare questo fatto mettiamoci in un caso particolare molto semplice: supponiamo che ci siano solo 2 giocatori, indichiamo le loro strategie con x, y (anziché x_1, x_2) e supponiamo che la funzione $f(s)$ sia lineare, cioè poniamo $f(s) = A(1 - s/M)$ (cosa non realistica). Si ha allora:

$$\begin{aligned} u_1(x, y) &= xf(x + y) - Cx = x\left(A - C - A(x + y)/M\right) \\ u_2(x, y) &= yf(x + y) - Cy = y\left(A - C - A(x + y)/M\right) \end{aligned}$$

L'equilibrio di Nash (x^*, y^*) si trova ponendo $\delta u_1/\delta x = \delta u_2/\delta y = 0$:

$$\begin{aligned} \delta u_1/\delta x &= A - C - A(x + y)/M - Ax/M = 0 & M(A - C) &= A(2x + y) \\ \delta u_2/\delta y &= A - C - A(x + y)/M - Ay/M = 0 & M(A - C) &= A(x + 2y) \end{aligned}$$

da cui $x = y = M(A - C)/3A$ e quindi l'equilibrio di Nash é dato da :

$$(x^*, x^*) = \left(M(A - C)/3A, M(A - C)/3A\right)$$

e l'utilità di ciascuno dei 2 giocatori nell'equilibrio di Nash vale:

$$u_1(x^*, x^*) = M(A - C)^2/9A$$

Invece l'utilità sociale $u(x, y) = u_1(x, y) + u_2(x, y)$ vale:

$$u(x, y) = (x + y) \left(A - C - A(x + y)/M \right) = s(A - C - As/M)$$

Il massimo dell'utilità sociale si ha nel punto $s' = (x', x')$ e si trova ponendo $du/ds = 0$:

$$du/ds = A - C - aS/M - aS/M = A - C - 2aS/M = 0$$

da cui

$$s' = M(A - C)/2A \quad x' = M(A - C)/4A$$

L'utilità di ciascuno dei 2 giocatori nel punto (x', x') vale $u_1(x', x') = M(A - C)^2/8A$ e quindi, é maggiore dell'utilità data dall'equilibrio di Nash.

I valori delle costanti A, C, M dipendono dalle unità di misura usate. Supponiamo per comodità che sia $A = M = 1, C = 0$. Allora, si ha:

$$\begin{aligned} u_1(x, y) &= x(1 - x - y) & u_2(x, y) &= y(1 - x - y) \\ (x^*, x^*) &= (1/3, 1/3) & u_1(x^*, x^*) &= u_2(x^*, x^*) = 1/9 \\ (x', x') &= (1/4, 1/4) & u_1(x', x') &= u_2(x', x') = 1/8 \end{aligned}$$

E la situazione può essere rappresentata dalla figura 4.2: il punto $P = (1/3, 1/3)$ rappresenta l'equilibrio di Nash. Il punto $P_0 = (1/4, 1/4)$ rappresenta il massimo sociale. Le 2 rette $y = (1 - x)/2$ e $x = (1 - y)/2$ rappresentano le *curve di miglior risposta* dei 2 giocatori. Questo significa che, fissata la strategia y del secondo giocatore, la miglior risposta del primo giocatore é quella che rende massima la sua funzione di utilità. Essa si indica con $r_1(y)$ e si ottiene ponendo $\delta u_1/\delta x = 1 - 2x - y = 0$, da cui $x = r_1(y) = (1 - y)/2$.

Analogamente la miglior risposta del secondo giocatore, indicata con $r_2(x)$, si ottiene ponendo $\delta u_2/\delta y = 1 - x - 2y = 0$, da cui $y = r_2(x) = (1 - x)/2$. L'insieme degli equilibri di Nash, per definizione, é dato dall'intersezione delle 2 curve di miglior risposta e nel nostro caso contiene un solo punto, il punto $P = (1/3, 1/3)$.

Supponiamo che i 2 giocatori si siano messi d'accordo per la strategia che dá il massimo sociale, rappresentata dal punto $P_0 = (1/4, 1/4)$, che dá a ciascuno un guadagno $u_1(1/4, 1/4) = u_2(1/4, 1/4) = 1/8 = 0.125$.

A questo punto il primo giocatore può essere tentato di "fare il furbo" e tradire l'accordo, spostandosi nel punto P_1 che sta sulla sua curva di

miglior risposta e ha coordinate $(3/8, 1/4)$. Così facendo il primo giocatore porta piú pecore al pascolo e aumenta il suo guadagno, guadagnando $9/64 = 0.141$ anziché $1/8 = 0.125$, mentre il secondo giocatore guadagna di meno ($3/32 = 0,094$ anziché $0,125$).

Ma naturalmente il secondo giocatore in seguito reagirá spostandosi a sua volta sulla sua curva di miglior risposta, nel punto $P_2 = (3/8, 5/16)$ e aumentando a sua volta il numero di pecore al pascolo. A questo punto, come mostra la figura 4.2, il primo giocatore, se vuole aumentare di nuovo il suo guadagno, sará costretto a diminuire il numero delle pecore al pascolo. Si puó andare avanti cosí indefinitamente, costruendo una successione di punti $P_1, P_2, P_3 \dots P_k, \dots$ e si puó dimostrare che essa (in questo caso) tende all'equilibrio di Nash, come é chiaro anche dalla figura 4.2.

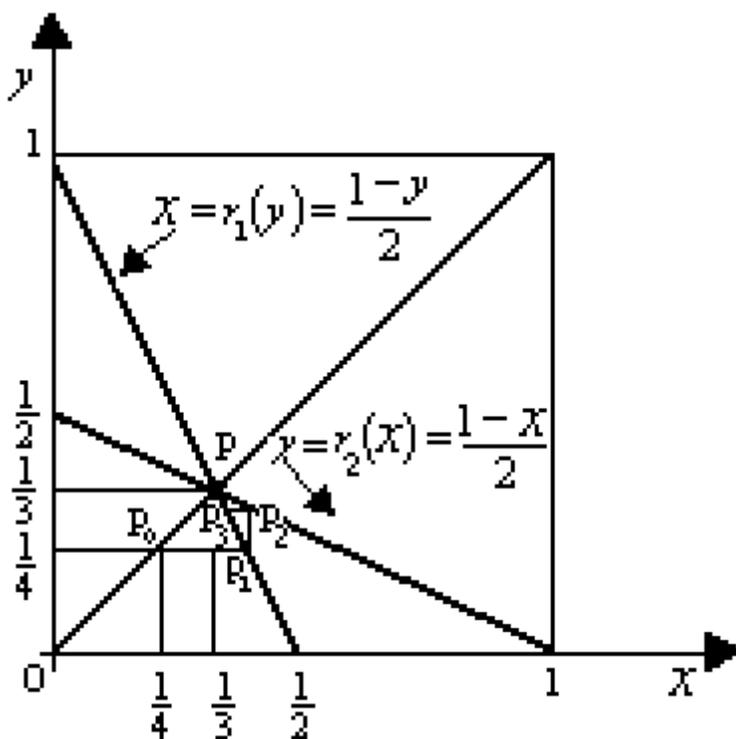


Figura4.2

I valori di x, y, u_1, u_2 per i punti $P_0, P_1, P_2, P_3 \dots, P$ sono i seguenti:

	x	y	u_1	u_2
P_0	1/4	1/4	0,125	0,125
P_1	3/8	1/4	0.141	0.094
P_2	3/8	5/16	0.117	0.098
P_3	11/32	5/16	0.118	0.107
P	1/3	1/3	0.111	0.111

Qual é il risultato finale? che entrambi i giocatori si trovano a guadagnare meno di prima (0.111 anziché 0.125) e che il pascolo viene sfruttato piú di prima (il numero totale di pecore al pascolo é $2M/3$ anziché $M/2$) e si esaurisce piú in fretta. É chiaro che da questa storiella si puó ricavare una morale “ecologica” fin troppo ovvia.

5 Giochi evolutivi

Introdurró adesso la nozione di *equilibrio evolutivo*, che é utile per studiare il comportamento degli esseri viventi in relazione al meccanismo dell'evoluzione e della selezione naturale. Mi limiteró al semplice caso in cui gli individui di una singola popolazione di organismi viventi (animali, esseri umani, piante) interagiscono fra loro a coppie (per esempio: lotte fra animali della stessa specie per il cibo, per il territorio o per l'accoppiamento).

In ciascuna lotta, ciascun animale si comporta in un modo ben determinato, "scegliendo" un' azione (che chiameró "strategia") da un certo "insieme di strategie" S , che é lo stesso per tutti gli animali. Supponiamo però che l'animale non scelga l'azione consciamente, ma seguendo un comportamento istintivo, geneticamente determinato. Cioé il comportamento dell'animale é dovuto ai geni che esso ha ereditato dai suoi antenati, oppure a delle mutazioni.

A seconda delle strategie dei 2 contendenti, ciascuno di essi ottiene un certo vantaggio o un certo svantaggio, che supponiamo possa essere quantificato mediante una certa "funzione di utilità", misurata non in termini delle possibilità di sopravvivenza dell'individuo, ma del suo successo riproduttivo futuro, ossia in termini delle possibilità di sopravvivenza della specie (in altre parole, il "successo" di un individuo consiste nel riprodursi il piú possibile e quindi diffondere il piú possibile i suoi geni nella popolazione).

Se un animale A incontra un animale B e adotta una certa strategia a , mentre l'avversario adotta una strategia b (a, b nell'insieme S), indichiamo con $u(a, b)$ l'utilità di A . Per simmetria, l'utilità di B sará data da $u(b, a)$, dove u é la stessa funzione. Quindi la situazione puó essere schematizzata mediante un gioco non cooperativo a 2 giocatori $G = (A, B, S, S, u_1, u_2)$, dove A, B sono i 2 giocatori, S é l'insieme delle strategie (uguale per i 2 giocatori), u_1 e u_2 sono le funzioni di utilità dei 2 giocatori e sono date da $u_1(a, b) = u(a, b)$, $u_2(a, b) = u(b, a)$. Un gioco che gode di queste proprietà si chiama *gioco simmetrico*.

A questo punto diamo la seguente definizione di *strategia evolutivamente stabile* o ESS (evolutionary stable strategy):

Definizione 5.1 *Una strategia evolutivamente stabile (ESS) b^* é una strategia tale che se in una popolazione di animali che adottano tutti la strategia b^* avviene una qualunque mutazione e se questa mutazione fa sí che una piccola percentuale della popolazione assuma una strategia b diversa da b^* , allora l'utilità attesa di un mutante é sempre minore di quella di un non mutante (qualunque sia b) e quindi i mutanti tendono ad estinguersi.*

In altre parole la definizione precedente significa che la strategia b^* non é una ESS se e solo se esiste almeno una strategia b diversa da b^* tale che, se avviene per caso una mutazione che fa assumere ad alcuni mutanti la

strategia b , questi mutanti proliferano ed invadono la popolazione. Ma che cosa significa “utilita’ attesa” ? Per capirlo dobbiamo fare alcune considerazioni di carattere probabilistico. Se abbiamo una popolazione in cui tutti gli animali adottano la stessa strategia b^* e se in essa avviene una mutazione che fa sí che una piccola percentuale di mutanti adotti una strategia diversa b , indichiamo con ε la percentuale iniziale dei mutanti (ε é un numero positivo compreso tra 0 e 1, ma “molto piccolo”, cioè molto piú vicino a 0 che a 1). É chiaro che, in una lotta tra 2 animali, l’utilità di ciascuno di essi dipende dal fatto che sia lui che il suo avversario siano mutanti o non mutanti, cioè dipende dalla probabilità che ciascun animale ha di incontrare un mutante (questa probabilità vale ε) oppure un non mutante ($1 - \varepsilon$). Quello che conta é l’utilità attesa, o utilità media, di ciascun animale. Per un non mutante (che adotta la strategia b^*) l’utilità attesa vale:

$$u = (1 - \varepsilon)u(b^*, b^*) + \varepsilon u(b^*, b)$$

in quanto se il non mutante incontra un altro non mutante ottiene un’utilità $u(b^*, b^*)$, mentre se incontra un mutante ottiene $u(b^*, b)$. Invece l’utilità attesa di un mutante vale:

$$u' = (1 - \varepsilon)u(b, b^*) + \varepsilon u(b, b)$$

Quindi la strategia b^* é una ESS se e solo se $u' < u$ per ogni strategia b diversa da b^* e per ogni ε sufficientemente piccolo. Possiamo allora riformulare la definizione di ESS in questo modo piu’ preciso (molto tecnico e matematico):

Definizione 5.2 *La strategia b^* é una ESS se e solo se per ogni strategia b diversa da b^* esiste $\delta > 0$ tale che per ogni ε nell’intervallo $(0, \delta)$ si abbia:*

$$(1 - \varepsilon)u(b, b^*) + \varepsilon u(b, b) < (1 - \varepsilon)u(b^*, b^*) + \varepsilon u(b^*, b) \quad (1)$$

Passando al limite per ε tendente a zero da questa disuguaglianza si ottiene:

$$u(b, b^*) \leq u(b^*, b^*) \quad \forall b \neq b^* \quad (2)$$

dalla quale segue che nel gioco $G = (A, B, S, S, u_1, u_2)$ la coppia di strategie (b^*, b^*) é un equilibrio di Nash. Infatti poiché $u_1(a, b) = u(a, b)$ e $u_2(a, b) = u(b, a)$, la (2) significa che $u_1(b, b^*) \leq u_1(b^*, b^*)$ e $u_2(b^*, b) \leq u_2(b^*, b^*)$ per ogni b diversa da b^* , che é proprio la definizione di equilibrio di Nash per (b^*, b^*) . (Ricordiamo che, in un gioco non cooperativo a 2 giocatori, un equilibrio di Nash é una coppia di strategie a partire dalla quale nessuno dei 2 giocatori puó migliorare la sua utilità cambiando unilateralmente strategia). Quindi la (2) é condizione necessaria perché valga la (1), cioè perché b^* sia una ESS, però non é condizione sufficiente. Infatti se nella (2) vale il segno

$<$ in senso stretto, allora é chiaro che per ε abbastanza piccolo vale anche la (1) (per il teorema della permanenza del segno sui limiti di una funzione), ma se nella (2) vale il segno $=$, allora la (1) diventa semplicemente:

$$\varepsilon u(b, b) < \varepsilon u(b^*, b)$$

e quindi é verificata se e solo se vale l'ulteriore condizione $u(b, b) < u(b^*, b)$. Abbiamo quindi dimostrato il seguente

Teorema 5.1 *La strategia b^* é evolutivamente stabile (ESS) se e solo se nel gioco $G = (A, B, S, S, u_1, u_2)$ la coppia di strategie (b^*, b^*) é un equilibrio di Nash che verifica l'ulteriore condizione:*

$$u(b, b^*) = u(b^*, b^*) \text{ con } b \text{ diversa da } b^* \Rightarrow u(b, b) < u(b^*, b) \quad (3)$$

(il gioco G si chiama anche “gioco evolutivo” e se b^* é una ESS, l'equilibrio di Nash (b^*, b^*) si chiama anche “equilibrio evolutivo”).

Esempio 5.1 Supponiamo che ogni animale abbia a disposizione solo 2 strategie a, b e che la tabellina dei valori della funzione di utilitá u sia la seguente:

$A \backslash B$	a	b
a	1	1
b	0	1

É facile verificare che il gioco $G = (A, B, S, S, u_1, u_2)$ ha 2 equilibri di Nash (a, a) e (b, b) , ma solo la strategia a é una ESS, mentre b non lo é. ■

Esempio 5.2 *Gioco Falchi - Colombe in strategie pure*

Un esempio classico di gioco evolutivo é il cosiddetto “gioco Falchi - Colombe”. Supponiamo che in una popolazione di animali della stessa specie si verifichino delle contese fra 2 animali per la conquista delle prede e poniamo uguale a 1 l'utilitá che un animale ottiene conquistando una preda. Supponiamo anche che ogni animale abbia a disposizione 2 sole strategie: un comportamento aggressivo, da “Falco” (F) e uno non aggressivo, da “Colomba” (C).

Se entrambi gli animali si comportano da Colomba, allora si dividono la preda e quindi ognuno ottiene $1/2$. Se uno si comporta da Falco e l'altro da Colomba, allora il Falco vince e ottiene 1, la Colomba scappa e ottiene 0. Se entrambi si comportano da Falco, allora combattono fra loro; in media ciascuno ottiene meta' della preda, ma deve pagare anche un certo prezzo c per il combattimento (in termini di fatica o di ferite ricevute) e quindi poniamo l'utilità di entrambi pari a $1/2 - c$. La tabellina dei valori della funzione di utilità u é quindi la seguente:

A \ B	F	C
F	$\frac{1}{2} - c$	1
C	0	$\frac{1}{2}$

A questo punto possiamo distinguere 3 casi:

Caso 1 : $c < 1/2$ (il costo della lotta é meno della metà del valore della preda). In tal caso é facile verificare che il gioco $G = (A, B, S, S, u_1, u_2)$ ha un solo equilibrio di Nash (F, F) , che é anche un equilibrio evolutivo, e inoltre la strategia F é evolutivamente stabile (mentre C non lo é). Notiamo però che in questo caso, come spesso succede, l'equilibrio di Nash ha un grave difetto: *non é efficiente*, nel senso che se entrambi gli animali adottassero la strategia C guadagnerebbero di piú (otterrebbero $1/2$ anziche' $1/2 - c$). Ma in questo caso le leggi di natura (la selezione naturale) orientano la popolazione verso la strategia F , come vedremo meglio fra poco.

Caso 2 : $c = 1/2$. In tal caso si vede subito che il gioco G ha 3 equilibri di Nash $(F, F), (F, C), (C, F)$, ma solo uno dei 3, e precisamente (F, F) , é un equilibrio evolutivo. La strategia F é l'unica ESS.

Caso 3 : $c > 1/2$ (il costo della lotta supera la metà del valore della preda). In tal caso il gioco G ha 2 equilibri di Nash (F, C) e (C, F) , ma naturalmente nessuno dei 2 é un equilibrio evolutivo (perché corrisponde a 2 strategie diverse). Ne segue che non esiste nessuna ESS.

Per verificare tutto questo, vediamo cosa succede se una mutazione fa comparire alcune Colombe (C) in una popolazione di Falchi (F), o alcuni Falchi

in una popolazione di Colombe. Anzi, consideriamo il caso generale di una popolazione mista di Falchi e Colombe, con una percentuale ε di Falchi e una percentuale $1 - \varepsilon$ di colombe (ε é un numero dell'intervallo aperto $(0, 1)$). Poiché in questa popolazione ogni animale ha una probabilità ε di incontrare un Falco e una probabilità $1 - \varepsilon$ di incontrare una colomba, le utilita' attese u, u' di un Falco e di una Colomba sono rispettivamente:

$$\begin{aligned} u &= \varepsilon(1/2 - c) + (1 - \varepsilon)1 = \varepsilon/2 - c\varepsilon + 1 - \varepsilon = 1 - c\varepsilon - \varepsilon/2 \\ u' &= \varepsilon 0 + (1 - \varepsilon)1/2 = 1/2 - \varepsilon/2 \end{aligned}$$

quindi i Falchi sono favoriti rispetto alle Colombe se e solo se $1 - c\varepsilon - \varepsilon/2 > -\varepsilon/2$, cioè $\varepsilon < 1/(2c)$.

Nei casi 1 e 2 ($c \leq 1/2$) si ha $1/(2c) \geq 1$ e quindi $\varepsilon < 1/(2c)$ é sempre vera, cioè i Falchi sono sempre favoriti: qualunque percentuale di Colombe nella popolazione é destinata a diminuire e ad estinguersi e rimarrá una popolazione di soli Falchi. Ciò conferma il fatto che F é la sola ESS.

Diversa é la situazione nel caso 3, cioè se il costo della lotta é troppo alto. In questo caso $1/(2c)$ é un numero positivo minore di 1 e i Falchi sono favoriti solo se la loro proporzione nella popolazione é minore di questo valore. Se si parte con una percentuale di Falchi uguale a $1/(2c)$, allora Falchi e Colombe sono ugualmente favoriti e la popolazione é una popolazione mista in equilibrio: le percentuali di Falchi e Colombe rimangono costanti. Se invece la percentuale di Falchi é minore di $1/(2c)$, aumenta e tende al valore di equilibrio $1/(2c)$. Infine, se la percentuale di Falchi é maggiore di $1/(2c)$, diminuisce e tende a $1/(2c)$. Questo conferma il fatto che nessuna delle 2 strategie F, C é evolutivamente stabile: non appena in una popolazione di puri Falchi appare un mutante Colomba, o in una popolazione di pure Colombe appare un mutante Falco, il mutante prolifera e si diffonde nella popolazione, finché la sua percentuale raggiunge il valore critico di equilibrio.

Questo risultato fa sorgere il sospetto che nel caso 3 ($c > 1/2$) gli animali possano adottare una "strategia mista" (cioé comportarsi a volte da Falco e a volte da Colomba) che sia evolutivamente stabile. Può darsi che gli animali abbiano un "grado di aggressività" fisso, determinato geneticamente, espresso da un numero x compreso tra 0 e 1, che non é altro che la probabilità con cui essi si comportano da Falchi (naturalmente $1 - x$ é la probabilità che si comportino da Colombe). Nel "gioco Falchi - Colombe in strategie miste" che ora esamineremo, le strategie a disposizione di ciascun animale non sono solo F e C (strategie pure), ma sono rappresentate da un qualsiasi numero x dell'intervallo $[0, 1]$ (strategie miste). ■

Esempio 5.3 *Gioco Falchi - Colombe in strategie miste*

In questo caso nel gioco $G = (A, B, S, S, u_1, u_2)$ l'insieme S delle strategie co-

incide con l'intervallo chiuso $[0, 1]$ per entrambi i giocatori. Indichiamo con x la strategia del primo giocatore e con y quella del secondo, cioè il primo giocatore si comporta da Falco con probabilità x e da Colomba con probabilità $1 - x$, mentre il secondo si comporta da Falco con probabilità y e da Colomba con probabilità $1 - y$. Allora le funzioni di utilità $u_1(x, y), u_2(x, y)$ dei 2 giocatori si calcolano come utilità medie, tenendo conto della probabilità che ciascuno dei 2 si comporti da Falco o da Colomba. Si ha:

$$\begin{aligned}
 u_1(x, y) = u(x, y) &= \left(\frac{1}{2} - c\right)xy + 1x(1 - y) + 0y(1 - x) + \frac{1}{2}(1 - x)(1 - y) = \\
 &= \frac{xy}{2} - cxy + x - xy + \frac{1}{2} - \frac{x}{2} - \frac{y}{2} + \frac{xy}{2} = \frac{(x - y + 1)}{2} - cxy \\
 u_2(x, y) &= u(y, x)
 \end{aligned}$$

Esaminiamo separatamente 3 casi, a seconda che il parametro c (costo della lotta rispetto al valore della preda) sia minore, uguale o maggiore di $1/2$:

Caso 1 : $c < 1/2$. Per trovare gli equilibri di Nash del gioco in strategie miste conviene considerare le *curve di miglior risposta* R_1, R_2 dei 2 giocatori. Definiamo la curva di miglior risposta R_1 del primo giocatore come l'insieme dei punti $(r_1(y), y)$, dove $r_1(y)$ indica una miglior risposta del primo giocatore alla strategia y del secondo giocatore. Cioé, fissata y , $r_1(y)$ indica un valore di x che rende massima $u(x, y)$ al variare di x . Diciamo "una miglior risposta" e non "la miglior risposta" perché di valori di x che rendono massima $u(x, y)$ possono essercene tanti, anche infiniti. Quindi la "funzione di miglior risposta" non é una funzione nel senso usuale di funzione ad un solo valore: é una funzione a piú valori, detta anche *multifunzione*.

In modo analogo si definisce la curva di miglior risposta R_2 del secondo giocatore come l'insieme dei punti $(x, r_2(x))$, dove $r_2(x)$ indica un valore di y che, fissata la strategia x del primo giocatore, renda massima l'utilità $u_2(x, y)$ del secondo giocatore, al variare di y . Ciò posto, l'insieme degli equilibri di Nash del gioco G é semplicemente $R_1 \cap R_2$, cioè l'intersezione delle 2 curve di miglior risposta.

Poiché $u(x, y)$, fissata y , é una funzione lineare di x , basta calcolare i valori che essa assume per $x = 0$ e per $x = 1$ per trovare la curva di miglior risposta $x = r_1(y)$: se $u(0, y) > u(1, y)$ si ha $r_1(y) = 0$, se $u(0, y) < u(1, y)$ si ha $r_1(y) = 1$ e infine se $u(0, y) = u(1, y)$ si ha $r_1(y) =$ l'intero intervallo $[0, 1]$. Quindi:

$$\begin{aligned}
 u(0, y) &= (1 - y)/2 & u(1, y) &= 1 - y/2 - cy \\
 u(1, y) &> u(0, y) & \text{se e solo se} & \quad 1 - y/2 - cy > 1/2 - y/2, \text{ cioè } y < 1/(2c)
 \end{aligned}$$

siccome $c < 1/2$ e quindi $1/(2c) > 1$, questa condizione é sempre soddisfatta, quindi la miglior risposta del primo giocatore é sempre

1, qualunque sia y . Per la simmetria del gioco G , la miglior risposta del secondo giocatore é sempre 1, qualunque sia x . Ma allora l'intersezione delle 2 curve di miglior risposta si riduce a un unico punto, il punto $(1, 1)$, che é l'unico equilibrio di Nash di G . Si verifica facilmente che si tratta di un equilibrio evolutivo, cioè la strategia $x = 1$ (comportarsi sempre da Falco) é evolutivamente stabile. Quindi in questo caso, come risultato della selezione naturale, dovremmo aspettarci una popolazione di puri Falchi. Se in una popolazione di puri Falchi avviene una mutazione che trasforma alcuni individui diminuendone l'aggressività, cioè facendoli adottare una strategia $x < 1$, i mutanti sono destinati a sparire.

Caso 2 : $c = 1/2$. In tal caso la funzione di utilità del primo giocatore é data da:

$$u_1(x, y) = u(x, y) = (x - y + 1)/2 - xy/2 = (x + 1)(1 - y)/2$$

Fissata y , il massimo di $u(x, y)$ si ha per $x = 1$ se $y < 1$, mentre se $y = 1$ la funzione $u(x, y)$ é identicamente nulla. Quindi la miglior risposta del primo giocatore é 1 se $y < 1$, é tutto l'intervallo $[0, 1]$ se $y = 1$. Analogamente, la miglior risposta del secondo giocatore alla strategia x del primo giocatore é 1 se $x < 1$, é tutto l'intervallo $[0, 1]$ se $x = 1$. Le 2 curve di miglior risposta $R1$ e $R2$ coincidono entrambe con l'unione dei 2 segmenti $(1, 1) - (1, 0)$ e $(1, 1) - (0, 1)$ nel piano xy . Quindi il gioco G ha infiniti equilibri di Nash: tutti i punti del tipo $(x, 1)$ o $(1, y)$, con x e y qualsiasi. Naturalmente fra questi equilibri, solo $(1, 1)$ può essere un equilibrio evolutivo, ed é facile verificare che effettivamente lo é. Anche qui la strategia $x = 1$ (puro Falco) é l'unica ESS e anche qui i mutanti in una popolazione di puri Falchi sono destinati a sparire.

Caso 3 : $c > 1/2$. Ripetendo gli stessi calcoli eseguiti nel caso $c < 1/2$, si trova che la miglior risposta del primo giocatore alla strategia y del secondo é 1 se $y < 1/(2c)$, é 0 se $y > 1/(2c)$, é tutto l'intervallo $[0, 1]$ se $y = 1/(2c)$. Queste tre situazioni sono tutte possibili, perché in questo caso il numero $1/(2c)$ é minore di 1. La curva di miglior risposta $R1$ del primo giocatore é l'unione dei 3 segmenti $(1, 0) - (1, 1/(2c))$, $(0, 1/(2c)) - (1, 1/(2c))$, $(0, 1/(2c)) - (0, 1)$ nel piano xy (fig.1). La curva di miglior risposta $R2$ del secondo giocatore é la simmetrica di $R1$ rispetto alla retta $y = x$ (fig.5.1).

Queste 2 curve s'intersecano nei 3 punti $(0, 1)$, $(1, 0)$, $(1/(2c), 1/(2c))$, che sono gli unici equilibri di Nash del gioco G . I primi 2 sono equilibri in strategie pure, il terzo é un equilibrio in strategie miste. Quest'ultimo equilibrio é l'unico che può essere un equilibrio evolutivo

ed effettivamente lo é, com'è facile verificare. La strategia $x = 1/(2c)$ é l'unica ESS. Quindi in questo caso ci dovremmo aspettare, come risultato della selezione naturale, non una popolazione di puri Falchi né una popolazione di pure Colombe, ma una popolazione di individui tutti uguali che adottano la strategia mista $x = 1/(2c)$, cioè si comportano da Falco con probabilità $1/(2c)$ e da colomba con probabilità $1 - 1/(2c)$ (é una situazione un po' diversa da quella che avevamo visto nell'esempio delle strategie pure, dove si poteva avere una popolazione in equilibrio con una percentuale $1/(2c)$ di puri Falchi e una percentuale $1 - 1/(2c)$ di pure Colombe). Questo dimostra l'importanza delle strategie miste.

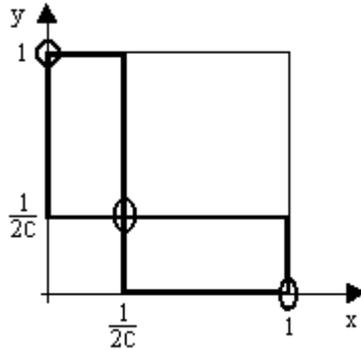


Figura 5.1

Il parametro c che determina il modello (rapporto tra costo della lotta e valore della preda) dipende dall'ambiente e può variare se variano le condizioni ambientali. In particolare, se il numero delle prede diminuisce, é chiaro che il valore di una singola preda aumenta e quindi (supponendo che il costo della lotta rimanga invariato) c diminuisce. Ci si aspetterebbe che questo aumenti la probabilità che gli animali si comportino come Falchi, ma noi supponiamo che gli animali non possano modificare volontariamente il loro comportamento, perché esso é determinato geneticamente. Bisogna aspettare che avvengano per caso delle mutazioni nella popolazione, poi l'ambiente eserciterá la sua influenza sui mutanti, facendoli estinguere o diffondere a seconda dei casi. Vediamo con un esempio cosa può succedere.

Supponiamo che all'inizio il parametro c abbia un valore $c_1 > 1/2$ e che tutta la popolazione si trovi in equilibrio e adotti l'unica ESS $x_1 = 1/(2c_1)$. Supponiamo che in seguito avvenga un cambiamento nell'ambiente: il numero delle prede diminuisca drasticamente (per qualche causa come incendi, alluvioni, malattie, cambiamenti di clima ...) e provochi la diminuzione di c dal valore iniziale c_1 a un valore finale $c_2 < c_1$ (supponiamo però che

$1/2 < c_2 < c_1$, in modo che col nuovo valore c_2 sia possibile un nuovo equilibrio in strategie miste, dato dalla nuova ESS $x_2 = 1/(2c_2)$, con $x_2 > x_1$. Cosa succede?

Intuitivamente si potrebbe pensare che gli animali modifichino il loro comportamento, adottando subito la nuova ESS piú aggressiva $x_2 = 1/(2c_2)$ (anche se non conoscono la teoria dei giochi e non sanno fare i calcoli!). Ma questo non é possibile, perché il loro comportamento é determinato geneticamente. Finché non si verificano mutazioni, gli animali continueranno tranquillamente ad adottare la loro vecchia strategia x_1 , anche se questa non é piú di equilibrio. Supponiamo pero' che ad un certo punto avvenga una mutazione che fa cambiare il comportamento degli animali, provocando la comparsa di una piccola percentuale ε di mutanti, che al posto della vecchia strategia x_1 adottano una nuova strategia x (scelta a caso nell'intervallo $[0, 1]$).

Cosa succede allora? La mutazione viene accettata (cioé la percentuale dei mutanti aumenta ed essi si diffondono nella popolazione) oppure rifiutata (cioé i mutanti tendono ad estinguersi)? Per saperlo occorre vedere se l'utilità attesa u' di un mutante é maggiore o minore dell'utilità attesa u di un non mutante. Per calcolare u, u' osserviamo che per ogni animale la probabilità di incontrare un mutante é ε , la probabilità di incontrare un non mutante é $1 - \varepsilon$. Perciò u, u' sono date da:

$$u = (1 - \varepsilon)u(x_1, x_1) + \varepsilon u(x_1, x)$$

$$u' = (1 - \varepsilon)u(x, x_1) + \varepsilon u(x, x)$$

la mutazione viene accettata se e solo se $u < u'$, cioè:

$$\varepsilon \left(u(x_1, x) - u(x, x) \right) < (1 - \varepsilon) \left(u(x, x_1) - u(x_1, x_1) \right)$$

ma $u(x, y) = (x - y + 1)/2 - c_2xy$ e quindi la mutazione viene accettata se e solo se:

$$\begin{aligned} \varepsilon \left((x_1 - x + 1)/2 - c_2x_1x - 1/2 + c_2x^2 \right) &< (1 - \varepsilon) \left((x - x_1 + 1)/2 - c_2x_1x - 1/2 + c_2x_1^2 \right) \\ \varepsilon \left((x_1 - x)/2 - c_2x(x_1 - x) \right) &< (1 - \varepsilon) \left((x - x_1)/2 - c_2x_1(x - x_1) \right) \\ \varepsilon(x_1 - x)(1/2 - c_2x) &< (1 - \varepsilon)(x - x_1)(1/2 - c_2x_1) \quad (4) \end{aligned}$$

A questo punto distinguiamo 3 casi, a seconda che sia $x < x_1$, $x_1 < x < x_2$, $x > x_2$:

Primo caso: $x < x_1$ (la nuova strategia x si allontana dalla nuova ESS x_2). In questo caso possiamo dividere ambo i membri della (4) per

$x_1 - x > 0$, ottenendo:

$$\begin{aligned} \varepsilon(1/2 - c_2x) &< (1 - \varepsilon)(c_2x_1 - 1/2) \\ \varepsilon/2 - \varepsilon c_2x &< c_2x_1 - 1/2 - \varepsilon c_2x_1 + \varepsilon/2 \\ \varepsilon c_2(x_1 - x) &< c_2x_1 - 1/2 \\ \varepsilon &< (c_2x_1 - 1/2)/(c_2(x_1 - x)) \\ \varepsilon &< (x_1 - 1/(2c_2))/(x_1 - x) \\ \varepsilon &< (x_1 - x_2)/(x_1 - x) \end{aligned}$$

ma poiché $x_1 - x_2 < 0$, questa disuguaglianza non é mai verificata, quindi la mutazione é sempre rifiutata, qualunque sia la percentuale ε dei mutanti. Perció i mutanti tendono ad estinguersi.



Figura 5.2

Secondo caso: $x_1 < x < x_2$ (la nuova strategia x si avvicina alla nuova ESS x_2). In questo caso $x_1 - x < 0$, quindi possiamo fare gli stessi calcoli del caso precedente, con l'avvertenza di cambiare senso alla disuguaglianza ogni volta che dividiamo per $x_1 - x$. Siccome però questa divisione viene fatta 2 volte, alla fine il risultato é lo stesso, cioè (cambiando segno a numeratore e denominatore, che sono entrambi negativi):

$$\varepsilon < (x_2 - x_1)/(x - x_1)$$

poiché il secondo membro é maggiore di 1, questa disuguaglianza é verificata qualunque sia la percentuale ε dei mutanti. Perció la mutazione viene accettata, i mutanti si diffondono nella popolazione e la loro percentuale tende addirittura a $1 = 100\%$.

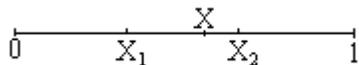


Figura 5.3

Terzo caso: $x > x_2$ (la nuova strategia x oltrepassa la nuova ESS x_2). In questo caso i calcoli sono perfettamente uguali a quelli del secondo caso e quindi si ottiene:

$$\varepsilon < (x_2 - x_1)/(x - x_1)$$

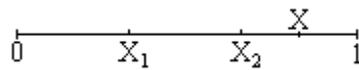


Figura 5.4

peró il secondo membro é un numero positivo minore di 1. Ciò significa che la mutazione viene accettata solo in parte: solo se la percentuale dei mutanti é minore del valore $P = (x_2 - x_1)/(x - x_1)$. All'inizio la percentuale dei mutanti é molto piccola, ma poi cresce e tende a un valore di equilibrio P . Si crea cosí una doppia popolazione: una percentuale P di animali usa la nuova strategia x , una percentuale $1 - P$ continua ad usare la vecchia strategia x_1 .

É chiaro che quando avverrá una seconda mutazione, essa sará accettata o meno a seconda dei casi e ciò provocherá un graduale avvicinamento della strategia della popolazione alla nuova ESS x_2 , ma questo avverrá in un modo discontinuo, molto complicato, con la possibile comparsa di popolazioni multiple.

■

6 Lotte tra animali come giochi evolutivi (da un articolo di “American Scientist” , luglio 1998)

La scienza del comportamento animale a volte si trova di fronte ad alcuni comportamenti paradossali, che sembrano contrari al “buon senso”, ma possono essere spiegati modellizzando la situazione mediante un gioco evolutivo e facendo uso delle strategie evolutivamente stabili (ESS). L’importante é capire “a quale gioco gli animali stanno giocando”.

Consideriamo un famoso esempio, noto come “principio dell’handicap”: a volte gli animali con interessi in conflitto sviluppano dei comportamenti che sono costosi per chi li attua, anche se essi abbassano le sue possibilità di sopravvivenza. Mostrando che può sopportare un handicap, l’animale mette in mostra la sua forza, un messaggio che gli altri animali dovrebbero rispettare. Per esempio, quando una gazzella vede un leone può spiccare un balzo in aria sulle 4 zampe diverse volte prima di scappare, dimostrando così di essere in ottime condizioni fisiche, di modo che il leone sprecherebbe invano le sue forze tentando di inseguirla. Quest’ipotesi fu inizialmente respinta da molti, perché contraddice l’intuizione biologica che l’evoluzione dovrebbe favorire i segnali che costano meno fatica agli animali che li producono. Per risolvere tali questioni può essere utile usare degli strumenti matematici, fra cui la teoria dei giochi evolutivi. Un gioco é un modello matematico dell’interazione strategica, che nasce quando il risultato dell’azione di un individuo dipende dalle azioni degli altri. In un gioco evolutivo giocato all’interno di una singola specie l’insieme dei giocatori é un ecotipo, cioè una popolazione di animali in un dato ambiente.

Ogni giocatore ha un insieme di strategie possibili. In un gioco evolutivo, questo insieme é lo stesso per ogni giocatore ed é vincolato dalla “struttura d’informazione” dell’interazione. Per esempio, gli animali possono modificare il loro comportamento in diverse circostanze, come essere padroni del territorio o intrusi, solo se sono consapevoli di tali ruoli. Le modalità d’interazione devono essere ben definite e accompagnate da una formula che esprime la funzione di utilità di ciascun giocatore, in funzione della sua strategia e di quelle degli altri giocatori. Nei giochi evolutivi l’utilità é misurata in termini di successo riproduttivo futuro.

Perché un gioco sia utile, dev’essere possibile riconoscere una o più fra le strategie possibili come “soluzione” per un certo scopo. Nel nostro caso, la soluzione é il comportamento che ci aspettiamo si sviluppi come conseguenza della selezione naturale. Se un comportamento é fissato in una popolazione reale, allora dev’essere vero almeno che ogni altro comportamento possibile, che compaia per mutazione in una piccola percentuale della popolazione, darebbe un’utilità minore, perché altrimenti il comportamento alternativo si sarebbe diffuso nella popolazione. Il concetto rilevante di soluzione, introdotto da John Maynard Smith, é quello di “strategia evolutivamente

stabile” (ESS), cioè una strategia che dá un’utilitá maggiore di ogni possibile strategia mutante.

Lo studio delle ESS può aiutare a risolvere alcuni apparenti paradossi nel comportamento animale. Un paradosso nasce quando l’evidenza contraddice l’intuizione, il che può succedere solo se l’intuizione si basa su ipotesi false, anche se implicite. Quindi l’unico modo di risolvere un paradosso é smascherare le ipotesi false. In altre parole, se esiste un paradosso nel comportamento di una certa specie di animali, allora noi abbiamo “indovinato” in modo sbagliato il gioco a cui gli animali stanno giocando. Per risolvere il paradosso noi dobbiamo cambiare il modello, piú volte se necessario, finché non troviamo il gioco giusto. Assumendo che il comportamento osservato corrisponda a una ESS, ci sono solo 5 cose che potrebbero essere sbagliate: l’ecotipo, la struttura informativa, l’insieme delle strategie, le modalitá d’interazione e la funzione di utilitá. Ciascuna di queste cose può essere importante, come dimostrano i seguenti 3 esempi.

Esempio 6.1 Gioco della guerra di logoramento (farfalle)

J.Marden e J.Waage studiarono una serie di lotte territoriali tra farfalle maschio della specie *Calopteryx Maculata*. Le farfalle avevano, come indicatori di forza, delle riserve di grasso variabili, diverse da individuo a individuo. Per queste lotte ci si aspetta che ogni animale paragoni la sua forza con quella dell’avversario e si ritiri quando pensa di perdere. Ciò é stato confermato da studi sperimentali su molte specie animali. La durata della lotta é piú lunga quando i contendenti hanno circa la stessa forza, cosicché é difficile giudicare chi vincerá.

Ma i duelli tra farfalle non seguono questa logica. Sebbene l’animale piú debole alla fine ceda al suo avversario in oltre il 90% dei casi, non c’ é nessuna correlazione negativa tra la durata della lotta e la differenza di forza tra i 2 contendenti. Questo é un paradosso, tanto piú sconcertante in quanto la relazione attesa si verifica invece in altre specie animali. Cosa c’ é di sbagliato nelle nostre ipotesi?

Una possibile ipotesi errata é semplicemente che le farfalle possano valutare la forza una dell’altra. Questa di solito é una buona ipotesi per animali con la vista acuta (come i cervi) e spesso viene assunta senza problemi nello studio delle lotte fra animali. Ma negli insetti il grasso é immagazzinato internamente, quindi una farfalla non può vedere direttamente le riserve del suo avversario (Marden non ha trovato nessuna correlazione tra le riserve di grasso e caratteristiche direttamente osservabili come la lunghezza del corpo o l’apertura alare). Può darsi che le farfalle non possano in alcun modo valutare la forza dell’avversario.

Gibbons, in una collaborazione con Marden, sviluppó un modello in cui un’ipotesi chiave é che i giocatori conoscono solo la propria forza. Il gioco risultante é noto come “guerra di logoramento” e una strategia é la percentuale delle sue riserve iniziali che l’animale é disposto a spendere in una

lotta prolungata per un certo sito. Una seconda ipotesi chiave è che il valore della vincita è proporzionale alle riserve di grasso rimanenti al vincitore. Più riserve un animale risparmia, più avrà successo nell'attrarre una compagna, nel trovare cibo o nel difendere il suo territorio (e quindi nel riprodursi). Queste ipotesi determinano l'insieme delle strategie e la formula della funzione di utilità.

Ma per quanto riguarda l'ecotipo? In generale un modello è soltanto una "caricatura" della natura, cioè una sua semplificazione estrema, in quanto riduce la complessità di un ambiente reale a pochi parametri numerici. Nel gioco della guerra di logoramento ci sono solo 2 parametri, ognuno dei quali è un numero compreso fra 0 e 1. Il primo, il coefficiente di variazione C , misura la dispersione delle riserve di energia intorno alla loro media. Il secondo parametro, il rapporto costi - benefici R , paragona il costo del perdente di spendere 1 unità di riserve di grasso con l'eventuale beneficio del vincitore per 1 unità risparmiata. Ciò significa che il valore delle riserve di energia si riduce di un fattore R passando dal vincitore al perdente. Nella realtà R varia da individuo a individuo, ma nel modello assumiamo R uguale per tutte le farfalle.

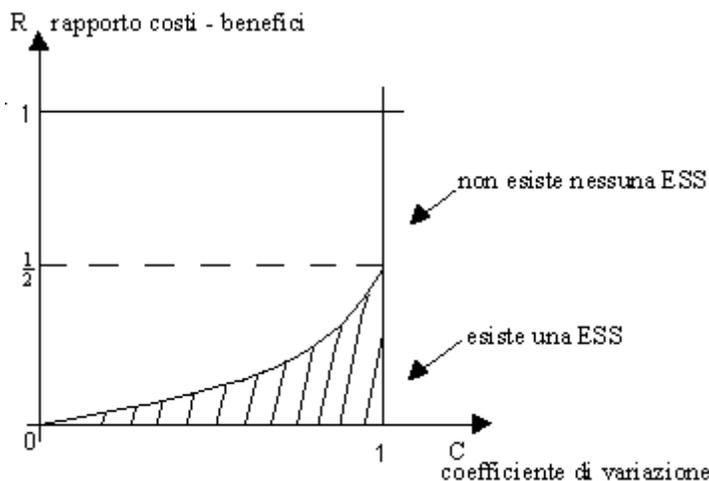


Figura 6.1

Quando ciascuna componente del gioco è stata identificata, possiamo cercare le ESS. In generale, i parametri del modello determinano se una ESS esiste o meno. In questo caso una ESS esiste se il rapporto costi - benefici R supera una certa soglia critica $f(C)$, che dipende dal coefficiente di variazione C e cresce con C (fig.6.1). Se una ESS esiste, è data da una certa strategia $x^* = g(C, R)$ che dipende dai parametri del modello (vedremo fra poco la

forma delle funzioni f, g). Se non esiste una ESS, allora ogni strategia $x > 0$ può essere invasa da mutanti che adottano la strategia $x = 0$ (cioè fuggono subito senza combattere).

La ESS x^* (quando esiste) è sempre maggiore di $2/3$, cioè le farfalle combattono sempre fin quasi all'esaurimento (continuando a svolazzare intorno al sito prescelto, senza contatto fisico), consumando almeno il 66.6% delle proprie energie. Anche questo è un comportamento paradossale, tanto più che si dimostra che la ESS x^* è sempre inefficiente, cioè se le farfalle adottassero tutte la strategia $x = 0$ otterrebbero un'utilità maggiore (ma naturalmente, se c'è solo una piccola percentuale di mutanti con $x = 0$ in una popolazione con $x = x^*$, i mutanti ottengono un'utilità minore dei non mutanti, altrimenti x^* non sarebbe una ESS).

Si può descrivere a parole la ragione per la stabilità della ESS. Essere disposti a spendere una percentuale troppo piccola delle riserve iniziali può significare cedere senza necessità a un avversario più debole (che altrimenti perderebbe), mentre una percentuale troppo alta può significare sprecare inutilmente delle risorse contro un avversario più forte (che vincerebbe comunque). Sebbene le farfalle non conoscano la forza dell'avversario, possono raggiungere un equilibrio tra queste 2 opposte esigenze adeguandosi alla distribuzione delle riserve nella popolazione.

In seguito Marden osservò altri duelli tra farfalle in collaborazione con Rollins e mise in comune con lui i dati. Sebbene la percentuale delle vittorie delle farfalle più grasse diminuisse da 90% a 86%, rimase molto alta. Ora, le approssimazioni ragionevoli sono la base dei modelli matematici (effetti che sono piccoli in una popolazione reale di solito vengono trascurati nel modello). Per esempio, potremmo aspettarci che l'86% delle vittorie delle farfalle più grasse si traduca in un 100% nel modello. E questo è proprio quel che accade, perché entrambi i contendenti sono disposti a spendere la stessa percentuale delle riserve iniziali. La farfalla più magra inevitabilmente si arrende prima.

Così le ipotesi della conoscenza e della non conoscenza (della forza dell'avversario) prevedono entrambe che la farfalla più grassa vince sempre. Ma c'è anche una differenza. L'ipotesi della conoscenza prevede una correlazione negativa tra la differenza di forza e la durata della lotta. Anche se, nei dati in comune, Marden trovò una tale correlazione nelle lotte più lunghe di 500 secondi, la variazione della differenza di forza poteva spiegare solo il 14% della variazione di durata della lotta. Invece, nel modello di non conoscenza, una lotta finisce quando il perdente se ne va dopo aver usato una percentuale fissa delle sue riserve. Quindi l'ipotesi di non conoscenza prevede una correlazione positiva tra le riserve finali del perdente e la durata della lotta. E, nelle lotte più lunghe di 500 secondi, si è visto che la variazione delle riserve del perdente spiega il 29% della variazione di durata della lotta.

Sarebbe attraente dedurre da questi risultati che l'ipotesi della non conoscenza è 2 volte più verosimile dell'ipotesi della conoscenza. Ma in realtà i risul-

tati sono inconcludenti, perché il 14% e il 29% sono entrambi molto minori del 100%. Ciò nonostante, i tentativi di risolvere il paradosso iniziale ci permettono di capire meglio il comportamento degli animali, cioè di capire che la vittoria del più forte non significa che le farfalle conoscono la forza dell'avversario. Inoltre si vede che il tipo di popolazione in cui ci possiamo aspettare una ESS nell'ipotesi della non conoscenza è una popolazione con rapporto costi - benefici R basso e coefficiente di variazione C alto. ■

Esempio 6.2 Gioco della minaccia (granchi)

Adams e Caldwell osservarono una serie di lotte tra stomatopodi (granchi) della specie *Gonodactylus Bredini*. Questi crostacei abitano in cavità della barriera corallina. Se uno invade la cavità di un altro, il residente spesso reagisce minacciando l'avversario con delle specie di artigli. Queste minacce spesso spaventano l'intruso, senza che ci sia nessun contatto fisico tra i 2 avversari. Un'osservazione sorprendente è che quando i granchi sono deboli e incapaci di combattere, minacciano più spesso di quando sono forti. Inoltre queste minacce spaventano degli intrusi molto più forti, che vincerebbero facilmente un combattimento, se ce ne fosse uno. In altre parole, il residente debole attua un "bluff". Ma se i granchi più deboli minacciano con successo, perché non lo fanno tutti? Se la minaccia viene fatta da animali che non possono attuarla, perché i loro avversari la rispettano?

Per spiegare il paradosso, consideriamo un gioco in cui il residente può minacciare o non minacciare e l'intruso può rispondere attaccando o fuggendo. Noi supponiamo che la forza degli animali vari nella popolazione e sia distribuita uniformemente tra 0 e 1. Introduciamo anche una funzione di utilità che è diversa da quella assunta da un modello di minaccia "onesta". Specificamente, la minaccia accresce la vulnerabilità del residente al danno che l'avversario gli farà (se ci sarà lotta). La minaccia non costa niente se il residente non è attaccato, ma ha un costo aggiuntivo, il "costo della minaccia" se il residente viene attaccato e perde. Inoltre se c'è lotta entrambi i contendenti pagano un "costo della lotta", che consiste in un "costo fisso", pagato anche dal vincitore, e in un "costo marginale", che cresce con la debolezza a un tasso costante. Il costo della lotta varia tra il costo fisso per un animale con forza 1 e il costo fisso + il costo marginale per un animale con forza 0.

Inoltre supponiamo che, se c'è lotta, l'animale più forte vinca sempre. Poiché la forza dei granchi non è visibile dall'esterno, assumiamo che (come nel gioco della guerra di logoramento) i granchi non conoscano la forza dell'avversario, e quindi il loro comportamento sia determinato solo dalla propria forza. La decisione del residente di minacciare o no precede la decisione dell'intruso di attaccare o scappare. Un sito ha lo stesso valore (posto uguale a 1 per convenzione) per tutti gli animali. La forza (sconosciuta) di ciascuno dei 2

avversari é presa a caso dalla distribuzione uniforme della popolazione. Per questo particolare gioco esiste sempre una e una sola ESS, che dipende dai parametri del modello (vedi fig.6.2, dove le strategie sono graficate in funzione del costo fisso, per valori costanti del costo marginale e del costo della minaccia). Se il costo marginale fosse 0, oppure se il costo della minaccia fosse pagato quando il residente é attaccato (anche se vince), non esisterebbe nessuna ESS. Quindi la ESS é una conseguenza di una funzione di utilitá molto speciale, identificata dal modello.

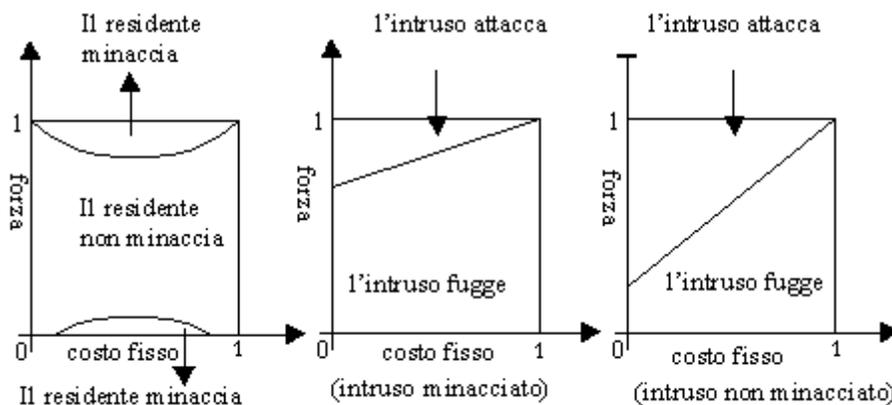


Figura 6.2

Secondo la ESS, gli animali piú deboli e piú forti minacciano entrambi (quando sono residenti e sono attaccati), mentre quelli di forza intermedia non minacciano (fig.6.2, sinistra). Perché questo comportamento paradossale, che é essenzialmente un esempio del cosiddetto “principio dell’handicap” ? Perché gli animali piú deboli non hanno altri mezzi per sopraffare l’avversario e perderebbero sicuramente se ci fosse lotta, mentre gli animali piú forti possono minacciare tranquillamente, perché é difficile che perdano e quindi paghino il costo della minaccia. Invece gli animali di forza intermedia non minacciano, perché il rischio di perdere la lotta supera i benefici della minaccia.

Per quanto riguarda la strategia dell’intruso di attaccare o scappare, esso attaccherà se la sua forza supera un certo valore soglia, che dipende dal costo della lotta e dal fatto che sia minacciato (fig.6.2, centro) oppure no (fig.6.2, destra). Il valore soglia é piú alto quando il residente minaccia. ■

Esempio 6.3 Gioco falchi - colombe ripetuto (ragni)

Le lotte tra animali di forza diversa sono di solito vinte dal piú forte. Ma che

dire delle lotte tra animali di forza uguale? Una possibilità nelle lotte per le risorse o per il territorio é quella di essere proprietario o intruso. In molti casi l'evidenza del comportamento animale mostra che gli animali sono consci del loro ruolo di proprietario o intruso. Quando lottano per una risorsa permanente, di solito il proprietario la mantiene con una lotta molto breve e leggera. Non c'è niente di paradossale in questo rispetto per la proprietà. Ma non sarebbe sorprendente se la lotta fosse breve e leggera perché a scappare é il proprietario e non l'intruso? Ebbene, questo é quel che accade in una specie di ragno messicano, *Oecobius Civitas*.

Secondo uno studio di Burgess, questi piccoli ragni vivono nell'oscurità, negli anfratti delle rocce. Se un ragno é disturbato nel suo nascondiglio, se ne va ed entra nel nascondiglio di un altro, il quale, se é in casa, non attacca l'intruso, ma se ne va subito a cercare un altro nascondiglio. Ne segue un "effetto domino", con successivi cambiamenti di proprietario, finché alla fine un ragno trova un buco vuoto e ci si infila. Questo comportamento é paradossale perché, assumendo che il nascondiglio dia un vantaggio, il premio va sempre a un intruso che potrebbe perdere la lotta, se ce ne fosse una. Come spiegare il paradosso?

Come al solito, quando un comportamento é fisso, ci aspettiamo di poterlo spiegare come la ESS di un gioco evolutivo, dove gli effetti che sono piccoli nella popolazione sono assenti dal modello. Così si può spiegare il comportamento di molte specie animali come una ESS in cui il proprietario vince sempre (invece che quasi sempre), mentre il comportamento di *Oecobius Civitas* si può spiegare come una ESS in cui il proprietario perde sempre. In entrambi i casi ci aspettiamo che le lotte siano assenti (anziché rare). Cioé, secondo la terminologia di Maynard Smith, ci aspettiamo che la maggioranza delle specie animali segua una strategia B (attaccare come proprietario, non attaccare come intruso), mentre invece *Oecobius Civitas* segue una strategia anti- B (non attaccare come proprietario, attaccare come intruso). Invece gli animali che attaccano sempre seguono la strategia Falco, quelli che non attaccano mai seguono la strategia Colomba. Queste 4 strategie sottintendono una struttura dell'informazione: gli animali devono conoscere i loro ruoli.

Usando un semplice gioco con queste 4 strategie, Maynard Smith mostrò che solo la strategia Falco é ESS se i costi sono abbastanza bassi, mentre le strategie B e anti- B sono ESS se i costi sono abbastanza alti. Ma quando esistono 2 ESS, la popolazione può trovarsi in una sola di esse, e il gioco di Maynard Smith non dice quale.

Gibbons ha incorporato l'insieme di strategie di Maynard Smith in un modello piú elaborato. Ogni animale ha un numero finito di periodi in cui trovare un sito, che é essenziale per la riproduzione. Non importa se il sito viene trovato presto o tardi, basta che sia posseduto prima della fine del periodo. Il numero dei periodi é sconosciuto, ma tutti gli animali hanno la stessa probabilità di sopravvivere alla predazione in ciascun periodo. Ci sono piú siti che animali, ma la ricerca é casuale. Perciò un animale può cercare

invano, e anche se trova un sito può essere occupato. In tal caso, l'animale può continuare a cercare oppure attaccare. Se vince, diventa proprietario e il precedente proprietario deve cercare ancora.

C'è comunque un guadagno in tutto ciò. Quando avviene una dura lotta, il perdente è ferito così gravemente che non può più cercare, e quindi non può riprodursi. Anche il vincitore può essere ferito e avere un successo riproduttivo minore di un animale sano. Quindi c'è un contrasto tra l'incertezza della ricerca e il rischio di essere feriti. La probabilità del proprietario di vincere una lotta si chiama "potenziale di mantenimento delle risorse" ed è un numero P compreso tra 0.5 e 1, perché è più facile che il proprietario vinca.

Se un animale combatte o no dipende dalla sua strategia. Un proprietario B e un intruso anti- B , o 2 Falchi, combattono sempre. Un proprietario anti- B e un intruso B , o 2 Colombe, non combattono mai, un Falco scaccia sempre una Colomba. Ne risulta un modello coi seguenti 6 parametri:

1. Probabilità S di sopravvivere alla predazione per un periodo di ricerca del sito.
2. Probabilità P del proprietario di vincere una lotta (potenziale di mantenimento delle risorse). Supponiamo $P > 0.5$.
3. Probabilità di trovare un sito in ciascun periodo di ricerca.
4. Rapporto tra numero di animali e numero di siti.
5. Probabilità di essere feriti quando si vince una lotta.
6. Rapporto tra il valore del sito per un proprietario ferito e per un proprietario sano.

Supponiamo anche che un sito abbia lo stesso valore (posto uguale a 1 per convenzione) per tutti gli animali e che tutti gli animali che cercano abbiano la stessa probabilità di trovare un sito qualsiasi.

Questo gioco differisce dagli altri 2 perché può avere ESS multiple. Le ESS dipendono da tutti e 6 i parametri, ma per spiegare il nostro paradosso conta soprattutto la loro dipendenza da P e da S , che è molto complicata (fig 6.3).

Se S è abbastanza bassa, solo la strategia Falco è ESS. Nonostante il rischio di essere feriti, è meglio combattere che sperare di trovare un buco libero. Se invece S è alta, o la strategia B o la anti- B sono ESS. Per vedere perché, supponiamo che P rimanga costante, ma che la predazione decresca lentamente nel tempo (quindi S cresce) e che la popolazione si adegui a questo mutamento nell'ambiente evolvendosi continuamente verso una nuova ESS. Allora noi pensiamo alla popolazione come a un punto che si muove orizzontalmente verso destra in fig.6.3. Quando il punto migra, alla fine la strategia

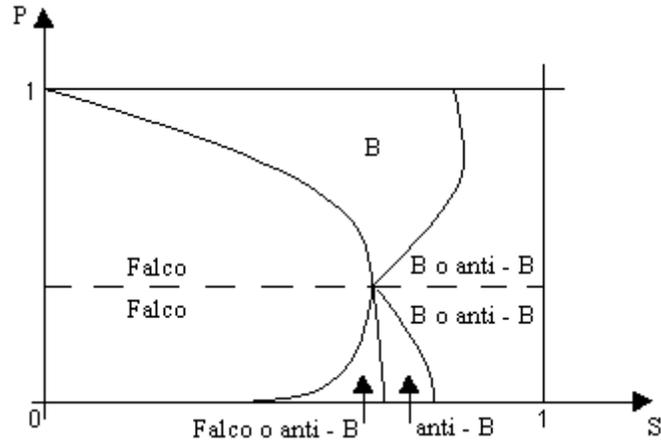


Figura 6.3

Falco cessa di essere una ESS.

Perché? Consideriamo un mutante B in una popolazione di Falchi. Poiché B si comporta come un Falco a casa sua, il fatto di essere proprietario non ha effetti sulla sua utilità. Quindi un mutante B ha un'utilità maggiore del resto della popolazione se un intruso B è favorito su un intruso Falco. Ma un intruso B scappa da un proprietario Falco, sperando di trovare alla fine un buco vuoto, mentre un intruso Falco combatte, sperando di occupare il buco subito. Se S e p sono abbastanza alte, tuttavia, la prima speranza è più facile da realizzare della seconda. Quindi il B è avvantaggiato sul Falco e si diffonde nella popolazione.

Similmente, un mutante anti- B in una popolazione di Falchi si comporta diversamente solo come proprietario. Esso scappa da un intruso Falco, sperando di trovare alla fine un buco libero, mentre un proprietario Falco combatte, sperando di rimanere nel suo buco. Se S è abbastanza alta e P abbastanza bassa, la prima speranza si realizza più facilmente della seconda, quindi l'anti- B è favorito sul Falco e si diffonde nella popolazione.

Riassumendo: se la predazione decresce, una popolazione di Falchi si evolve in B se P è superiore a un certo valore critico P_0 (linea tratteggiata in fig. 6.3), si evolve in anti- B se $P < P_0$ (pur essendo sempre $P > 0.5$). Se la predazione cala abbastanza, la popolazione rimarrà alla ESS B o anti- B a seconda del valore iniziale di P . In entrambi i casi, da una popolazione di combattenti si evolve una popolazione di non combattenti.

Poiché una bassa predazione accresce la probabilità che non combattere e cercare permetta alla fine di trovare un sito senza danno, questi risultati concordano con quelli di Maynard Smith. Ma questo nuovo modello va oltre e suggerisce una soluzione del nostro paradosso. Se esiste un'asimmetria di comportamento dovuta al fatto di essere proprietario, allora P è quasi

sempre abbastanza alta da portare la popolazione in una strategia B , se la predazione é bassa. Tuttavia, in alcuni rari ambienti P é abbastanza bassa da portare la popolazione in una strategia anti- B . Il ragno *Oecobius Civitas* rappresenta una tale eccezione? Forse. Ma senza una prova empirica che P é davvero bassa per *Oecobius Civitas*, questa rimane nulla piú che una congettura. ■

6.1 Il valore della teoria dei giochi

Spesso i modelli della teoria dei giochi rimangono parzialmente inconcludenti, tuttavia hanno il merito di suggerire modi di verificare nuove idee. Il modello delle lotte tra farfalle suggerisce un test per una ESS in cui le farfalle non conoscono la forza dell'avversario. Il modello dei granchi suggerisce un test per una ESS con un "bluff" parziale. Infine, il modello dei ragni suggerisce un test per una ESS *Anti - B*. Però suggerire un test non é la stessa cosa che eseguirlo e la difficultá di eseguirlo non deve essere sottovalutata. Per esempio, l'unico modo di misurare il potenziale di mantenimento delle risorse (P) dei ragni *Oecobius Civitas* é osservare molte lotte, il che non é per niente facile per una specie che vive nascosta come *Oecobius Civitas* (come spesso succede non si riesce a cavare un ragno dal buco!)

La difficultá di verificare le previsioni della teoria dei giochi evolutivi ha indotto qualcuno a mettere in dubbio il suo valore. Ma i giochi non hanno valore solo perché suggeriscono il modo di verificare nuove idee. Essi hanno valore anche perché ci permettono di verificare rigorosamente, con calcoli precisi la logica di argomentazioni puramente verbali, determinando se sono vere o false. Come dimostrano gli esempi visti, spesso la teoria dei giochi dimostra cose che sono difficili da intuire. Studiando i duelli delle farfalle, abbiamo capito che la vittoria dell'animale piú forte non significa che le farfalle conoscono la forza dell'avversario. Studiando le lotte tra granchi, abbiamo capito come il "bluff" puó verificarsi con alta frequenza. E analizzando le contese tra ragni, abbiamo visto che una popolazione di usurpatori non combattenti puó evolversi (se la predazione decresce) da una popolazione di combattenti che ignora l'assimetria proprietario - intruso.

Forse la non conoscenza della forza dell'avversario, il bluff sistematico ed i ripetuti cambiamenti di sito sono tutti comportamenti molto rari in natura. Per esempio "l'effetto domino" non é stato osservato in nessun'altra specie di ragni *Oecobius Civitas*. Ma é raro che i comportamenti normali attirino la nostra attenzione: sono piú interessanti i comportamenti strani, che stimolano nuove e piú profonde indagini. In questo senso, la teoria dei giochi evolutivi si é dimostrata spesso molto utile per la scienza del comportamento animale.

7 Il gioco della guerra di logoramento

Alcuni anni fa Marden e Waage studiarono una serie di lotte territoriali fra farfalle maschio della specie *Calopteryx Maculata*.

Di solito nelle lotte tra animali ognuno dei 2 contendenti paragona la sua forza a quella dell'avversario e si ritira quando pensa di perdere. Ciò é stato confermato da studi sperimentali su molte specie animali. La durata della lotta é piú lunga quando i contendenti hanno la stessa forza cosicché é difficile capire chi vincerá. Ma i duelli tra farfalle non seguono questa logica. Marden non trovó nessuna correlazione tra la durata della lotta e la differenza di forza tra le farfalle (la forza delle farfalle é misurata dalle loro riserve di grasso, che sono usate come fonte di energia).

Questo comportamento, apparentemente paradossale puó essere spiegato semplicemente ammettendo che le farfalle non conoscano la forza una dell'altra.

Nel cosiddetto "gioco della guerra di logoramento" i giocatori conoscono solo la propria forza e una strategia é la percentuale delle riserve di energia iniziali che l'animale é disposto a spendere in una lotta prolungata per conquistare un sito. Piú risorse un animale risparmia, piú avrá successo nell'attrarre una compagna, nel trovare cibo o nel difendere il suo territorio, e quindi, nel riprodursi. Queste ipotesi ci permettono di individuare lo spazio delle strategie e la formula della funzione di utilitá.

Se vogliamo costruire un modello nell'esempio delle farfalle abbiamo bisogno di 2 parametri C , R , entrambi numeri reali appartenenti all'intervallo $[0, 1]$. C (coefficiente di variazione) misura la dispersione delle riserve di energia intorno alla loro media (posta uguale a 1 per convenzione), cioè misura la variabilitá della forza delle farfalle. Se C é grande, significa che c'è una grande varietá di farfalle.

R (rapporto costi - benefici) é il rapporto tra il valore di un'unitá di energia per il perdente e il suo valore per il vincitore, cioè paragona il costo della lotta al beneficio di conquistare il sito, entrambi in termini di successo riproduttivo futuro. Se R é piccolo, significa che il valore del sito é elevato. Nella realtá R varia da individuo a individuo, ma noi supponiamo che sia uguale per tutti.

Schematizziamo il gioco cosí: $G = (A, B, S, S, u_1, u_2)$ dove A, B sono le 2 farfalle che lottano, $S = [0, 1]$ é lo spazio delle strategie, uguale per i 2 giocatori (la strategia di una farfalla é la percentuale di energia che essa é disposta a spendere prima di abbandonare la lotta). Siano x, y le strategie delle 2 farfalle e $u(x, y)$ la funzione di utilitá di una farfalla se essa usa la strategia x , mentre l'avversaria usa y . Allora, $u_1(x, y) = u(x, y)$ e $u_2(x, y) = u(x, y)$. Indichiamo con e_1, e_2 le riserve di energia iniziali delle 2 farfalle. Supponiamo che esse siano distribuite uniformemente nella popolazione delle farfalle, tra un minimo $1 - C$ e un massimo $1 + C$, indichiamo cioè con 1 l'energia media e con C la semiampiezza dell'intervallo di variazione (chiaramente C sta in $[0, 1]$).

L'utilità $u(x, y)$ é calcolata come utilità attesa, cioè come utilità media della farfalla A al variare della sua energia e_1 e di quella dell'avversaria e_2 , in tutti i modi possibili nella popolazione. Se supponiamo di conoscere e_1, e_2 , l'utilità di ciascuna farfalla dipende dal fatto che essa vinca o perda la lotta. Supponiamo che A e B consumino energia con la stessa velocità e indichiamo con $v(x, y, e_1, e_2)$ l'utilità della farfalla A conoscendo i valori di e_1, e_2 . Poiché A é disposta a spendere un'energia xe_1 e B é disposta a spendere un'energia ye_2 , si ha:

- A vince se $xe_1 > ye_2$
- A perde se $xe_1 < ye_2$
- A vince con probabilità $1/2$ se $xe_1 = ye_2$

$$v(x, y, e_1, e_2) = e_1 - ye_2 \quad \text{se } A \text{ vince, ovvero se } xe_1 > ye_2$$

$$v(x, y, e_1, e_2) = Re_1(1 - x) \quad \text{se } A \text{ perde, ovvero se } xe_1 < ye_2$$

$$v(x, y, e_1, e_2) = (R + 1)e_1(1 - x)/2 \quad \text{se } xe_1 = ye_2$$

Per calcolare $u(x, y)$ dobbiamo fare una media pesata di $v(x, y, e_1, e_2)$ al variare di e_1, e_2 in $[1 - C, 1 + C]$.

La probabilità che A abbia un'energia compresa nell'intervallo infinitesimo $[e_1, e_1 + de_1]$ e B abbia un'energia in $[e_2, e_2 + de_2]$ é $de_1/de_2/(4C^2)$ (perché la probabilità totale deve valere 1; $4C^2$ é l'area del quadrato $Q = [1 - C, 1 + C]^2$). Quindi, $u(x, y)$ dev'essere calcolata come integrale doppio:

$$u(x, y) = 1/(4C^2) \int \int_Q v(x, y, e_1, e_2) de_1 de_2$$

Data l'espressione di $v(x, y, e_1, e_2)$ dobbiamo distinguere 4 casi, a seconda di come é disposta la retta $e_2 = (x/y)e_1$ rispetto al quadrato Q nel piano cartesiano con assi e_1, e_2 (A vince sotto la retta, perde sopra la retta):

- CASO 1: $x/y \leq (1 - C)/(1 + C)$. Allora la retta sta tutta sotto il

quadrato (A perde sempre) e quindi, si ha:

$$\begin{aligned}
 u(x, y) &= \frac{1}{4C^2} \int \int_Q R e_1 (1-x) d e_1 d e_2 \\
 &= R \frac{(1-x)}{4C^2} \int_{1-C}^{1+C} e_1 d e_1 \int_{1-C}^{1+C} d e_2 \\
 &= R \frac{(1-x)}{4C^2} 2C \int_{1-C}^{1+C} e_1 d e_1 \\
 &= R \frac{(1-x)}{2C} \left[\frac{e_1^2}{2} \right]_{1-C}^{1+C} \\
 &= \frac{R(1-x)}{2C} \frac{(1+C)^2 - (1-C)^2}{2} \\
 &= \frac{R(1-x)}{2C} \frac{1 + 2C + C^2 - 1 + 2C - C^2}{2} \\
 &= R(1-x)
 \end{aligned}$$

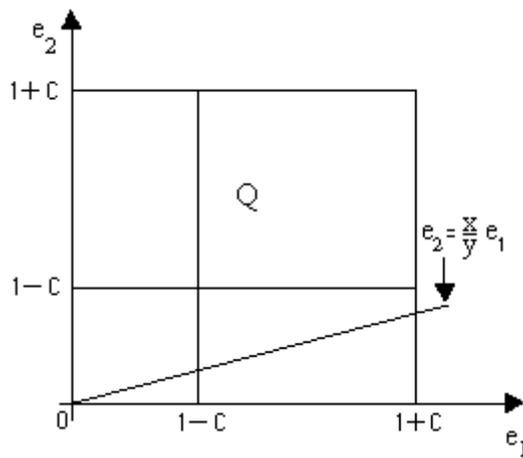


Figura 7.1

- CASO 2: $x/y \geq (1-C)/(1+C)$. Allora la retta sta tutta sopra il

quadrato (A vince sempre) e quindi, si ha:

$$\begin{aligned}
 u(x, y) &= \frac{1}{4C^2} \int \int_Q (e_1 - ye_2) de_1 de_2 \\
 &= \frac{1}{4C^2} \int_{1-C}^{1+C} e_2 de_1 \int_{1-C}^{1+C} de_2 - \frac{y}{4C^2} \int_{1-C}^{1+C} e_2 de_2 \int_{1-C}^{1+C} de_1 \\
 &= (\text{con calcoli analoghi al caso 1}) \frac{1}{4C^2} 4C^2 - \frac{y}{4C^2} 4C^2 \\
 &= 1 - y
 \end{aligned}$$

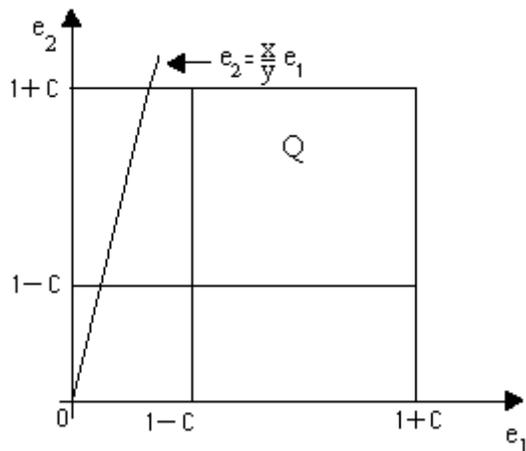


Figura 7.2

- CASO 3: $(1+C)/(1-C) \leq x/y \leq 1$. Allora la retta taglia il quadrato come in figura (dove le coordinate dei punti A, B sono rispettivamente $((1-C)y/x, 1-C)$ e $(1+C, (1+C)y/x)$) e si ha:

$$\begin{aligned}
 u(x, y) &= \frac{1}{4C^2} \int \int_Q Re_1(1-x) de_1 de_2 + \frac{1}{4C^2} \int \int_T (e_1 + \\
 &\quad -ye_2 - Re_1 + Re_1x) de_1 de_2 \\
 &= R(1-x) + \frac{1-R+Rx}{4C^2} \int \int_T e_1 de_1 de_2 - \frac{y}{4C^2} \int \int_T e_2 de_1 de_2 \\
 &= R(1-x) + \frac{1-R+Rx}{4C^2} a(x, y) - \frac{y}{4C^2} b(x, y)
 \end{aligned}$$

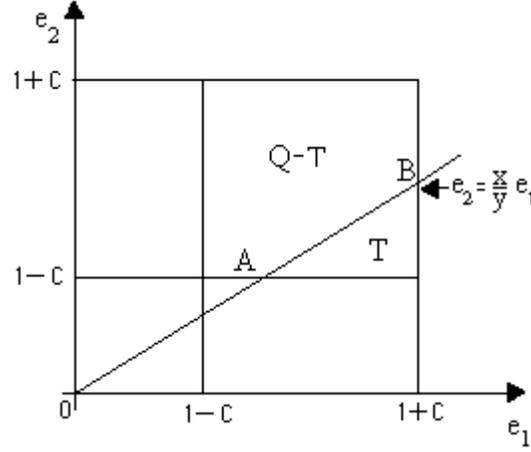


Figura 7.3

Avendo posto:

$$\begin{aligned}
 a(x, y) &= \int \int_T e_1 de_1 de_2 = \int_{\frac{y}{x}(1-C)}^{1+C} de_1 \int_{1-C}^{\frac{x}{y}e_1} e_1 de_2 = \int_{\frac{y}{x}(1-C)}^{1+C} e_1 de_1 \left[\frac{x}{y} e_1 - (1-C) \right] \\
 &= \left[\frac{x}{y} \frac{e_1^3}{3} - (1-C) \frac{e_1^2}{2} \right]_{e_1=\frac{y}{x}(1-C)}^{e_1=1+C} \\
 &= \frac{x}{y} \frac{(1+C)^3}{3} - \frac{(1-C)(1+C)^2}{2} - \frac{x}{3y} \frac{y^3}{x^3} (1-C)^3 + \frac{1-C}{2} \frac{y^2}{x^2} (1-C)^2 \\
 &= \frac{y^2}{6x^2} (1-C)^3 + \frac{x}{3y} (1+C)^3 - \frac{(1-C)(1+C)^2}{2}
 \end{aligned}$$

$$\begin{aligned}
 (x, y) &= \int \int_T e_2 de_1 de_2 = \int_{\frac{y}{x}(1-C)}^{1+C} de_1 \int_{1-C}^{\frac{x}{y}e_1} e_2 de_2 = \int_{\frac{y}{x}(1-C)}^{1+C} de_1 \left[\frac{e_2^2}{2} \right]_{e_2=1-C}^{e_2=\frac{x}{y}e_1} \\
 &= \int_{\frac{y}{x}(1-C)}^{1+C} \left[\frac{x^2 e_1^2}{y^2} - \frac{(1-C)^2}{2} \right] de_1 = \left[\frac{x^2 e_1^3}{2y^2} - \frac{(1-C)^2}{2} e_1 \right]_{e_1=\frac{y}{x}(1-C)}^{e_1=1+C} \\
 &= \frac{x^2}{6y^2} (1+C)^3 - \frac{(1-C)^2 (1+C)}{2} - \frac{x^2}{6y^2} \frac{y^3}{x^3} (1-C)^3 + \frac{(1-C)^3}{2} \frac{y}{x} \\
 &= \frac{x^2}{6y^2} (1+C)^3 + \frac{y}{3x} (1-C)^3 - \frac{(1+C)(1-C)^2}{2}
 \end{aligned}$$

- CASO 4: $1 \leq x/y \leq (1+C)/(1-C)$. Allora la retta taglia il quadrato come in figura (dove le coordinate dei punti C, D sono rispettivamente

$((1+C)y/x, 1+C)$ e $(1-C, (1-C)x/y)$ e si ha:

$$\begin{aligned}
 u(x, y) &= \int \int_V \frac{Re_1(1-x)}{4C^2} de_1 de_2 + \int \int_{Q-V} \frac{e_1 - ye_2}{4C^2} de_1 de_2 \\
 &= \int \int_Q \frac{e_1 - ye_2}{4C^2} de_1 de_2 + \int \int_V \frac{Re_1 - Re_1x - e_1 + ye_2}{4C^2} de_1 de_2 \\
 &= 1 - y + \frac{R - Rx - 1}{4C^2} \int \int_V e_1 de_1 de_2 + \frac{y}{4C^2} \int \int_V e_2 de_1 de_2 \\
 &= 1 - y + \frac{R - Rx - 1}{4C^2} c(x, y) + \frac{y}{4C^2} d(x, y)
 \end{aligned}$$

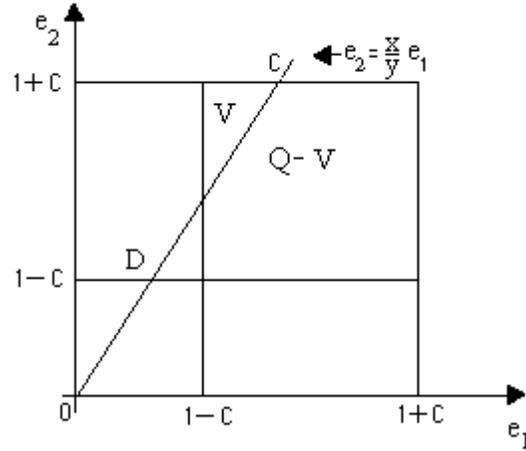


Figura 7.4

$$\begin{aligned}
 c(x, y) &= \int \int_V e_1 de_1 de_2 = \int_{1-C}^{\frac{y}{x}(1+C)} de_1 \int_{\frac{x}{y}e_1}^{1+C} e_1 de_2 = \int_{1-C}^{\frac{y}{x}(1+C)} e_1 de_1 [1 + C - \frac{x}{y}e_1] \\
 &= [(1+C)\frac{e_1^2}{2} - \frac{x}{y}\frac{e_1^3}{3}]_{e_1=(1-C)}^{e_1=\frac{y}{x}(1+C)} = \frac{(1+C)^3 y^2}{2 x^2} - \frac{x}{3y} \frac{y^3}{x^3} (1+C)^3 - \frac{(1+C)(1-C)^2}{2} + \\
 &+ \frac{x}{3y} (1-C)^3 = \\
 &= \frac{y^2}{6x^2} (1+C)^3 + \frac{x}{3y} (1-C)^3 - \frac{(1+C)(1-C)^2}{2}
 \end{aligned}$$

$$\begin{aligned}
d(x, y) &= \int \int_V e_2 de_1 de_2 = \\
&= \int_{1-C}^{\frac{y}{x}(1+C)} de_1 \int_{\frac{x}{y}e_1}^{1+C} e_2 de_2 = \\
&= \int_{1-C}^{\frac{y}{x}(1+C)} de_1 \left[\frac{e_2^2}{2} \right]_{e_2=\frac{x}{y}e_1}^{e_2=1+C} = \\
&= \int_{1-C}^{\frac{y}{x}(1+C)} \left[\frac{(1+C)^2}{2} - \frac{x^2 e_1^2}{y^2 2} \right] de_1 = \left[\frac{(1+C)^2}{2} e_1 - \frac{x^2}{6y^2} e_1^3 \right]_{e_1=1-C}^{e_1=\frac{y}{x}(1+C)} = \\
&= \frac{(1+C)^3}{2} \frac{y}{x} - \frac{x^2}{6y^2} \frac{y^3}{x^3} (1+C)^3 + \\
&\quad - \frac{(1+C)^2(1-C)}{2} + \frac{x^2}{6y^2} (1-C)^3 = \frac{x^2}{6y^2} (1-c)^3 + (1+C)^3 \frac{y}{3x} - \frac{(1+C)^2(1-C)}{2}
\end{aligned}$$

Mancano ancora 3 casi limite: $u(x, 0)$ con $x > 0$, $u(0, y)$ con $y > 0$ e $u(0, 0)$ (se $y = 0$ i calcoli precedenti non hanno significato):

- CASO 5: $u(x, 0)$ con $x > 0$. In tal caso la farfalla A vince sempre ed ottiene un'utilità $e_1 - ye_2 = e_1$, che in media vale 1. Quindi, $u(x, 0) = 1$ se $x > 0$.
- CASO 6: $u(0, y)$ con $y > 0$. In tal caso la farfalla A perde sempre ed ottiene un'utilità $Re_1(1-x) = Re_1$, che in media vale R . Quindi, $u(0, y) = R$ se $y > 0$.
- CASO 7: $u(0, 0)$. In tal caso la farfalla A vince nel 50% dei casi ed ottiene e_1 (in media 1),perde nel 50%dei casi ed ottiene Re_1 (in media R). Quindi, $u(0, 0) = (R + 1)/2$.

Concludendo, la funzione di utilità $u(x, y)$ ha la seguente espressione molto complicata, che é diversa nelle 4 zone del quadrato $[0, 1]^2$ del piano cartesiano con assi x, y (vedi figura):

$$\begin{aligned}
u(x, y) &= R(1-x) \quad \text{nella zona 1} \\
u(x, y) &= R(1-x) + (1-R+Rx)a(x, y)/(4C^2) - yb(x, y)/(4C^2) \quad \text{nella zona 2} \\
u(x, y) &= 1-y + (R-Rx-1)c(x, y)/(4C^2) + yd(x, y)/(4C^2) \quad \text{nella zona 3} \\
u(x, y) &= 1-y \quad \text{nella zona 4} \\
u(x, 0) &= 1 \quad \text{se } x > 0 \\
u(0, y) &= R \quad \text{se } y > 0 \\
u(0, 0) &= (R + 1)/2
\end{aligned}$$

Avendo posto:

$$a(x, y) = (1 - C)^3 y^2 / (6x^2) + (1 - C)^3 x / (3y) - (1 - C)(1 + C)^2 / 2$$

$$b(x, y) = (1 + C)^3 x^2 / (6y^2) + (1 - C)^3 y / (3x) - (1 + C)(1 - C)^2 / 2$$

$$c(x, y) = (1 + C)^3 y^2 / (6x^2) + (1 - C)^3 x / (3y) - (1 + C)(1 - C)^2 / 2$$

$$d(x, y) = (1 - C)^3 x^2 / (6y^2) + (1 - C)^3 y / (3x) - (1 - C)(1 + C)^2 / 2$$

Con ulteriori calcoli (facili, ma lunghi) si vede che $u(x, y)$ e la sua derivata parziale $\delta u / \delta x$ rispetto a x sono continue in tutto il quadrato $[0, 1]^2$ del piano x, y , tranne che nell'origine.

Una volta trovata la forma della funzione di utilità $u(x, y)$ possiamo cercare le eventuali strategie evolutivamente stabili (ESS). (Figura 7.5)

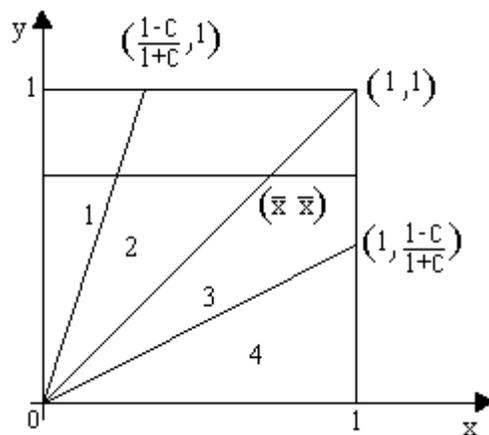


Figura 7.5

Affinché $\bar{x} \in [0, 1]$ sia una ESS occorre innanzitutto che (\bar{x}, \bar{x}) sia un equilibrio di Nash del gioco G e quindi, \bar{x} sia un punto di massimo assoluto per la funzione $g(x) = u(x, \bar{x})$.

Distinguiamo i 3 casi $\bar{x} = 0$, $\bar{x} \in (0, 1)$, $\bar{x} = 1$

- primo caso: $\bar{x} = 0$. Allora, $g(x) = u(x, 0) = 1$ se $x > 0$, mentre $g(0) = u(0, 0) = \frac{R+1}{2}$. Poiché R deve essere < 1 (altrimenti la farfalla che perde non avrebbe alcun danno), $g(x)$ non può avere massimo in 0. Quindi, 0 non è una ESS.
- secondo caso: $\bar{x} \in (0, 1)$. In questo caso condizione necessaria affinché \bar{x} sia una ESS è che $g'(\bar{x}) = \frac{\delta u}{\delta x}(\bar{x}, \bar{x}) = 0$. Calcolando $\frac{\delta u}{\delta x}(x, x)$ mediante

l'espressione di $u(x, y)$ precedente si trova che:

$$\frac{\delta u}{\delta x}(x, x) = \frac{(1-R)(c^2+3)(1-x) - Rc(3-c)}{6cx}$$

uguagliando quest'espressione a 0, si ha:

$$\begin{aligned} (1-R)(c^2+3)(1-x) &= Rc(3-c) \\ (1-R)(c^2+3) - x(1-R)(c^2+3) &= Rc(3-c) \\ (1-R)(c^2+3) &= x((1-R)(c^2+3) + Rc(3-c)) \\ 1 &= x \left(1 + \frac{Rc(3-c)}{(1-R)(c^2+3)} \right) \end{aligned}$$

E quindi, l'unico valore \bar{x} di x che può essere una ESS é:

$$\bar{x} = \frac{1}{1 + \frac{Rc(3-c)}{(1-R)(c^2+3)}} \quad (1)$$

Peró non siamo ancora sicuri che \bar{x} sia una ESS, perché la condizione $g'(\bar{x}) = 0$ é necessaria, ma non sufficiente.

- terzo caso: $\bar{x} = 1$. In questo caso siamo nell'estremo destro dell'intervallo $[0, 1]$ e quindi, una condizione necessaria perché 1 sia una ESS non é piú $g'(1) = 0$ ma $g'(1) \geq 0$. Dall'espressione precedente di $u(x, y)$ si trova che $g'(1) = \frac{\delta u}{\delta x}(1, 1) = \frac{R(c-3)}{6}$. Se $R > 0$, questo valore é negativo e quindi, 1 non può essere una ESS. Se $R = 0$, si ha $g'(1) = 0$ e quindi, 1 potrebbe essere una ESS. Questo risultato concorda col caso 2, perché la formula (1) dá $\bar{x} = 1$ per $R = 0$, $\bar{x} < 1$ per $R > 0$.

Ma in quali casi il valore di \bar{x} dato dalla (1) é effettivamente una ESS?. Affinché \bar{x} sia una ESS un'altra condizione necessaria é che sia $u(\bar{x}, \bar{x}) \geq u(0, \bar{x}) = R$.

Ma dall'espressione di $u(x, y)$ si vede che $u(x, x) = \frac{c+3+R(3-c)-x(R+1)(3-c)}{6}$ e allora imponendo che $u(\bar{x}, \bar{x}) \geq R$ si trova :

$$\begin{aligned} c + 3 + 3R - Rc - \bar{x}(R+1)(3-c) &\geq 6R \\ c + 3 - 3R - Rc &\geq \bar{x}(R+1)(3-c) \\ \bar{x} &\leq \frac{(c+3)(1-R)}{(R+1)(3-c)} \\ \frac{(1-R)(c^2+3) + Rc(3-c)}{(1-R)(c^2+3)} &\geq \frac{(R+1)(3-c)}{(1-R)(C+3)} \\ (1-R)(c+3)(c^2+3) + Rc(9-c^2) &\geq (R+1)(3-c)(c^2+3) \\ (c+3)(c^2+3) - (3-c)(c^2+3) &\geq R((3-c)(c^2+3) + (c+3)(c^2+3) - c(9-c^2)) \\ 2c(c^2+3) &\geq R(6c^2+18-9c+c^3) \end{aligned}$$

$$R \leq \frac{2c(c^2 + 3)}{c^2(c + 6) + 9(2 - c)}$$

Posto $f(c) = \frac{2c(c^2+3)}{c^2(c+6)+9(2-c)}$, si vede facilmente che $f(c)$ é una funzione crescente di c e cresce da 0 a $1/2$ quando c cresce da 0 a 1 (vedi figura 7.6).

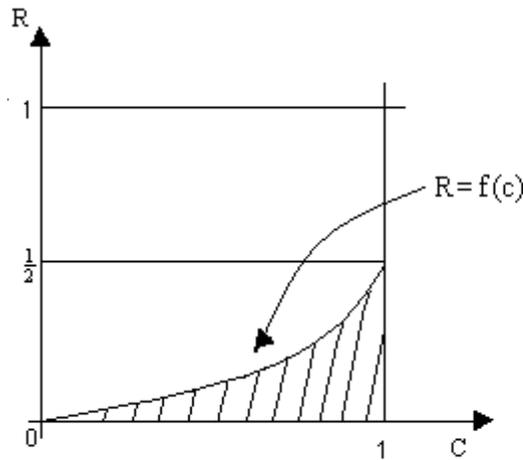


Figura 7.6

Si può poi dimostrare che *la condizione $R \leq f(c)$ é anche sufficiente perché \bar{x} sia una ESS* (vedi dimostrazione a seguire) e quindi, concludiamo che se $R \leq f(c)$ esiste un'unica ESS \bar{x} , mentre se $R > f(c)$ non esiste nessuna ESS. In quest'ultimo caso si ha $u(0, \bar{x}) > u(\bar{x}, \bar{x})$ e quindi, se in una popolazione di farfalle che adottano tutte la strategia \bar{x} compare un mutante che adotta la strategia 0 (cioè non vuole combattere), i mutanti sono favoriti e invadono la popolazione.

Notiamo che se $c = 1$, $r = 1/2$ (punto A in figura) si ha $\bar{x} = 2/3$, $\frac{u(\bar{x}, \bar{x})}{u(0,0)} = 2/3$, mentre é facile verificare che in tutti gli altri casi in cui esista una ESS (zona tratteggiata in figura si ha $\bar{x} \geq 2/3$, $\frac{u(\bar{x}, \bar{x})}{u(0,0)} \leq 2/3$ Ciò significa che, quando esiste una ESS, le farfalle lottano sempre ad oltranza (consumando almeno i $2/3$ della loro energia) e comunque, la ESS non é efficiente, in quanto se tutte le farfalle non combattessero (cioè adottassero la strategia $x = 0$, otterrebbero un'utilità maggiore.

LA CONDIZIONE $R \leq f(c)$ É ANCHE SUFFICIENTE PERCHÉ \bar{x} SIA UNA ESS. **Dimostrazione.** Basta dimostrare che \bar{x} é un punto di massimo assoluto per $g(x) = u(x, \bar{x})$ in $[0, 1]$.

Se $R \leq f(c)$ si verifica con facili calcoli che $\bar{x} \geq \frac{1-c}{1+c}$ e quindi, $u(x, \bar{x})$ può avere 3 espressioni diverse a seconda che il punto (x, \bar{x}) cada nella zona 1,2,3 della figura (7.5). Usando l'espressione di $u(x, y)$ si vede che $g(x)$ é del tipo

:

$$g(x) = \begin{cases} R(1-x) & \text{se } x \in \left[0, \frac{1-c}{1+c}\bar{x}\right] \\ A_1x^2 + B_1x + C_1 + D_1/x + E_1/x^2 & \text{se } x \in \left[\frac{1-c}{1+c}\bar{x}, \bar{x}\right] \\ A_2x^2 + B_2x + C_2 + D_2/x + E_2/x^2 & \text{se } x \in [\bar{x}, 1] \end{cases}$$

dove $A_1, B_1 \dots E_2$ sono opportune costanti. Si verifica inoltre che $g(x)$ e la sua derivata prima $g'(x)$ sono continue in $[0, 1]$ e che $g'\left(\frac{1-c}{1+c}\bar{x}\right) = -R, g'(\bar{x}) = 0, g'(1) \leq 0$. Per studiare l'andamento di $g(x)$ consideriamo la sua derivata terza $g'''(x)$:

$$g'''(x) = \begin{cases} -6D_1/x^4 - 24E_1/x^5 & \text{se } x \in \left(\frac{1-c}{1+c}\bar{x}, \bar{x}\right) \\ -6D_2/x^4 - 24E_2/x^5 & \text{se } x \in (\bar{x}, 1) \end{cases}$$

dall'espressione di $u(x, y)$ si ricavano i valori delle 4 costanti D_1, E_1, D_2, E_2 :

$$D_1 = \frac{(1-c)^3\bar{x}^2}{24c^2}(R-2) \quad E_1 = \frac{(1-c)^3\bar{x}^2}{24c^2}(1-R)$$

$$D_2 = \frac{(1+c)^3\bar{x}^2}{24c^2}(2-R) \quad E_2 = \frac{(1+c)^3\bar{x}^2}{24c^2}(R-1)$$

si ha quindi:

$$\begin{aligned} g'''(x) &= \frac{-6}{x^5}(D_1x + 4E_1) \\ &= \frac{-6}{x^5} \frac{(1-c)^3\bar{x}^2}{24c^2} ((R-2)x + 4(1-R)) \leq 0 \quad \text{se } x \in \left(\frac{1-c}{1+c}\bar{x}, \bar{x}\right) \end{aligned}$$

$$\begin{aligned} g'''(x) &= \frac{-6}{x^5}(D_2x + 4E_2) \\ &= \frac{-6}{x^5} \frac{(1+c)^3\bar{x}^2}{24c^2} ((2-R)x + 4(R-1)) \geq 0 \quad \text{se } x \in (\bar{x}, 1) \end{aligned}$$

Ne deriva che la funzione $g'(x)$ é sempre concava in $\left[\frac{1-c}{1+c}\bar{x}, \bar{x}\right]$ (avendo derivata seconda negativa) e sempre convessa in $[\bar{x}, 1]$. Poiché $g'\left(\frac{1-c}{1+c}\bar{x}\right) = -R \leq 0$, $g'(\bar{x}) = 0, g'(1) \leq 0$, se ne deduce che il grafico di $g'(x)$ puó essere di uno dei 2 tipi illustrati in figura, cioè possono presentarsi 2 casi:

- 1 CASO: $g'(x) \leq 0$ in $[0, 1]$. Allora, $g(x)$ é sempre decrescente e quindi, ha un massimo assoluto per $x = 0$.
- 2 CASO: esiste un unico punto $x_0 \in \left(\frac{1-c}{1+c}\bar{x}, \bar{x}\right)$ tale che $g'(x_0) = 0$. In tal caso $g(x)$ é decrescente in $[0, x_0]$ e in $[\bar{x}, 1]$ e crescente in $[x_0, \bar{x}]$, quindi ha 2 punti di massimo relativo: 0 e \bar{x} .

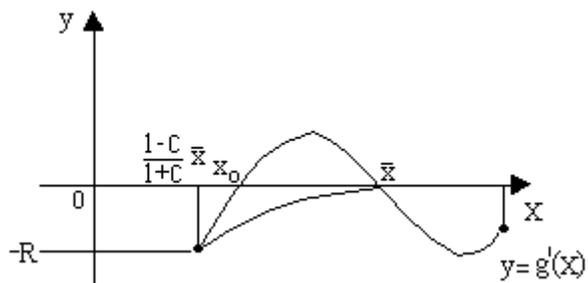


Figura 7.7

In entrambi i casi se ne deduce che \bar{x} é punto di massimo assoluto per $g(x)$ se e solo se $g(\bar{x}) \geq g(0)$, cioè $R \leq f(c)$.

Per vedere che se $R \leq f(c)$, allora \bar{x} é una ESS bisogna verificare ancora una cosa: se vale l'uguaglianza $u(0, \bar{x}) = u(\bar{x}, \bar{x})$ (cioé $g(\bar{x}) = g(0)$ cioè $R = f(c)$) dev'essere $u(0, 0) < u(\bar{x}, 0)$, cioè $\frac{R+1}{2} < 1$, cioè $R < 1$.

Poiché questo é sempre vero, la dimostrazione é conclusa. ■

8 Equazioni differenziali autonome

Definizione 8.1 Un'equazione differenziale autonoma del primo ordine in forma normale nella funzione incognita $x(t)$ é un'equazione differenziale del tipo :

$$x'(t) = f(x(t)) \quad (1) \quad (\text{o semplicemente } x' = f(x))$$

Supponiamo che la funzione $f(x)$ sia definita in un certo intervallo I e sia di classe C^1 (cioé abbia derivata continua) in I . In tal caso vale il *teorema di esistenza e unicitá in piccolo*: se fissiamo una certa condizione iniziale $x(t_0) = x_0$ (con x_0 in I), allora, il problema di Cauchy dato dall'equazione differenziale (1) con la condizione iniziale $x(t_0) = x_0$ ha una e una sola soluzione in un opportuno intorno $(t_0 - \delta, t_0 + \delta)$ di t_0 .

Se inoltre supponiam o che $f(x)$ oppure $f'(x)$ siano limitate in I , allora vale anche il *teorema di esistenza e unicitá in grande*: il problema di Cauchy ha una e una sola soluzione $x(t)$ definita in tutto R .

Se in un punto x^* si ha $f(x^*) = 0$, chiaramente la funzione costante $x(t) = x^*$ soddisfa l'equazione differenziale. Si dice che x^* é un *punto di equilibrio* e che $x(t) = x^*$ é una soluzione di equilibrio.

Le equazioni differenziali autonome del primo ordine in forma normale sono un caso particolare delle *equazioni differenziali a variabili separabili*, la cui forma generale é :

$$x'(t) = dx/dt = f(x(t))g(t) \quad (2)$$

Per un'equazione differenziale a variabili separabili la soluzione del problema di Cauchy dato dall'equazione (2) con la condizione iniziale $x(t_0) = x_0$ si calcola formalmente cosí:

$$dx/f(x) = g(t)dt \quad \int_{x_0}^x dx/f(x) = \int_{t_0}^t g(t)dt$$

Se gli integrali che si ottengono sono calcolabili elementarmente, indicando con $F(x)$, $G(t)$ rispettivamente una primitiva di $1/f(x)$ e una primitiva di $g(t)$, l'ultima uguaglianza diventa:

$$F(x) - F(x_0) = G(t) - G(t_0)$$

Se poi la funzione $F(x)$ é invertibile, indicando con F^{-1} la sua inversa, si puó trovare l'espressione esplicita della soluzione del problema di Cauchy:

$$x(t) = F^{-1}(G(t) - G(t_0) + F(x_0))$$

Applichiamo questo procedimento alla soluzione della generica equazione differenziale autonoma del primo ordine in forma normale $x'(t) = dx/dt =$

$f(x(t))$, con la condizione iniziale $x(0) = x_0$.

Supponiamo, per fissare le idee, che $f(x)$ sia definita in tutto R , si annulli in 2 punti a, b ($a < b$), sia positiva nell'intervallo (a, b) e negativa per $x < a$ e $x > b$ (fig.8.1). Allora, l'equazione differenziale $x'(t) = f(x(t))$ ha le 2 soluzioni di equilibrio $x(t) = a$, $x(t) = b$.

Queste 2 soluzioni sono rappresentate dalle 2 rette orizzontali $x = a$, $x = b$ in fig.8.2, dove è disegnato il grafico di $x(t)$ in funzione di t .

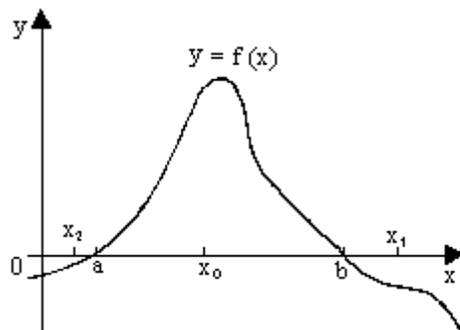


Figura 8.1

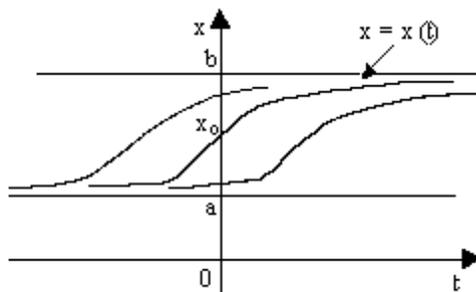


Figura 8.2

Consideriamo ora un punto qualsiasi x_0 in (a, b) e risolviamo il problema di Cauchy $dx/dt = f(x)$ con la condizione iniziale $x(0) = x_0$:

$$dx/f(x) = dt \quad \int_{x_0}^x dx/f(x) = \int_0^t dt \quad F(x) - F(x_0) = t$$

$$x = F^{-1}(t + F(x_0))$$

Dove $F(x)$ è una primitiva di $1/f(x)$. Poiché $1/f(x) > 0$ in (a, b) , $F(x)$ è crescente in (a, b) e quindi è invertibile in (a, b) ; indichiamo con F^{-1} la sua inversa. Inoltre si può dimostrare che $F(x)$ tende a $-\infty$ per $x \rightarrow a^+$ e tende a $+\infty$ per $x \rightarrow b^-$.

Infatti, indichiamo con M il massimo di $|f'(x)|$ in $[a, b]$, che esiste sicuramente per il teorema di Weierstrass sulle funzioni continue, essendo $|f'(x)|$ continua. Scelto un qualsiasi punto x nell'intervallo (a, b) , per il teorema di Lagrange sulle funzioni derivabili applicato all'intervallo (x, b) esiste un punto c in (x, b) tale che:

$$\left| \frac{(f(b) - f(x))}{(b - x)} \right| = |f'(c)| \leq M$$

ma poiché $f(b) = 0$, se ne deduce:

$$f(x)/(b-x) \leq M \quad f(x) \leq M(b-x) \quad 1/f(x) \geq 1/(M(b-x))$$

da cui, se anche x_0 sta in (a, b) , si ha (se $x_0 < x$):

$$F(x) - F(x_0) = \int_{x_0}^x dx/f(x) \geq \int_{x_0}^x dx/(M(b-x)) = -(1/M)\log|(b-x)/(b-x_0)|$$

ma per $x \rightarrow b^-$ il secondo membro di quest'uguaglianza tende a $+\infty$ e quindi, anche $F(x)$ tende a $+\infty$. Analogamente si dimostra che $F(x)$ tende a $-\infty$ per $x \rightarrow a^+$.

Ne deriva che la soluzione $x(t)$ del problema di Cauchy, essendo data da $x(t) = F^{-1}(t + C)$ (dove $C = F(x_0)$, é definita in tutto \mathbb{R} , é crescente e tende ad a per $t \rightarrow -\infty$, tende a b per $t \rightarrow +\infty$ (fig. 8.2).

Notiamo che tutte le soluzioni dell'equazione differenziale $dx/dt = f(x)$ tali che $a < x(t) < b$ sono del tipo $x(t) = F^{-1}(t + C)$ e quindi, dato il grafico di una di esse, il grafico di tutte le altre si ottiene traslando questo grafico parallelamente all'asse t , verso destra o verso sinistra di un tratto qualsiasi (fig.8.2). Quindi *le soluzioni di un'equazione differenziale autonoma del primo ordine in forma normale sono invarianti per traslazioni temporali*, il che significa che se $x(t)$ é una soluzione, allora anche $x(t + C)$ é soluzione qualunque sia la costante C . Questa é una proprietá generale, che vale per tutte le equazioni differenziali autonome, come é facile verificare.

Consideriamo ora un punto qualsiasi $x_1 > b$ e risolviamo il problema di Cauchy $dx/dt = f(x)$, con la condizione iniziale $x(0) = x_1$. Si trova $x = F^{-1}(t + F(x_0))$, dove $F(x)$ é una generica primitiva di $1/f(x)$ in $(b, +\infty)$ e F^{-1} é l'inversa di F . Poiché $1/f(x) < 0$ in $(b, +\infty)$, $F(x)$ é decrescente in $(b, +\infty)$ e tende a $+\infty$ per $x \rightarrow b^+$ si dimostra analogamente a prima) mentre per $x \rightarrow +\infty$, $F(x)$ puó tendere a $-\infty$ oppure ad un valore finito $t_1 < 0$.

Quindi, la soluzione $x(t)$ del problema di Cauchy, essendo data da $x(t) = F^{-1}(t + C)$, é decrescente e ha come codominio l'intervallo $(b, +\infty)$ e inoltre tende a b per $t \rightarrow +\infty$.

Peró il suo dominio puó essere tutto \mathbb{R} oppure un intervallo del tipo $(t_1, +\infty)$ e quindi, il suo grafico puó essere di uno dei 2 tipi disegnati in fig. 8.3. Per esempio se $f(x) = (x-1)(2-x)x$ e se supponiamo $x(t) > 2$, la soluzione é del tipo di fig. 8.3A, mentre se $f(x) = (x-1)(2-x)/$, la soluzione é del tipo di fig. 8.3B.

Infine, consideriamo un punto qualsiasi $x_2 < a$ e risolviamo il problema di Cauchy $dx/dt = f(x)$ con la condizione iniziale $x(0) = x_2$. Con considerazioni analoghe a quelle fatte prima, si vede che la soluzione $x(t)$ é decrescente, ha come codominio l'intervallo $(-\infty, a)$ e tende ad a per $t \rightarrow -\infty$. Il

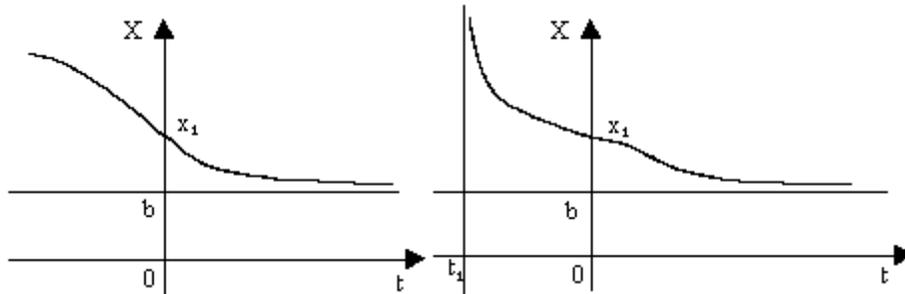


Figura 8.3A

Figura 8.3B

suo dominio può essere tutto \mathbb{R} oppure un intervallo del tipo $(-\infty, t_2)$ con $t_2 > 0$ e il suo grafico può essere di uno dei 2 tipi disegnati in fig.8.4

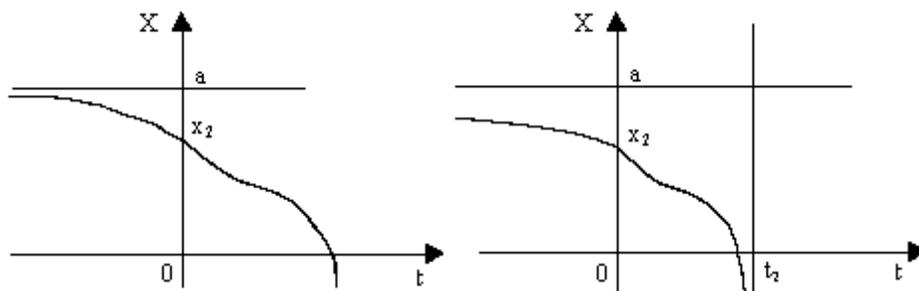


Figura 8.4A

Figura 8.4B

Ogniqualevolta si ha un intervallo in cui $f(x)$ ha segno costante, ci troviamo di fronte ad una situazione analoga a quella che abbiamo descritto. Il comportamento delle soluzioni (crescenza, decrescenza, limiti agli estremi del dominio) si può studiare anche senza risolvere esplicitamente l'equazione (*studio qualitativo* delle soluzioni).

Nel caso in cui $f(x)$ ha segno costante su tutto \mathbb{R} , (ad esempio $f(x) > 0$ per ogni x in \mathbb{R}) si può solo dire che ogni soluzione $x(t)$ è crescente e ha come codominio tutto \mathbb{R} , ma il suo dominio può essere di uno di questi 4 tipi: \mathbb{R} , (t_1, t_2) , $(-\infty, t_2)$, $(t_1, +\infty)$ ed il suo grafico può essere di uno dei 4 tipi di fig.8.5. Un esempio per ciascuno di questi 4 comportamenti si può avere prendendo $f(x) = 1$, $f(x) = 1 + x^2$, $f(x) = 1 + x^2 + x|x|$, $f(x) = 1 + x^2 - x|x|$.

Nel caso, già considerato prima che sia $f(a) = f(b) = 0$, $f(x) > 0$ in (a, b) , $f(x) < 0$ in $(-\infty, a)$, e in $(b, +\infty)$, se fissiamo una condizione iniziale $x(0) = x_0$ possono verificarsi 4 casi:

1. Se $x_0 = a$ oppure $x_0 = b$, la soluzione $x(t)$ é costante (equilibrio).
2. Se $a < x_0 < b$, la soluzione $x(t)$ é crescente e tende a b per $t \rightarrow +\infty$.
3. Se $x_0 > b$, la soluzione $x(t)$ é decrescente e tende a b per $t \rightarrow +\infty$.
4. Se $x_0 < a$, la soluzione $x(t)$ é decrescente e tende a $-\infty$ per $t \rightarrow +\infty$, oppure per t tendente ad un certo valore $t_2 > 0$.

In questo caso si dice che a é un punto di *equilibrio instabile* e che b é un punto di *equilibrio asintoticamente instabile*. Piú in generale, possiamo dare le seguenti definizioni:

Definizione 8.2 Si dice che un punto di equilibrio $x = a$ per l'equazione differenziale $x' = f(x)$ é stabile se, fissati comunque t_0 in \mathbb{R} e $\varepsilon > 0$, esiste $\delta > 0$ tale che, per ogni soluzione $x(t)$, se $|x(t_0) - a| < \delta$, allora $|x(t) - a| < \varepsilon$ per ogni $t > t_0$.

Si dice che un punto di equilibrio é instabile se non é stabile.

Definizione 8.3 Si dice che un punto di equilibrio $x = a$ per l'equazione differenziale $x' = f(x)$ é asintoticamente stabile se, fissato comunque t_0 in

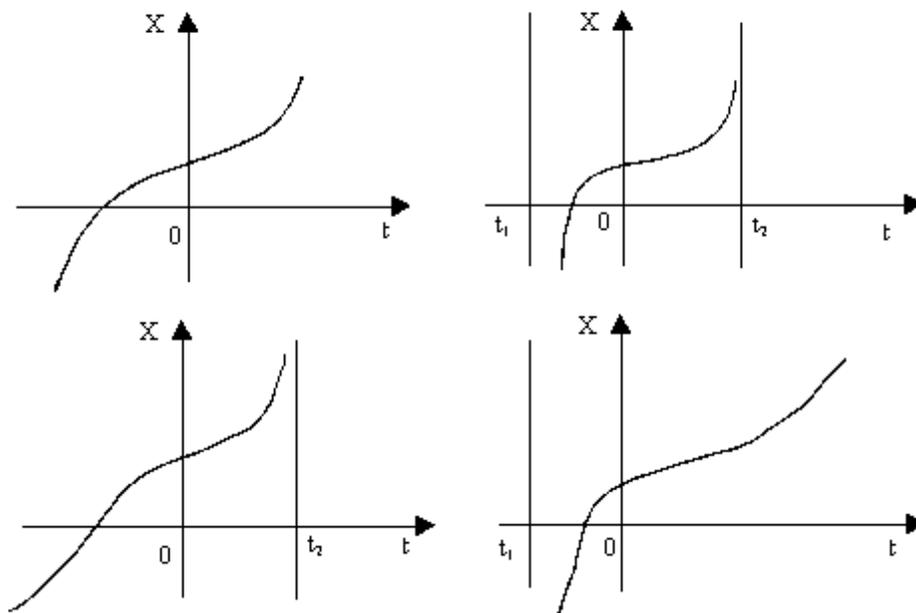


Figura 8.5

Se un punto di equilibrio é asintoticamente stabile, allora é stabile (ma non viceversa). Si dimostra facilmente che :

1. Se $f(a) = 0$ e inoltre $f(x) > 0$, in un intorno sinistro di a , $f(x) < 0$, in un intorno destro di a , allora, a é un punto di equilibrio asintoticamente stabile.
2. Se $f(a) = 0$ e inoltre $f(x)$ é diversa da 0 in un intorno di a , ma il segno di $f(x)$ non é come nel caso 1 (cioé $f(x)$ é positiva in tutto l'intorno, o negativa in tutto l'intorno, o positiva a destra di a e negativa a sinistra), allora a é un punto di equilibrio instabile.
3. Se $f'(a) > 0$, allora a é un punto di equilibrio instabile. Esempio: $f(x) = x^2 - x$, $a = 1$. Le soluzioni sono del tipo $x(t) = 1/(1 - Ce^t)$ e hanno l'andamento di figura 8.6
4. Se $f'(a) < 0$, allora a é un punto di equilibrio asintoticamente stabile. Esempio: $f(x) = x^2 - x$, $a = 0$. Le soluzioni sono del tipo $x(t) = 1/(1 - Ce^t)$ e hanno l'andamento di figura 8.6
5. Se $f'(a) = 0$, non si puó dire niente. Il punto di equilibrio a puó essere instabile, asintoticamente stabile, oppure stabile ma non asintoticamente stabile. Tre esempi di queste 3 situazioni sono i seguenti: $f(x) = x^3$, $a = 0$ (le soluzioni sono del tipo $x(t) = \pm 1/(C - 2t)^{1/2}$ e hanno l'andamento di fig.7)., $f(x) = -x^3$, $a = 0$ (le soluzioni sono del tipo $x(t) = \pm 1/(C - 2t)^{1/2}$ e hanno l'andamento di fig.8), $f(x) = x^3 - |x^3|$, $a = 0$ (le soluzioni sono del tipo $x(t) = \text{costante}$ se $x(t) > 0$, $x(t) = -1/(C - 4t)^{1/2}$ se $x(t) < 0$ e hanno l'andamento di fig.9).

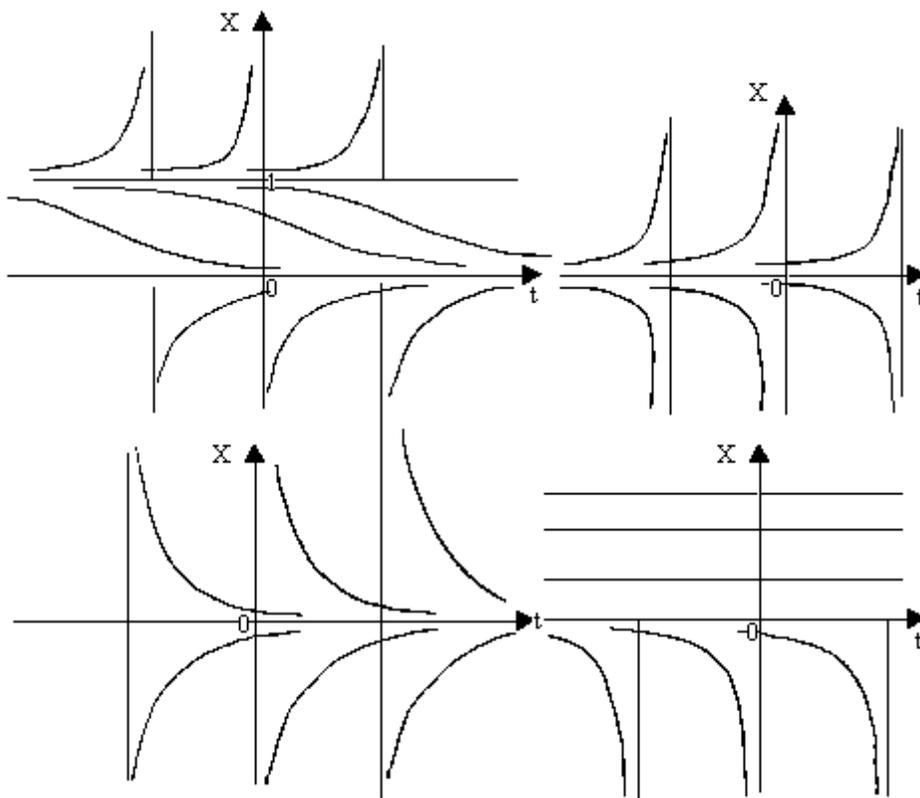


Figura 8.6

Figura 8.7

Figura 8.8

Figura 8.9

8.1 Crescita di popolazioni

Consideriamo una popolazione di esseri viventi ed indichiamo con $n(t)$ il numero di individui della popolazione al tempo t .

La funzione $n(t)$ é una funzione definita sui numeri reali ($t \in \mathbb{R}$) e a valori nei numeri naturali ($n(t) \in \mathbb{N}$), perciò é una funzione *costante a tratti*, cioè una funzione che varia a scatti negli istanti di tempo in cui si verifica una nascita o una morte e quindi, non può essere continua e tanto meno derivabile. Se però si considerano sempre popolazioni con un numero di individui abbastanza grande e si approssima $n(t)$ con una funzione derivabile $x(t)$ che differisca da $n(t)$ per meno di 1 unità ($|n(t) - x(t)| \leq 1$), il che é sempre possibile, l'errore relativo che si commette é molto piccolo.

Ciò premesso, consideriamo un generico intervallo di tempo $I = (t, t + \Delta t)$ e siano n_1, n_2 il numero di nascite ed il numero di morti che si verificano nell'intervallo I . Allora, la popolazione all'istante $t + \Delta t$ é data da $x(t + \Delta t) = x(t) + n_1 - n_2$ (supponiamo che non ci siano né immigrazione né emigrazione).

Definiamo allora:

- tasso di natalità medio in I , $r_1(t, \Delta t) = \frac{n_1}{x(t)\Delta t}$
- tasso di mortalità medio in I , $r_2(t, \Delta t) = \frac{n_2}{x(t)\Delta t}$
- tasso di crescita medio in I , $r(t, \Delta t) = r_1(t, \Delta t) - r_2(t, \Delta t) = \frac{x(t+\Delta t) - x(t)}{x(t)\Delta t}$
- tasso di crescita istantaneo all'istante t :

$$r(t) = \lim_{\Delta t \rightarrow 0} r(t, \Delta t) = \lim_{\Delta t \rightarrow 0} \frac{x(t + \Delta t) - x(t)}{x(t)\Delta t} = \frac{x'(t)}{x(t)}$$

Il tasso di crescita di una popolazione si può esprimere ad esempio in aumento del numero di individui per ogni anno per milione di individui. Ad esempio si può dire che $r(t) = 0.003$, il che significa che la popolazione aumenta di $3/1000$ all'anno, cioè di 3000 individui all'anno per milioni di individui.

Sul tasso di crescita si possono fare varie ipotesi, ad esempio le seguenti 3, che danno luogo a 3 modelli di crescita di popolazioni molto semplici (*crescita esponenziale*, *crescita logistica*, *crescita esplosiva*):

- $r(t) = \text{costante} = r \in \mathbb{R} \rightarrow$ crescita esponenziale
- $r(t) = a - bx(t)$ ($a, b \in \mathbb{R}, a, b > 0$) \rightarrow crescita logistica
- $r(t) = ax(t) - b$ ($a, b \in \mathbb{R}, a, b > 0$) \rightarrow crescita esplosiva

Esempio 8.1 CRESCITA ESPONENZIALE (Malthus, 1798)

Si suppone che il tasso di crescita $r(t)$ sia costante ed assuma un valore $r = r_1 - r_2$ dove r_1 ed r_2 sono il tasso di natalità ed il tasso di mortalità e possono essere supposti anch'essi costanti nel tempo (naturalmente, $r_1 \geq 0$, $r_2 \geq 0$). Se $r_1 > r_2$ si ha $r > 0$ e la popolazione cresce nel tempo, se $r_1 < r_2$ si ha $r < 0$ e la popolazione decresce nel tempo, se $r_1 = r_2$ si ha $r = 0$ e la popolazione rimane fissa al valore $x(t) = x_0$ che aveva per $t = 0$.

L'equazione differenziale che esprime il modello é:

$$\frac{dx}{dt} = rx(t) \quad \text{con la condizione iniziale } x(0) = x_0$$

Risolvendo, si trova:

$$\int_{x_0}^x \frac{dx}{x} = \int_0^t r dt \quad \log \frac{x}{x_0} = rt$$

$$\frac{x}{x_0} = e^{rt} \quad x(t) = x_0 e^{rt}$$

Se $r > 0$, (crescita esponenziale) la popolazione $x(t)$ tende a $+\infty$ per $t \rightarrow +\infty$. Se $r < 0$ (decadimento esponenziale), $x(t)$ decresce e tende a

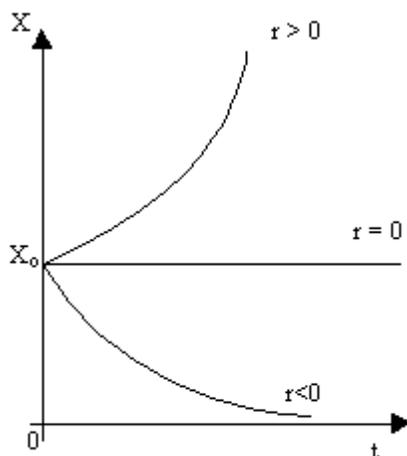


Figura 8.6

0 per $t \rightarrow +\infty$. Se $r = 0$, $x(t) = x_0$ per ogni t . (fig. 8.10).

Questo modello naturalmente, non può essere valido in ogni situazione perché in natura la popolazione è contenuta in un certo ambiente che le fornisce il nutrimento e non può crescere oltre ogni limite. La crescita esponenziale è valida con buona approssimazione per popolazioni immerse in un ambiente molto ricco di risorse, finché la popolazione è abbastanza piccola dall'essere ben lontane dall'esaurire le risorse dell'ambiente (esempio: batteri in coltura). Il modello di decadimento esponenziale è valido quasi esattamente per il numero di atomi di un campione di sostanza radioattiva, per il quale il "tasso di natalità" è 0 ed il "tasso di mortalità" (cioè il tasso di decadimento) è costante nel tempo. Anche qui però dopo un certo tempo il numero di atomi radioattivi del campione diventa molto piccolo (ad esempio $n < 100$) ed il modello non vale più perché non è possibile prevedere il momento esatto in cui un singolo atomo decadrà (in tal caso si rivela più adeguato modellizzare il processo di decadimento mediante un cosiddetto "processo aleatorio" o "processo stocastico", cioè una variabile aleatoria che varia nel tempo, secondo una certa legge di probabilità). ■

Esempio 8.2 CRESCITA LOGISTICA (Verhulst, 1837)

Si suppone che il tasso di crescita $r(t)$ di una popolazione $x(t)$ sia dato da :

$$r(t) = a - bx(t) \quad (a, b \text{ costanti} > 0)$$

Il significato di questo modello è che la popolazione è immersa in un certo ambiente che può nutrire un numero massimo M di individui. Se $x(t)$ è

molto piccolo (molto minore di M), il tasso di crescita é circa uguale ad a ; quindi, a ha il significato di tasso di crescita che la popolazione avrebbe se avesse a disposizione una quantità illimitata di risorse. Se invece le risorse sono limitate, si fa l'ipotesi che il tasso di crescita $r(t)$ sia proporzionale alla percentuale di risorse non ancora utilizzate, cioè a $1 - \frac{x(t)}{M}$:

$$r(t) = a \left(1 - \frac{x(t)}{M} \right) = a - \frac{a}{M}x(t) = a - bx(t)$$

dove si é posto $\frac{a}{M} = b$, ovvero il numero massimo M di individui é uguale ad $\frac{a}{b}$.

L'equazione differenziale che esprime il modello (con la condizione iniziale $x(0) = x_0$) é:

$$\frac{dx}{dt} = r(t)x(t) = x(t) (a - bx(t)) \quad x(0) = x_0$$

e risolvendo:

$$\int_{x_0}^x \frac{dx}{x(a - bx)} = \int_0^t dt = t \quad \left[\frac{1}{a} \log \left| \frac{x}{a - bx} \right| \right]_{x_0}^x = t$$

$$\frac{1}{a} \log \left| \frac{x(a - bx_0)}{x_0(a - bx)} \right| = t$$

a questo punto osserviamo che l'equazione differenziale é autonoma del primo ordine in forma normale e ha 2 soluzioni di equilibrio: $x(t) = 0$ e $x(t) = M = \frac{a}{b}$. Per quanto detto prima sulle equazioni differenziali autonome, se $x_0 < \frac{a}{b}$, allora $x(t) < \frac{a}{b} \forall t \in \mathbf{R}$ e se $x_0 > \frac{a}{b}$, allora $x(t) > \frac{a}{b} \forall t \in \mathbf{R}$, perciò possiamo togliere il modulo nell'ultima uguaglianza e abbiamo:

$$\frac{1}{a} \log \frac{x(a - bx_0)}{x_0(a - bx)} = t \quad \frac{x}{a - bx} = e^{at} \frac{x_0}{a - bx_0}$$

$$x(t) = \frac{ax_0e^{at}}{a - bx_0 + bx_0e^{at}}$$

Si possono verificare 3 casi:

1. Se $x_0 = 0$ oppure $x_0 = M = \frac{a}{b}$, allora $x(t)$ é costante e vale 0 oppure M (soluzioni di equilibrio).
2. Se $0 < x_0 < M$ allora, $x(t)$ é crescente e tende al valore di saturazione $M = \frac{a}{b}$ per $t \rightarrow +\infty$ (fig. 11).
3. se $x_0 > M$, allora $x(t)$ é decrescente e tende al valore di saturazione $M = \frac{a}{b}$ per $t \rightarrow +\infty$ (fig. 8.11).

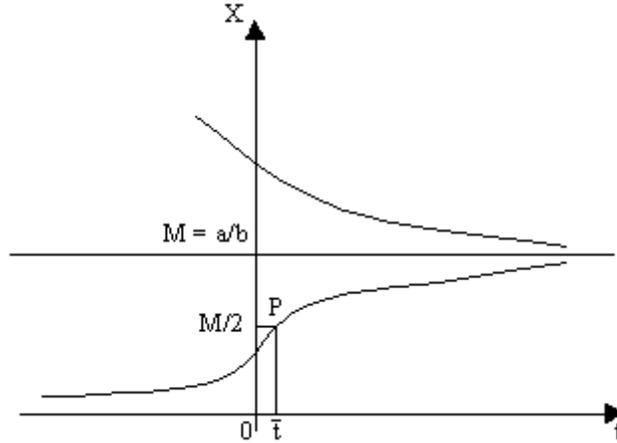


Figura 8.11

Osserviamo che nel caso 2., posto $\bar{t} = \frac{1}{a} \log \frac{a-bx_0}{bx_0}$, la soluzione $x(t)$ si può riscrivere così:

$$x(t) = \frac{a}{b} \frac{e^{a(t-\bar{t})}}{1 + e^{a(t-\bar{t})}}$$

da cui si deduce che il grafico di $x(t)$ (*curva logistica*) si ottiene dalla curva $x = \frac{e^t}{1+e^t}$ mediante opportuni cambiamenti di scala sugli assi coordinati. Il grafico di $x(t)$ ha un flesso nel punto $P = (\bar{t}, \frac{M}{2})$ ed é simmetrico rispetto a P (nel punto P la popolazione raggiunge un valore pari a metà del valore di saturazione M). ■

Esempio 8.3 CRESCITA ESPLOSIVA

Si suppone che il tasso di crescita $r(t)$ della popolazione $x(t)$ sia dato da:

$$r(t) = ax(t) - b \quad (a, b \text{ costanti } > 0)$$

Questo modello si adatta bene ad una popolazione con pochi individui dispersi su un vasto territorio. Il tasso di crescita $r(t)$ é la somma di un tasso di nascita $ax(t)$ che si suppone proporzionale alla popolazione (questo potrebbe verificarsi perché la probabilità di un incontro tra un maschio e una femmina é proporzionale a $x(t)$) e di un termine negativo costante $-b$ (tasso di morte).

L'equazione differenziale che esprime il modello (con la condizione iniziale $x(0) = x_0$) é:

$$\frac{dx}{dt} = r(t)x(t) = x(t)(ax(t) - b) \quad x(0) = x_0$$

Si tratta di un'equazione differenziale autonoma del primo ordine in forma normale che ha 2 soluzioni di equilibrio: $x(t) = 0$ e $x(t) = \frac{b}{a}$. Se $x_0 < \frac{b}{a}$, la soluzione $x(t)$ é decrescente e tende a 0 per $t \rightarrow +\infty$. Se $x_0 > \frac{b}{a}$, si ha la "crescita esplosiva": la soluzione $x(t)$ é crescente e tende a $+\infty$ in un tempo finito $\bar{t} = \frac{-1}{b} \log \frac{ax_0 - b}{ax_0}$ (esplode). Il valore di \bar{t} é positivo e tanto piú piccolo quanto piú grande é x_0 (fig. 8.12)

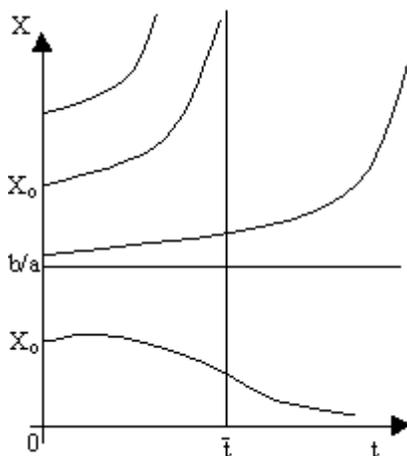


Figura 8.12

L'equazione differenziale si puó risolvere separando le variabili:

$$\begin{aligned} \frac{dx}{dt} &= x(a - x) \\ \int_{x_0}^x \frac{dx}{x(a - x)} &= \int_0^t dt = t \\ \left[\frac{1}{b} \log \left| \frac{ax - b}{ax} \right| \right]_{x_0}^x &= t \\ \frac{1}{b} \log \frac{x_0(ax - b)}{x(ax_0 - b)} &= t \quad a - \frac{b}{x} = \left(a - \frac{b}{x_0}\right) e^{bt} \\ x(t) &= \frac{bx_0}{ax_0 - (ax_0 - b)e^{bt}} \end{aligned}$$

Si possono verificare 3 casi:

1. Se $x_0 = 0$ oppure $x_0 = \frac{b}{a}$, allora $x(t) = x_0 = \text{costante}$.
2. Se $x_0 < \frac{b}{a}$, la popolazione si estingue. Il valore $\frac{b}{a}$ é un valore critico, al di sotto del quale si ha l'estinzione.

3. Se $x_0 > \frac{b}{a}$ ($ax_0 - b > 0$), il denominatore si annulla per $t = \bar{t} = -\frac{1}{b} \log \frac{ax_0 - b}{ax_0}$ e $x(t) \rightarrow +\infty$ per $t \rightarrow \bar{t}$ da sinistra (crescita esplosiva).
É chiaro però che nella realtà quando $x(t)$ cresce troppo, si esauriscono le risorse dell'ambiente ed il modello non vale più.

■

9 Datazione radioattiva

Supponiamo di avere un campione di una certa sostanza che al tempo $t = 0$ contiene un certo numero x_0 di atomi di un certo elemento radioattivo x , i quali decadono con un tasso costante r . Ciò significa che, indicando con $x(t)$ il numero di atomi di x al variare del tempo, esso diminuisce secondo l'equazione differenziale:

$$x'(t) = \frac{dx}{dt} = -rx(t)$$

Risolvendo l'equazione differenziale, con la condizione iniziale $x(0) = x_0$, si trova la legge del decadimento esponenziale:

$$x(t) = x_0 e^{-rt} \quad (1)$$

Di solito, anziché mediante il tasso di decadimento r , questa legge si esprime attraverso il cosiddetto *periodo di dimezzamento* T . Se nell'equazione precedente poniamo $x(t) = \frac{x_0}{2}$ otteniamo:

$$x_0 e^{-rt} = \frac{x_0}{2} \quad e^{-rt} = \frac{1}{2} \quad e^{rt} = 2 \quad t = \frac{\log 2}{r} \simeq \frac{0,693}{r}$$

Indichiamo con T questo valore, cioè poniamo $T = \frac{\log 2}{r}$. T rappresenta il tempo necessario per passare da x_0 a $\frac{x_0}{2}$ ed è sempre lo stesso, indipendentemente da x_0 . Poiché $r = \frac{\log 2}{T}$, la legge (1) si può esprimere anche mediante T e diventa :

$$x(t) = x_0 2^{-\frac{t}{T}} \quad (2)$$

Ciò significa che per $t = 0, T, 2T, 3T, \dots$, $x(t)$ vale rispettivamente $x_0, \frac{x_0}{2}, \frac{x_0}{4}, \frac{x_0}{8}, \dots$ eccetera. Se la sostanza X è molto radioattiva (cioè decade rapidamente) ha r grande e T piccolo, se viceversa è poco radiattiva (cioè decade lentamente) ha r piccolo e t grande. Il periodo di dimezzamento T può variare entro limiti molto ampi, come mostra la seguente tabella:

Consideriamo ora il *problema della datazione radioattiva*, cioè il problema di determinare l'“età” di una roccia (o di un qualsiasi campione di materiale) contenente uno o più elementi radioattivi e i loro prodotti di decadimento, misurando le quantità di tutte queste sostanze (espresse come numero di atomi) contenute nel campione. Si tratta di un problema molto difficile, per 2 motivi.

Per prima cosa bisogna definire bene cosa s'intende per “età” del campione: di solito s'intende il tempo trascorso da un istante iniziale t_0 , opportunamente fissato nel passato ad oggi. Per esempio, per “età” di una roccia di

elemento	periodo di dimezzamento
xeno 133	5 giorni
piombo 210	22 anni
radio 226	1600 anni
carbonio 14	5568 anni
plutonio	23103 anni
uranio 238	4,5 miliardi di anni

Tabella 9.1

solito o s'intende il tempo trascorso dalla sua ultima solidificazione.

Secondo: a priori non si ha nessuna informazione sulle quantità iniziali presenti nel campione, all'istante t_0 , dei vari elementi radioattivi e dei loro prodotti di decadimento. É chiaro che, senza tali informazioni, il problema della datazione radioattiva é quasi sempre irresolubile. Nella maggioranza dei casi si riesce solo a dare una maggiorazione dell'età, cioè a stabilire che essa dev'essere minore o uguale di un certo valore \bar{t} .

Solo in alcuni casi particolari, in cui si hanno informazioni sulle quantità iniziali, si possono fare delle datazioni abbastanza precise.

Vediamo alcuni esempi, distinguendo 3 casi:

- 1 CASO: *Decadimento di un isotopo radiattivo X in un isotopo stabile Y*

Questo é il caso piú facile. Consideriamo un campione contenente in tutto N atomi, fra cui un numero $x(t)$ variabile col tempo, di atomi dell'isotopo radioattivo X ed un numero $y(t)$, anch'esso variabile, di atomi dell'isotopo stabile Y . Supponiamo che un atomo di X decade in un atomo di Y con tasso r e periodo di dimezzamento T :

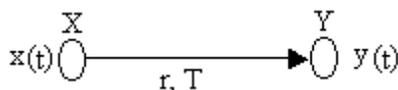


Figura 9.1

Supponiamo inoltre che al tempo iniziale $t = 0$ si abbiano certi valori iniziali $x(0) = x_0$, $y(0) = y_0$ e che in un certo istante successivo t

si osservino dei valori $x(t) = x_1$, $y(t) = y_1$. Vogliamo vedere come variano nel tempo $x(t)$, $y(t)$ ed inoltre vogliamo risolvere il problema della datazione radioattiva : conoscendo x_1 , y_1 , ricavare "età" t del campione, senza avere a priori nessuna informazione sui valori iniziali x_0 , y_0 .

É chiaro che le leggi con cui variano $x(t)$, $y(t)$ nel tempo sono le seguenti :

$$\begin{aligned}x(t) &= x_0 e^{-rt} \\y(t) &= y_0 + x_0 - x_0 e^{-rt} = y_0 + x_0 - x(t)\end{aligned}$$

Si ha quindi, che $x(t) + y(t)$ é costante e uguale a $x_0 + y_0$ (come é ovvio poiché un atomo di X decade in un atomo di Y). Scrivendo x_1 , y_1 invece di $x(t)$, $y(t)$ e supponendo noti x_1 , y_1 , e sconosciuti x_0 , y_0 , possiamo stimare l'età t del campione :

$$e^{-rt} = \frac{x_1}{x_0} \quad e^{rt} = \frac{x_0}{x_1} \quad t = \frac{1}{r} \log \frac{x_0}{x_1} \quad (3)$$

Peró in questa formula non conosciamo x_0 e quindi, non possiamo ricavare t . Poiché $y_0 + x_0 = y_1 + x_1$, sappiamo solo che x_0 deve essere compreso tra x_1 e $y_1 + x_1$, da cui :

$$0 \leq t \leq \frac{1}{r} \log \left(1 + \frac{y_1}{x_1} \right)$$

Il valore $\bar{t} = \frac{1}{r} \log \left(1 + \frac{y_1}{x_1} \right)$ é noto e ci dá una maggiorazione dell'età, in quanto deve essere $t \leq \bar{t}$. Se noi sapessimo che all'inizio nel campione non erano presenti atomi di Y (cioé tutti gli atomi di Y derivano dal decadimento di X) e quindi, $y_0 = 0$ e $x_0 = x_1 + y_1$, potremmo dire che:

$$t = \bar{t} = \frac{1}{r} \log \left(1 + \frac{y_1}{x_1} \right) \quad (4)$$

Esempio 9.1 *Il metodo del potassio - Argo*

Supponiamo ad esempio che sia $x =$ potassio 40 e $Y =$ argo 40. X decade in Y con periodo di dimezzamento $T=1,3$ miliardi di anni, Y é stabile. Poiché l'argo é un gas, é quasi impossibile che atomi di argo fossero presenti in una roccia al momento della sua formazione e quindi, possiamo essere abbastanza sicuri dell'età data dalla formula (4). Se per esempio in una roccia troviamo che il rapporto $\frac{\text{argo40}}{\text{potassio40}}$ = $\frac{y_1}{x_1}$ vale $\frac{1}{10}$, poiché $r = \frac{\log 2}{T}$, la formula (4) ci dá:

$$t = \frac{T}{\log 2} \log \left(1 + \frac{1}{10} \right) = 1300 \frac{\log 1,1}{\log 2} \text{milioni di anni} = 178 \text{ milioni di anni}$$

L'unica incertezza in questo caso é che l'argo, essendo un gas, potrebbe essere sfuggito dalla roccia per innalzamento della temperatura. Quindi, la quantità di argo prodotta dal decadimento radioattivo potrebbe essere maggiore di quella osservata, cioè potrebbe essere $\frac{y_1}{x_1} > \frac{1}{10}$. In tal caso l'età sarebbe maggiore di 178 milioni di anni (contrariamente a quel che succede di solito, si troverebbe un limite inferiore per l'età).

■

Esempio 9.2 Il metodo del Carbonio 14

Il carbonio 14 (avente 6 protoni e 8 neutroni nel nucleo) é un isotopo radioattivo del carbonio, usato per datare reperti archeologici animali o vegetali. Il carbonio 14 si forma nell'atmosfera dal normale azoto 14 (7 protoni e 7 neutroni) per effetto dei raggi cosmici che trasformano 1 protone in neutrone nel nucleo di azoto 14. Ma il carbonio 14 é radioattivo e decade con radioattività β^- (nel nucleo 1 neutrone si trasforma in protone, emettendo un elettrone) ridando azoto 14. Perciò nell'atmosfera il rapporto tra carbonio 14 ed il normale carbonio 12 (6 protoni e 6 neutroni) é costante, perché si é raggiunto un equilibrio tra la formazione di carbonio 14 ed il suo decadimento.

Le piante assorbono CO_2 dall'aria con la fotosintesi clorofilliana. Alla morte della pianta (o dell'animale che si nutre di piante o di animali erbivori) cessa ogni apporto di carbonio e poiché il carbonio 12 é stabile, il rapporto tra carbonio 14 e carbonio 12 nella pianta (o nell'animale) si limita a decadere esponenzialmente, con periodo di dimezzamento 5670 anni. Per datare il reperto basta allora, bruciarne un frammento e misurare il rapporto tra carbonio 14 e carbonio 12 nella CO_2 ottenuta.

Nel caso del carbonio 14 si ha $X =$ carbonio 14, $Y =$ azoto 14, $T = 5670$ anni. In questo caso, l'età del reperto si determina direttamente con la formula (3) $t = \frac{1}{r} \log \frac{x_0}{x_1}$, in quanto sappiamo che x_0 (numero iniziale di atomi di carbonio 14 nel campione) é dato da $x_0 = kz$, dove z é il numero di atomi di carbonio 12 nel campione e k é il rapporto di equilibrio tra carbonio 14 e carbonio 12. Si ha quindi:

$$t = \frac{1}{r} \log \frac{kz}{x_1}$$

Per esempio, sapendo che il rapporto attuale $\frac{x_1}{z}$ tra carbonio 14 e carbonio 12 é $\frac{k}{20}$ ($\frac{1}{20}$ del valore di equilibrio), si ricava $x_1 = \frac{kz}{20}$ e poiché $r = \frac{\log 2}{T}$ si trova:

$$t = \frac{T}{\log 2} \log 20 = 5670 \frac{\log 20}{\log 2} \text{anni} = 24512 \text{anni}$$

■

- 2 CASO: *Decadimento di un isotopo radiattivo X in un altro isotopo radiattivo Y che decade piú rapidamente di X.*

É questo il caso piú frequente. Consideriamo un campione contenente in tutto N atomi, fra cui $x(t)$ atomi di X e $y(t)$ atomi di Y .

Supponiamo che un atomo di X decada in un atomo di Y con tasso r_1 e periodo di dimezzamento T_1 e che un atomo di Y decada in un atomo di un altro isotopo Z (che puó essere radioattivo o no) con tasso $r_2 > r_1$ e periodo di dimezzamento $T_2 < T_1$:

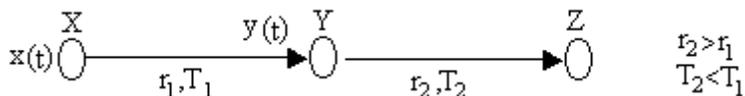


Figura 9.2

Supponiamo che all'istante iniziale $t = 0$ si abbiano certi valori iniziali $x(0) = x_0$, $y(0) = y_0$ e che in un certo istante successivo t si osservino dei valori $x(t) = x_1$, $y(t) = y_1$. Vogliamo vedere come variano nel tempo $x(t), y(t)$ ed il loro rapporto $\frac{y(t)}{x(t)}$ ed inoltre vogliamo valutare l'età t del campione, conoscendo x_1, y_1 senza sapere a priori nulla di x_0, y_0 .

Le leggi con cui variano $x(t), y(t)$ si trovano risolvendo un sistema di 2 equazioni differenziali:

$$\begin{cases} x'(t) = -r_1x(t) & x(0) = x_0 \\ y'(t) = r_1x(t) - r_2y(t) & y(0) = y_0 \end{cases}$$

Infatti, $x(t)$ decresce con velocità $-r_1x(t)$ proporzionale a $x(t)$, mentre $y(t)$ cresce con velocità $y'(t)$ data dalla somma di un termine positivo $r_1x(t)$ (dovuto ad atomi di X che decadono in Y) e di un termine negativo $-r_2y(t)$ (dovuto ad atomi di Y che decadono in Z). Risolvendo la prima equazione si trova $x(t) = x_0e^{-r_1t}$ e sostituendo nella seconda equazione si ha:

$$y'(t) + r_2y(t) = r_1x_0e^{-r_1t} \quad y(0) = y_0$$

Quest'equazione si puó risolvere moltiplicando ambo i membri per e^{r_2t} :

$$y'e^{r_2t} + r_2ye^{r_2t} = r_1x_0e^{(r_2-r_1)t}$$

$$D(ye^{r_2t}) = r_1x_0e^{(r_2-r_1)t}$$

$$ye^{r_2t} = \frac{r_1x_0}{r_2-r_1}e^{(r_2-r_1)t} + C$$

$$y(t) = Ce^{-r_2t} + \frac{r_1x_0}{r_2-r_1}e^{-r_1t}$$

Imponendo la condizione iniziale $y(0) = y_0$ si trova $C = y_0 - \frac{r_1x_0}{r_2-r_1}$ e quindi, posto $a = \frac{r_1}{r_2-r_1}$, si hanno le leggi seguenti:

$$x(t) = x_0e^{-r_1t}$$

$$y(t) = (y_0 - ax_0)e^{-r_2t} + ax_0e^{-r_1t} \quad \left(a = \frac{r_1}{r_2-r_1} \right) \quad (5)$$

$$\frac{y(t)}{x(t)} = \left(\frac{y_0}{x_0} - a \right) e^{(r_1-r_2)t} + a$$

Nella formula che dá $\frac{y(t)}{x(t)}$ l'esponente é negativo e quindi, l'esponenziale tende a 0 per $t \rightarrow +\infty$. $x(t)$ e $y(t)$ tendono entrambi a 0 (come é ovvio), ma il loro rapporto $\frac{y(t)}{x(t)}$ tende al valore di equilibrio a , qualunque sia il valore iniziale $\frac{y_0}{x_0}$. La differenza $\frac{y(t)}{x(t)} - a$ decresce esponenzialmente con tasso $r_2 - r_1$, che corrisponde ad un periodo di dimezzamento $\frac{T_1T_2}{T_1-T_2} = aT_1$.

I grafici di $x(t)$ e $\frac{y(t)}{x(t)}$ sono del tipo:

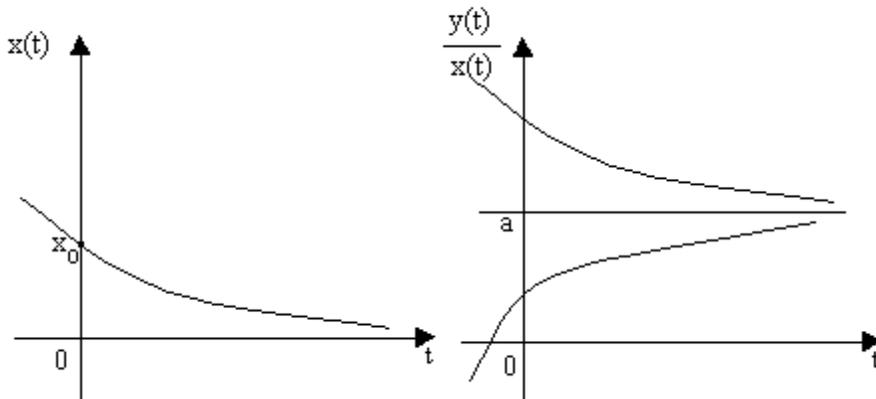


Figura 9.3

Se $\frac{y_0}{x_0} < a$, il rapporto $\frac{y(t)}{x(t)}$ cresce asintoticamente verso il valore di equilibrio a , se $\frac{y_0}{x_0} = a$ il rapporto $\frac{y(t)}{x(t)}$, rimane costante $= a$, se $\frac{y_0}{x_0} > a$, il rapporto $\frac{y(t)}{x(t)}$ decresce asintoticamente verso a .

L'andamento di $y(t)$ é piú complicato. Se $\frac{y_0}{x_0} \geq a$, $y(t)$ é sempre decrescente, tende a 0 per $t \rightarrow +\infty$ e tende a $+\infty$ per $t \rightarrow -\infty$. Se invece $\frac{y_0}{x_0} < a$, $y(t)$ tende a 0 per $t \rightarrow +\infty$ e tende a $-\infty$ per $t \rightarrow -\infty$, quindi esiste un valore $t_1 < 0$ del tempo per cui $y(t_1) = 0$. $y(t)$ dapprima cresce da 0 ad un certo valore massimo per $t = t_2$, poi decresce fino a 0. Con facili calcoli si vede che $t_1 = \frac{1}{r_2 - r_1} \log \left(1 - \frac{y_0}{ax_0} \right)$,

$t_2 = \frac{1}{r_2 - r_1} \log \frac{r_2}{r_1} \left(1 - \frac{y_0}{ax_0} \right)$ e che $t_2 > 0 \Leftrightarrow y_0 < \frac{r_1}{r_2} x_0$.

L'andamento di $y(t)$ puó essere di uno dei 5 tipi illustrati in figura 9.4, a seconda che $\frac{y_0}{x_0}$ sia minore di $\frac{r_1}{r_2}$ (curva 1), uguale a $\frac{r_1}{r_2}$ (curva 2), compreso tra $\frac{r_1}{r_2}$ e a (curva 3), uguale ad a , (curva 4), maggiore di a (curva 5):

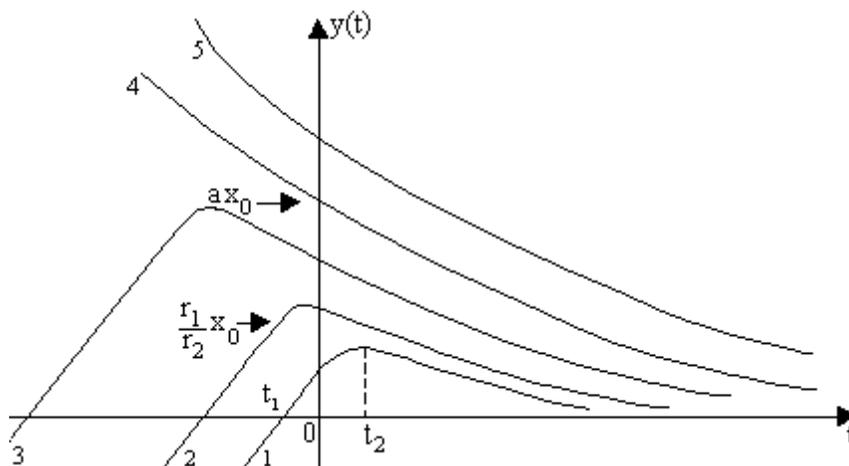


Figura 9.4

Veniamo adesso al problema della datazione radioattiva, cioè supponiamo di aver misurato i valori di $x(t)$, $y(t)$ ad un certo istante t , ottenendo 2 valori x_1, y_1 , e cerchiamo di valutare l'età t del campione, senza sapere nulla sui valori iniziali x_0, y_0 . Conviene usare la formula che dá $\frac{y(t)}{x(t)}$ e ricavare da essa t :

$$\frac{y_1}{x_1} = \left(\frac{y_0}{x_0} - a \right) e^{(r_1-r_2)t} + a \quad e^{(r_1-r_2)t} = \frac{a - \frac{y_1}{x_1}}{a - \frac{y_0}{x_0}}$$

$$t = \frac{1}{r_1 - r_2} \log \frac{a - \frac{y_1}{x_1}}{a - \frac{y_0}{x_0}} = \frac{1}{r_2 - r_1} \log \frac{a - \frac{y_0}{x_0}}{a - \frac{y_1}{x_1}}$$

A questo punto, conoscendo $\frac{y_0}{x_0}$, si troverebbe direttamente l'età t del campione. Non conoscendo $\frac{y_0}{x_0}$, si può solo dire che t dev'essere inferiore ad un certo valore \bar{t} . Si possono verificare 2 casi: o $\frac{y_1}{x_1} < a$ oppure $\frac{y_1}{x_1} > a$ (evidentemente, se fosse $\frac{y_1}{x_1} = a$, non si potrebbe concludere nulla).

Se $\frac{y_1}{x_1} < a$, deve essere anche $\frac{y_0}{x_0} < a$ e quindi:

$$t \leq \frac{1}{r_2 - r_1} \log \frac{a}{a - \frac{y_1}{x_1}} \quad (6)$$

Se $\frac{y_1}{x_1} > a$, dev'essere anche $\frac{y_0}{x_0} > a$, ma non si può concludere nulla di preciso sull'età se non si stabilisce un numero superiore M per $\frac{y_0}{x_0}$ (potrebbe essere ad esempio $M = \frac{N}{x_1}$, dove N è il numero totale di atomi del campione).

Si ha allora:

$$t = \frac{1}{r_2 - r_1} \log \frac{\frac{y_0}{x_0} - a}{\frac{y_1}{x_1} - a} \leq \frac{1}{r_2 - r_1} \log \frac{M - a}{\frac{y_1}{x_1} - a} \quad (7)$$

Esempio 9.3 Consideriamo la seguente coppia di isotopi radioattivi:

$X =$ radio 226 ($T = 1600$ anni), $Y =$ piombo 210 ($T = 22$ anni). X decade in Y , quindi, in una roccia contenente X e Y il rapporto $\frac{y(t)}{x(t)}$ tende rapidamente al valore di equilibrio $a = \frac{r_1}{r_2 - r_1} = \frac{T_2}{T_1 - T_2} = \frac{22}{1600 - 22} = 0.014$. La differenza tra $\frac{y(t)}{x(t)}$ e 0.014 decresce esponenzialmente con un periodo di dimezzamento $\frac{T_1 T_2}{T_1 - T_2} = 22.3$ anni.

La maggior parte dei minerali terrestri è di gran lunga più vecchia di 22.3 anni quindi in essi il rapporto tra radio 226 e piombo 210 è circa uguale a 0.014.

Se però una roccia è di formazione molto recente e in essa, all'epoca della sua formazione, sono stati inglobati X e Y in un rapporto diverso da 0.014, allora è possibile ricavare delle informazioni sull'età della roccia purché essa non sia troppo grande (altrimenti $\frac{y(t)}{x(t)}$ sarebbe ritornato al valore di equilibrio). Lo stesso succede se viene estratto un minerale contenente X e Y e poi esso viene trattato chimicamente, rimuovendo parte di X o di Y . In seguito, osservando il valore di $\frac{y(t)}{x(t)}$

nel minerale, si possono avere informazioni sul tempo trascorso dalla sua estrazione. ■

Questo esempio é interessante perché é stato usato per scoprire dei quadri falsi. Infatti, il piombo 210 é spesso contenuto nei colori usati in pittura; al momento della preparazione della vernice, quasi tutto il radio 226 contenuto nel minerale da cui é stato estratto il piombo 210 viene rimosso e quindi, il valore $\frac{y_0}{x_0}$ del rapporto tra piombo 210 e radio 226 nella vernice al momento della sua fabbricazione é molto maggiore del valore di equilibrio 0.014. c Però non tutto il radio 226 può essere rimosso, perciò il rapporto $\frac{y_0}{x_0}$ dev'essere inferiore ad un certo valore massimo $M \simeq 200$. Misurando il valore $\frac{y_1}{x_1}$ del rapporto $\frac{y(t)}{x(t)}$ in un quadro é possibile ricavare delle informazioni sull'età del quadro e quindi scoprire eventuali falsi.

Questo metodo é stato applicato nel 1945 ad alcuni quadri del seicento che si sospettava fossero stati dipinti in seguito e quindi, fossero falsi. In uno di essi si aveva $\frac{y_1}{x_1} = 8.5$. Usando la formula (7), si ricava che l'età t del quadro deve verificare la disuguaglianza :

$$t \leq \frac{1}{r_2 - r_1} \log \frac{200 - 0.014}{8.5 - 0.014} = \frac{1}{\frac{\log 2}{T_2} - \frac{\log 2}{T_1}} \log 23.6 = \frac{1600 \cdot 22 \text{anni}}{(1600 - 22) \log 2} \log 23.6 = 101.7 \quad \text{anni}$$

Quindi, il quadro poteva risalire al massimo al 1843, quindi era falso. Se ad esempio, si fosse trovato $\frac{y_1}{x_1} = 0.007$ (inferiore al valore di equilibrio 0.014) allora dalla formula (6) si sarebbe dedotto che :

$$t \leq \frac{1}{r_2 - r_1} \log \frac{0.014}{0.014 - 0.007} = \frac{1600 \cdot 22 \text{anni}}{(1600 - 22) \log 2} \log 2 = 22.3 \quad \text{anni (quadro falso)}.$$

In entrambi i casi, si riesce ad ottenere solo una maggiorazione dell'età.

- 3 CASO: *decadimento di un isotopo radiattivo X in un altro isotopo radiattivo y che decade piú lentamente di X*

Questo caso é perfettamente analogo al caso precedente, con l'unica differenza che $r_2 < r_1, T_2 > T_1, a = \frac{r_1}{r_2 - r_1} < 0$. Posto allora $b = -a = \frac{r_1}{r_1 - r_2} > 0$, le formule (5) di pagina 97 diventano

$$\begin{aligned} x(t) &= x_0 e^{-r_1 t} \\ y(t) &= (y_0 + b x_0) e^{-r_2 t} - b x_0 e^{-r_1 t} \quad \left(b = \frac{r_1}{r_1 - r_2} \right) \quad (8) \\ \frac{y(t)}{x(t)} &= \left(\frac{y_0}{x_0} + b \right) e^{(r_1 - r_2)t} - b \end{aligned}$$

$x(t), y(t)$ tendono entrambi a 0 per $t \rightarrow +\infty$ (com'è ovvio), ma il loro rapporto $\frac{y(t)}{x(t)}$ tende a $\rightarrow +\infty$ perché l'esponente dell'esponenziale è positivo. Perciò, col passare del tempo, X tende a sparire ed in pratica rimane soltanto Y .

Per $t \rightarrow -\infty$, $y(t)$ tenderebbe a $-\infty$, ma in realtà esiste un istante $t_1 < 0$ in cui era $y(t_1) = 0$. $y(t)$ è crescente per t che varia da t_1 a un certo istante $t_2 > t_1$, nel quale $y(t)$ raggiunge un valore massimo. Per $t > t_2$, $y(t)$ è decrescente e tende a 0 per $t \rightarrow +\infty$. Con facili calcoli si vede che $t_1 = \frac{1}{r_1 - r_2} \log \frac{bx_0}{y_0 + bx_0}$, $t_2 = \frac{1}{r_1 - r_2} \log \frac{bx_0 r_1}{(y_0 + bx_0)r_2}$ e inoltre $t_2 > 0$ se e solo se $y_0 < \frac{r_1 x_0}{r_2}$. Il rapporto $\frac{y(t)}{x(t)}$ è sempre crescente e si annulla per $t = t_1$. I grafici di $x(t), y(t), \frac{y(t)}{x(t)}$ sono del tipo

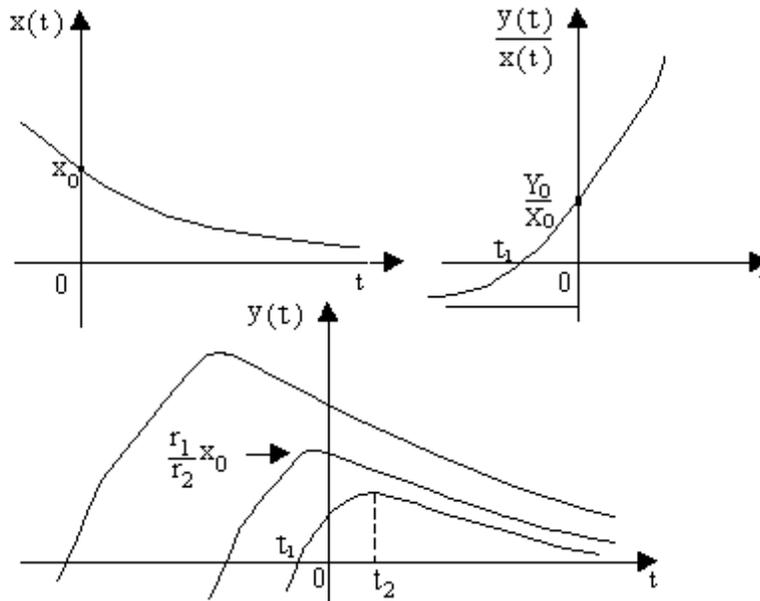


Figura 9.5

Per quanto riguarda il problema della datazione radiattiva, anche qui conviene usare la formula che dá $\frac{y(t)}{x(t)}$. Se ad un certo istante t misuriamo un valore $\frac{y_1}{x_1}$ del rapporto $\frac{y(t)}{x(t)}$, si ha

$$\frac{y_1}{x_1} = \left(\frac{y_0}{x_0} + b \right) e^{(r_1 - r_2)t} - b$$

e da qui si può ricavare l'età t del minerale:

$$e^{(r_1 - r_2)t} = \frac{b + \frac{y_1}{x_1}}{b + \frac{y_0}{x_0}} \quad t = \frac{1}{r_1 - r_2} \log \frac{b + \frac{y_1}{x_1}}{b + \frac{y_0}{x_0}}$$

conoscendo il valore iniziale $\frac{y_0}{x_0}$, si potrebbe ricavare il valore esatto di t . Ma anche non sapendo nulla di $\frac{y_0}{x_0}$, si sa che $\frac{y_0}{x_0} \geq 0$ e quindi si può ricavare una maggiorazione di t :

$$t \leq \frac{1}{r_1 - r_2} \log\left(1 + \frac{y_1}{bx_1}\right)$$

Esempio 9.4 $X =$ bismuto 210, $Y =$ polonio 210.

X decade in Y con periodo di dimezzamento $T_1 = 5$ giorni e Y decade in piombo 206 (stabile) con periodo di dimezzamento $T_2 = 138,4$ giorni. Supponiamo che un minerale contenente X e Y venga trattato chimicamente per eliminare quasi tutto Y , ma poi, dopo un certo tempo t , si misuri un rapporto $\frac{y_1}{x_1} = 100$. Allora per "l'età" del minerale (cioè il tempo trascorso dopo il trattamento) si ricava la maggiorazione seguente (tenendo conto che $r_1 = \frac{\log 2}{T_1}$, $r_2 = \frac{\log 2}{T_2}$)

$$t \leq \frac{1}{\frac{\log 2}{T_1} - \frac{\log 2}{T_2}} \log\left(1 + \frac{100}{b}\right) = \frac{T_1 T_2}{\log 2 (T_2 - T_1)} \log\left(1 + \frac{100 T_2}{T_2 - T_1}\right) = 34,75 \text{ giorni}$$

Quest'esempio è puramente ipotetico, perché i periodi di dimezzamento sono molto brevi. D'altronde in natura si verifica molto raramente che Y decada più lentamente di X , e quando questo si verifica di solito i periodi di dimezzamento sono molto brevi.

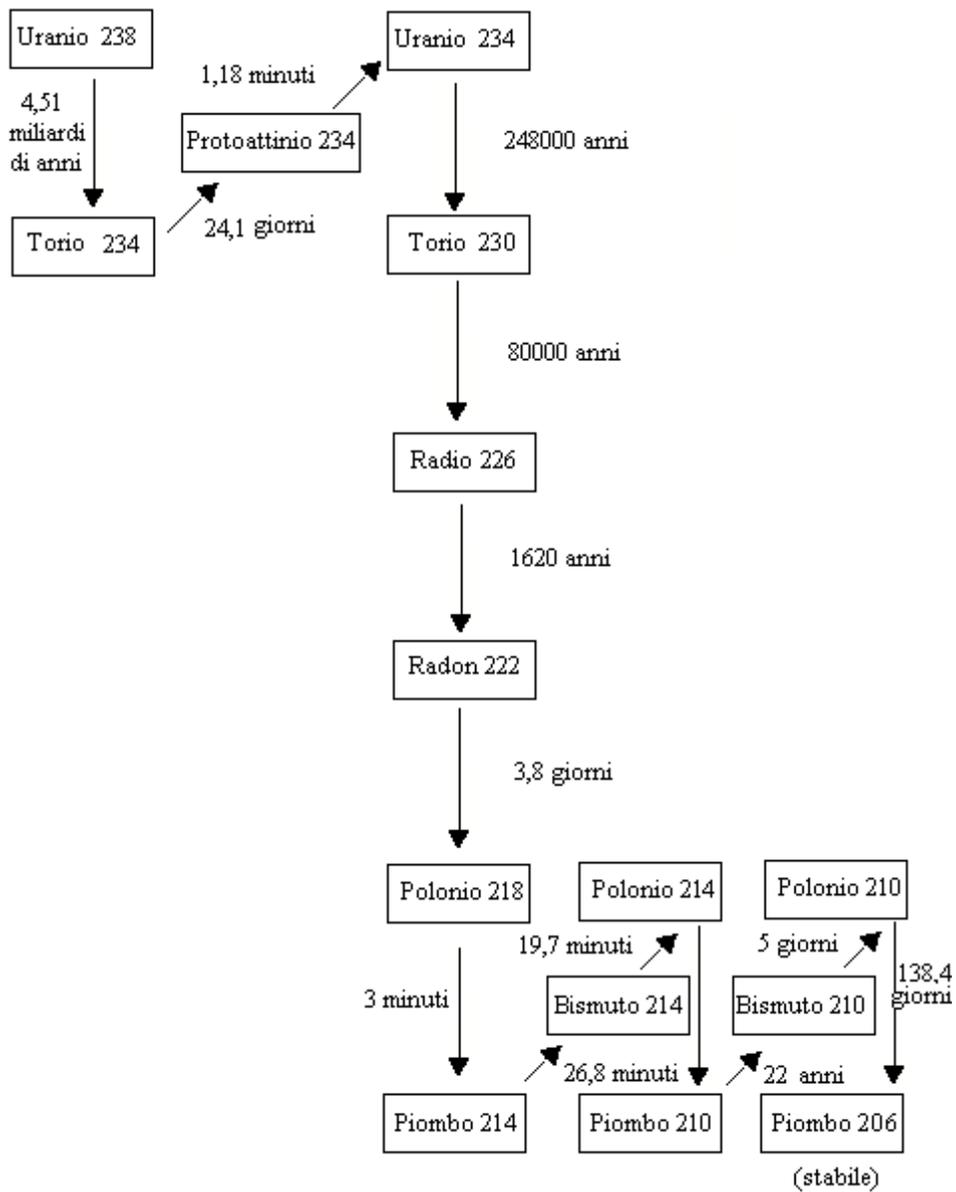
Consideriamo ad esempio lo schema che illustra il decadimento completo dell'uranio 238. Esso avviene in 14 passi successivi, che comprendono 8 decadimenti con raggi α (un nucleo radiattivo espelle una particella α formata da 2 protoni e neutroni, diminuendo di 2 la sua carica e di 4 la sua massa) e 6 decadimenti con raggi β^- (un nucleo radiattivo espelle un elettrone e contemporaneamente trasforma un neutrone in protone, aumentando di 1 la sua carica e lasciando quasi inalterata la sua massa).

Dapprima l'uranio 238 si trasforma in uranio 234 con 1 decadimento α e 2 decadimenti β^- successivi (periodi di dimezzamento = 4.51 miliardi di anni, 24.1 giorni 1.18 minuti) passando attraverso il torio 234 e il protoattinio 234. Poi avvengono 5 decadimenti α successivi (periodi di dimezzamento = 248000 anni, 80000 anni, 1620 anni, 3.8 giorni, 3 minuti), che trasformano l'uranio 234 in torio 230, radio 226, radon 222, polonio 218 e piombo 214.

Poi il piombo 214 si trasforma in piombo 210 con 2 decadimenti β^- e 1 decadimento α successivi (periodi di dimezzamento = 26.8 minuti, 19.7 minuti, 0.16 millisecondi), passando attraverso il bismuto 214 ed il polonio 214.

Infine il piombo 210 si trasforma in piombo 206 (stabile) con altri 2 decadimenti β^- e 1 decadimento α (periodi di dimezzamento = 22 anni, 5 giorni, 138, 4 giorni), passando attraverso il bismuto 210 ed il polonio 210.

■



Schema 9.1

10 Sistemi differenziali autonomi

Definizione 10.1 Un sistema differenziale autonomo del 1° ordine in forma normale nelle 2 funzioni incognite $x(t), y(t)$ è un sistema del tipo:

$$\begin{cases} x'(t) = f(x(t), y(t)) \\ y'(t) = g(x(t), y(t)) \end{cases}$$

o semplicemente

$$\begin{cases} x' = f(x, y) \\ y' = g(x, y) \end{cases} \quad (1)$$

Supponiamo che le funzioni f, g siano definite in tutto \mathbb{R}^2 e siano di classe C^1 (cioè abbiano derivate parziali continue). In tal caso vale il *teorema di esistenza ed unicità in piccolo*: se fissiamo una condizione iniziale $x(t_0) = x_0, y(t_0) = y_0$ allora il problema di Cauchy:

$$\begin{cases} x' = f(x, y) & x(t_0) = x_0 \\ y' = g(x, y) & y(t_0) = y_0 \end{cases}$$

ha 1 ed 1 sola soluzione in un opportuno intorno $(t_0 - \delta, t_0 + \delta)$ di t_0 .

Se inoltre supponiamo che f, g , oppure le loro derivate parziali, siano limitate, allora vale anche il *teorema di esistenza ed unicità in grande*, cioè il problema di Cauchy ha 1 ed 1 sola soluzione $(x(t), y(t))$ definita in tutto \mathbb{R} .

Interpretando la variabile indipendente t come tempo e le funzioni incognite $x(t), y(t)$ come le coordinate di un punto mobile nel piano x, y (detto *piano delle fasi*), il sistema differenziale (1) equivale ad assegnare un *campo di velocità*, cioè ad assegnare per ogni punto $P = (x, y)$ del piano un vettore $\vec{v}(P) = (f(x, y), g(x, y)) = (x'(t), y'(t))$, che rappresenta la velocità che deve avere il punto mobile quando passa per P . Risolvere il sistema significa trovare le funzioni $x(t), y(t)$, cioè trovare traiettoria e legge oraria del punto mobile. La traiettoria γ del punto mobile è una linea che si chiama *orbita* della soluzione considerata.

Se in un punto $\bar{P} = (\bar{x}, \bar{y})$ si ha $f(\bar{x}, \bar{y}) = g(\bar{x}, \bar{y}) = 0$, chiaramente la coppia di funzioni costanti $x(t) = \bar{x}, y(t) = \bar{y}$ soddisfa il sistema (1). Si dice che \bar{P} è un *punto di equilibrio*. La soluzione $(x(t), y(t)) = (\bar{x}, \bar{y})$ si chiama soluzione di equilibrio e la sua orbita si riduce al solo punto \bar{P} .

Valgono le seguenti proprietà

Teorema 10.1 Le soluzioni di un sistema differenziale autonomo sono invarianti per traslazioni temporali, cioè se $x(t), y(t)$ è soluzione, lo è anche $(x(t+c), y(t+c))$ qualunque sia la costante c .

Teorema 10.2 *Due orbite diverse non possono intersecarsi, cioè se $(x_1(t), y_1(t))$ e $(x_2(t), y_2(t))$ sono 2 soluzioni del sistema (1), se y_1 e y_2 sono le rispettive orbite e se $P \in \gamma_1 \cap \gamma_2$, allora $\gamma_1 = \gamma_2$ ed esiste una costante C tale che $x_2(t) = x_1(t + c)$, $y_2(t) = y_1(t + c)$, cioè le due soluzioni coincidono a meno di una traslazione temporale.*

Teorema 10.3 *Se l'orbita γ di una soluzione $(x(t), y(t))$ ripassa per uno stesso punto $P_1(x_1, y_1)$ in 2 istanti diversi t_1 e t_2 ($t_1 < t_2$), cioè se $x(t_1) = x(t_2) = x_1$, $y(t_1) = y(t_2) = y_1$, allora γ è una curva chiusa e le funzioni $x(t), y(t)$ sono periodiche di periodo $T = t_2 - t_1$.*

Teorema 10.4 *Se l'orbita γ di una soluzione $(x(t), y(t))$ si avvicina asintoticamente ad un punto $\bar{P} = (\bar{x}, \bar{y})$ cioè se $\lim_{t \rightarrow +\infty} x(t) = \bar{x}$ e $\lim_{t \rightarrow +\infty} y(t) = \bar{y}$, allora \bar{P} è un punto di equilibrio.*

Anche per i sistemi differenziali autonomi si possono dare le definizioni di punto di equilibrio stabile, instabile, asintoticamente stabile, del tutto analoghe a quelle viste per le equazioni differenziali autonome:

Definizione 10.2 *Si dice che un punto di equilibrio $P = (\bar{x}, \bar{y})$ per il sistema differenziale (1) è stabile se fissati comunque t_0 ed $\epsilon > 0$, esiste $\delta > 0$ tale che per ogni soluzione $(x(t), y(t))$ si abbia :*

$$|x(t_0) - \bar{x}| < \delta, \quad |y(t_0) - \bar{y}| < \delta \Rightarrow |x(t) - \bar{x}| < \epsilon, \quad |y(t) - \bar{y}| < \epsilon \quad \forall t > t_0$$

si dice che un punto di equilibrio è instabile se non è stabile

Definizione 10.3 *si dice che un punto di equilibrio $P = (\bar{x}, \bar{y})$ per il sistema differenziale (1) è asintoticamente stabile se fissato comunque t_0 esiste $\delta > 0$ tale che per ogni soluzione $(x(t), y(t))$ si abbia:*

$$|x(t_0) - \bar{x}| < \delta, \quad |y(t_0) - \bar{y}| < \delta \Rightarrow \lim_{t \rightarrow +\infty} x(t) = \bar{x} \text{ e } \lim_{t \rightarrow +\infty} y(t) = \bar{y}$$

Se un punto di equilibrio è asintoticamente stabile, allora è stabile (ma non viceversa).

10.1 Sistemi differenziali lineari omogenei a coefficienti costanti

Un tipo particolare di sistema differenziale autonomo è dato dai *sistemi differenziali lineari omogenei a coefficienti costanti*, la cui forma generale nel caso di due funzioni incognite $x(t), y(t)$ è la seguente:

$$\begin{cases} x' = ax + by \\ y' = cx + dy \end{cases} \quad a, b, c, d \in R \quad (2)$$

Lo studio di questo sistema é importante perché ci permette di dire qualcosa anche sui sistemi autonomi piú generali e in particolare dei loro punti di equilibrio.

Le proprietá del sistema (2) dipendono dalla *matrice dei coefficienti* $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ e dai suoi *autovalori*. Ricordiamo che un numero complesso λ si dice

autovalore di A se esiste un vettore non nullo $\vec{v} = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}$ tale che $\mathbf{A}\vec{v} = \lambda\vec{v}$

cioé $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} \lambda\alpha \\ \lambda\beta \end{pmatrix}$; il vettore \vec{v} si chiama *autovettore associato all'autovalore* λ .

É facile vedere che la coppia di funzioni $x(t) = \alpha e^{\lambda t}, y(t) = \beta e^{\lambda t}$ é soluzione del sistema(2) se e solo se λ é un autovalore di \mathbf{A} e $\vec{v} = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}$ é un autovettore associato a λ .

Si dimostra che gli autovalori λ_1, λ_2 di \mathbf{A} sono le radici dell'equazione di secondo grado $\det(\mathbf{A} - \lambda\mathbf{I}) = \det \begin{pmatrix} a - \lambda & b \\ c & d - \lambda \end{pmatrix} = 0$ nell'incognita λ . Essi possono essere reali e distinti, reali e coincidenti, oppure complessi coniugati. Nel piano delle fasi x, y , l'origine $(0, 0)$ é un punto di equilibrio del sistema (2). Posto $\delta = \det(\mathbf{A}) = \det \begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad - bc$, se $\delta \neq 0$ non ci sono altri punti di equilibrio. La natura di questo punto di equilibrio (stabile, instabile, asintoticamente stabile) e la forma delle orbite dipendono soprattutto dal segno della parte reale degli autovalori di \mathbf{A} . Si possono presentare in tutto 11 casi diversi:

- CASO 1: λ_1, λ_2 reali positivi ($0 < \lambda_1 < \lambda_2$).

La soluzione generale del sistema (2) é del tipo

$$\begin{cases} x(t) = C_1 e^{\lambda_1 t} + C_2 e^{\lambda_2 t} + \\ y(t) = C_3 e^{\lambda_1 t} + C_4 e^{\lambda_2 t} \end{cases}$$

Calcolando $x'(t), y'(t)$ da queste formule e sostituendole nel sistema (2), si trovano le 2 costanti C_3, C_4 in funzione di C_1, C_2 e cosí nell'espressione di $x(t), y(t)$ restano solo 2 costanti arbitrarie C_1, C_2 (questo succede sempre, perché l'insieme delle soluzioni del sistema (2) é uno spazio vettoriale di dimensione 2).

Le orbite delle soluzioni sono curve che si allontanano dall'origine 0 e "provengono asintoticamente da 0" (cioé $x(t) \rightarrow 0$ e $y(t) \rightarrow 0$ per $t \rightarrow -\infty$), mentre per $t \rightarrow +\infty$ si allontanano indefinitamente da 0. Se \vec{v}_1, \vec{v}_2 sono 2 autovettori associati agli autovalori λ_1, λ_2 rispettivamente, le orbite provengono dalla direzione di \vec{v}_1 (cioé in 0 sono tutte tangenti a \vec{v}_1) e tendono tutte a disporsi nella direzione di \vec{v}_2 (cioé per

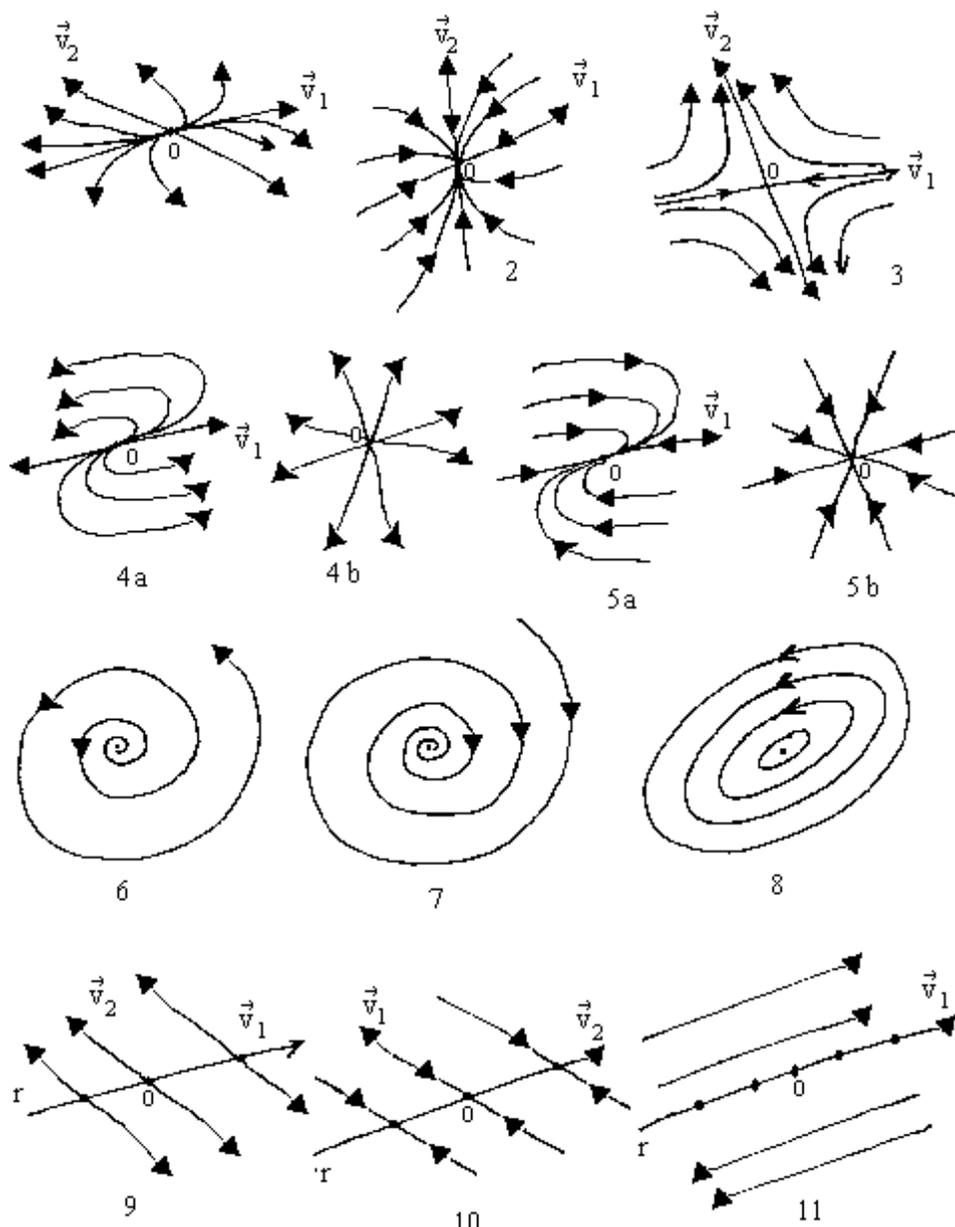


Figura 10.1

$t \rightarrow +\infty$ la tangente all'orbita tende ad essere parallela a \vec{v}_2). 0 é un punto di equilibrio instabile e prende il nome di *nodo instabile*.

Esempio numerico : $A = \begin{pmatrix} 1 & -2 \\ 1 & 4 \end{pmatrix}$

- CASO 2: λ_1, λ_2 reali negativi ($\lambda_1 < \lambda_2 < 0$).

Questo caso é perfettamente analogo al precedente, salvo che per il verso di percorrenza delle orbite. Per $t \rightarrow +\infty, x(t) \rightarrow 0$ e $y(t) \rightarrow 0$, le orbite si avvicinano asintoticamente all'origine e nell'origine sono tutte tangenti a \vec{v}_2 . Si dice che l'origine é un *nodo asintoticamente stabile*.

Esempio numerico $A = \begin{pmatrix} -2 & 1 \\ 2 & -3 \end{pmatrix}$

- CASO 3: λ_1, λ_2 reali discordi ($\lambda_1 < 0 < \lambda_2$).

Anche in questo caso la soluzione generale del sistema é del tipo

$$\begin{cases} x(t) = C_1 e^{\lambda_1 t} + C_2 e^{\lambda_2 t} + \\ y(t) = C_3 e^{\lambda_1 t} + C_4 e^{\lambda_2 t} \end{cases}$$

Peró le orbite provengono dall'infinito (dalla direzione di \vec{v}_1) e tendono ancora all'infinito (nella direzione di \vec{v}_2). L'origine é un punto di equilibrio instabile e prende il nome di *punto sella*.

Esempio numerico $A = \begin{pmatrix} 1 & 4 \\ 3 & 2 \end{pmatrix}$

- CASO 4: Autovalori reali coincidenti positivi ($\lambda_1 = \lambda_2 = \lambda > 0$)

In questo caso possono presentarsi 2 sottocasi diversi: 4a e 4b.

CASO 4 a: gli autovettori associati all'autovalore λ formano uno spazio vettoriale di dimensione 1, cioè sono tutti multipli di un certo vettore $\vec{v}_1 \neq 0$. In tal caso l'andamento delle orbite é quello indicato nella figura 10.1 - 4a: le orbite provengono asintoticamente dall'origine e tendono all'infinito, nell'origine sono tutte tangenti a \vec{v}_1 e per $t \rightarrow +\infty$ tendono nuovamente a disporsi nella direzione di \vec{v}_1 . La soluzione generale del sistema (2) é del tipo

$$\begin{cases} x(t) = C_1 e^{\lambda t} + C_2 t e^{\lambda t} \\ y(t) = C_3 e^{\lambda t} + C_4 t e^{\lambda t} \end{cases}$$

Calcolando $x'(t), y'(t)$ da queste formule e sostituendo nel sistema (2) si trovano le 2 costanti C_3, C_4 in funzione di C_1, C_2 .

Esempio numerico: $A = \begin{pmatrix} 3 & 1 \\ -1 & 1 \end{pmatrix}$.

CASO 4 b: gli autovettori associati a λ formano uno spazio vettoriale di dimensione 2, cioè qualunque vettore $\begin{pmatrix} \alpha \\ \beta \end{pmatrix} \in \mathbb{R}^2$ é un autovettore. Questo é possibile solo se la matrice \mathbf{A} coincide con $\begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix}$. In

tal caso il sistema (2) diventa $\begin{cases} x' = \lambda x \\ y' = \lambda y \end{cases}$ e la sua soluzione generale

$\begin{cases} x = C_1 e^{\lambda t} \\ y = C_2 e^{\lambda t} \end{cases}$. Le orbite sono semirette uscenti dall'origine (da cui provengono asintoticamente) e tendono all'infinito (vedi figura 10.1 - 4b). In entrambi i casi, l'origine è un punto di equilibrio instabile (*nodo instabile degenere*).

- CASO 5: Autovalori reali coincidenti negativi ($\lambda_1 = \lambda_2 = \lambda < 0$)
 Questo caso è perfettamente analogo al caso 4, salvo che per il verso di percorrenza delle orbite. Possono darsi 2 sottocasi 5 a e 5 b, a seconda che l'insieme degli autovettori associati a λ abbia dimensione 1 o 2. In entrambi i casi l'origine è un punto di equilibrio asintoticamente stabile (*nodo degenere asintoticamente stabile*).

Esempi numerici $A = \begin{pmatrix} -2 & -1 \\ 1 & -4 \end{pmatrix}$ per il caso 5 a, $A = \begin{pmatrix} -3 & 0 \\ 0 & -3 \end{pmatrix}$ per il caso 5b.

- CASO 6: Autovalori complessi coniugati con parte reale positiva ($\lambda_1 = a + ib, \lambda_2 = a - ib$ con $a > 0$)

In tal caso la soluzione del sistema (2) è del tipo

$$\begin{cases} x(t) = C_1 e^{at} \cos bt + C_2 e^{at} \sin bt \\ y(t) = C_3 e^{at} \cos bt + C_4 e^{at} \sin bt \end{cases}$$

Calcolando $x'(t), y'(t)$ da queste formule e sostituendo al sistema (2), si trovano le costanti C_3, C_4 in funzione di C_1, C_2 . Per $t \rightarrow -\infty, x(t)$ e $y(t)$ tendono a 0 e per $t \rightarrow +\infty, x^2(t) + y^2(t) \rightarrow +\infty$, quindi le orbite si allontanano asintoticamente dall'origine e tendono all'infinito. La presenza dei seni e dei coseni fa sì che le orbite "spiraleggino", facendo infiniti giri intorno all'origine fig. 10.1 -6. Il fatto che spiraleggino in senso orario o antiorario dipende dal segno delle costanti C_1, C_2, C_3, C_4 . L'origine è un punto di equilibrio instabile (*fuoco instabile*).

Esempio numerico $A = \begin{pmatrix} 1 & 1 \\ -1 & 3 \end{pmatrix}$.

- CASO 7: Autovalori complessi coniugati con parte reale negativa ($\lambda_1 = a + ib, \lambda_2 = a - ib$ con $a < 0$) La situazione è perfettamente analoga a quella del caso 6, salvo che per il verso di percorrenza delle orbite. Le orbite provengono dall'infinito e si avvicinano asintoticamente all'origine, spiraleggiando infinite volte intorno ad essa. Fig. 10.1 -7. L'origine è un punto di equilibrio asintoticamente stabile (*fuoco asintoticamente stabile*).

Esempio numerico $A = \begin{pmatrix} -1 & 2 \\ -2 & -1 \end{pmatrix}$.

- CASO 8: Autovalori immaginari puri ($\lambda_1 = ib, \lambda_2 = -ib$ con $b > 0$) In tal caso la soluzione del sistema (2) è del tipo

$$\begin{cases} x(t) = C_1 \cos bt + C_2 \sin bt \\ y(t) = C_3 \cos bt + C_4 \sin bt \end{cases}$$

Calcolando $x'(t), y'(t)$ da queste formule e sostituendo al sistema (2), si trovano C_3, C_4 in funzione di C_1, C_2 . Le orbite sono curve chiuse (per la precisione ellissi, figura 10.1 - 8.) e le funzioni $x(t), y(t)$ sono periodiche di periodo $T = \frac{2\pi}{b}$. Il fatto che le orbite siano percorse in senso orario o antiorario dipende dal segno delle costanti C_1, C_2, C_3, C_4 . L'origine è un punto di equilibrio stabile, ma non asintoticamente stabile e si chiama *centro*.

Esempio numerico $A = \begin{pmatrix} 1 & 1 \\ -2 & -1 \end{pmatrix}$.

Restano da determinare i casi in cui $\det A = 0$, il che accade se e solo se almeno uno dei 2 autovalori $\lambda_1 \lambda_2$ è nullo. Se questo accade gli autovalori sono reali, l'origine non è l'unico punto di equilibrio e si possono avere i seguenti casi 9, 10, 11.

- CASO 9: $\lambda_1 = 0, \lambda_2 > 0$ In tal caso la soluzione generale del sistema (2) è data da $(x(t), y(t)) = C_1 \vec{v}_1 + C_2 \vec{v}_2 e^{\lambda_2 t}$, dove \vec{v}_1, \vec{v}_2 sono 2 autovettori associati rispettivamente a λ_1, λ_2 . Tutti i punti della retta r passante per l'origine e parallela a \vec{v}_1 sono punti di equilibrio instabile. Le orbite sono semirette parallele a \vec{v}_2 figura 10.1 - 9. che si allontanano asintoticamente dai punti di equilibrio (nei 2 versi opposti) e tendono all'infinito.

Esempio numerico $A = \begin{pmatrix} 1 & 2 \\ 3 & 6 \end{pmatrix}$.

- CASO 10: $\lambda_1 < 0, \lambda_2 = 0$ È perfettamente analogo al caso 9, salvo il verso di percorrenza delle orbite. Tutti i punti della retta r passante per l'origine e parallela a \vec{v}_2 sono punti di equilibrio stabile, ma non asintoticamente stabile. Le orbite sono semirette parallele a \vec{v}_1 , che tendono asintoticamente ad un punto di equilibrio.

Esempio numerico $A = \begin{pmatrix} 1 & 2 \\ -2 & -4 \end{pmatrix}$.

- CASO 11: $\lambda_1 = \lambda_2$ Possono presentarsi due sottocasi diversi (11 a e 11 b)

CASO 11 a: gli autovettori associati all'autovalore 0 sono tutti multipli di un certo vettore $\vec{v}_1 \neq 0$. In tal caso la soluzione generale del

sistema (2) é del tipo

$$\begin{cases} x(t) = C_1 + C_2 t \\ y(t) = C_3 + C_4 t \end{cases}$$

Calcolando $x'(t), y'(t)$ da queste formule e sostituendo nel sistema (2) si trovano C_3, C_4 , in funzione di C_1, C_2 . Tutti i punti della retta r passante per l'origine e parallela a \vec{v}_1 sono punti di equilibrio instabile. Le orbite sono rette parallele a r , percorse in un verso o nel verso opposto opposto a seconda che si trovino in uno o nell'altro dei 2 semipiani delimitati da r figura 10.1 - 11.

Esempio numerico $A = \begin{pmatrix} 6 & 3 \\ -12 & -6 \end{pmatrix}$

CASO 11 b: qualunque vettore $\vec{v} = \begin{pmatrix} a \\ b \end{pmatrix}$ del piano é un autovettore associato all'autovalore 0. Questo é possibile solo se la matrice A é la matrice nulla $\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$. Allora il sistema (2) si riduce a $x' = y' = 0$, cio'è $x = C_1, y = C_2$ (C_1, C_2 costanti qualsiasi). Tutti i punti del piano sono punti di equilibrio (stabile, ma non asintoticamente stabile).

10.2 Sistemi differenziali autonomi: studio qualitativo

Sia $\bar{P} = (\bar{x}, \bar{y})$ un punto di equilibrio del generico sistema differenziale autonomo:

$$\begin{cases} x' = f(x, y) \\ y' = g(x, y) \end{cases} \quad (1)$$

In certi casi la natura del punto \bar{P} (cioé la sua stabilitá, la sua stabilitá asintotica, l'andamento delle orbite in un intorno di \bar{P}) puó essere dedotta dallo studio della matrice jacobiana:

$$J(x, y) = \begin{pmatrix} \frac{\partial f}{\partial x} & \frac{\partial f}{\partial y} \\ \frac{\partial g}{\partial x} & \frac{\partial g}{\partial y} \end{pmatrix}$$

nel punto \bar{P} . Si puó mostrare infatti il seguente:

Teorema 10.5 *Se f, g sono di classe C^2 (cioé hanno derivate parziali seconde continue) indichiamo con λ_1, λ_2 gli autovalori della matrice jacobiana $J(\bar{x}, \bar{y})$, calcolata nel punto di equilibrio $\bar{P} = (\bar{x}, \bar{y})$. Allora :*

1. *Se λ_1, λ_2 hanno entrambi parte reale negativa, \bar{P} é asintoticamente stabile.*

2. Se almeno uno dei 2 autovalori ha parte reale positiva, \bar{P} é instabile.
3. Negli altri casi (autovalori reali di cui uno nullo e l'altro ≤ 0 , oppure autovalori immaginari puri) non si puó dire nulla.

Inoltre, se entrambi gli autovalori hanno parte reale $\neq 0$, esiste un intorno U di \bar{P} in cui le orbite hanno un comportamento del tipo di quello indicato nelle figura 10.1 (nodo instabile - nodo asintoticamente stabile - sella - fuoco instabile - fuoco asintoticamente stabile).

Notiamo che, in generale il sistema (2) non si sa risolvere (se non con metodi numerici approssimanti al computer); però in alcuni casi é possibile determinare esplicitamente l'equazione delle orbite. Quando questo non é possibile, a volte si puó studiare l'andamento qualitativo delle orbite esaminando il segno di x', y' , cioè di $f(x, y), g(x, y)$.

Infatti consideriamo nel piano xy le linee γ_1, γ_2 di equazione rispettivamente $f(x, y) = 0, g(x, y) = 0$. Queste linee sono dette *nullicline*; su γ_1 si ha $x' = 0$, su γ_2 si ha $y' = 0$. L'intersezione $\gamma_1 \cap \gamma_2$ non é altro che l'insieme dei punti di equilibrio del sistema (2).

In genere γ_1 e γ_2 dividono il piano in un certo numero di regioni connesse, all'interno di ciascuna delle quali f, g hanno segno costante. Sia A una di queste regioni: studiando il segno di f, g in A (il che é molto facile: basta vedere che segno hanno f, g in un punto qualsiasi $P \in A$) si trova il segno di $x'(t), y'(t)$ in A e quindi si hanno informazioni sull'andamento delle orbite in A .

Per esempio, se in A si ha $f > 0, g > 0$ allora $x(t), y(t)$ devono essere entrambe crescenti in A , cioè le orbite in A "si dirigono verso Nord - Est nel piano xy ". Negli altri 3 casi ($f > 0, g < 0$, $f < 0, g > 0$, $f < 0, g < 0$) le orbite in A si dirigono rispettivamente verso Sud - Est, Nord - Ovest, Sud - Ovest. Inoltre, nei punti in cui $f(x, y) \neq 0$ (cio'è in $\mathbb{R}^2 - \gamma_1$) é possibile dividere la seconda equazione del sistema (2) per la prima, ottenendo:

$$\frac{g(x, y)}{f(x, y)} = \frac{y'}{x'} = \frac{\frac{dy}{dx}}{\frac{dy}{dt}} = \frac{dy}{dx} \quad \text{cioé} \quad \frac{dy}{dx} = \frac{g(x, y)}{f(x, y)}$$

Questa é un'equazione differenziale del 1 ordine nella funzione incognita $y(x)$, che rappresenta l'equazione di una generica orbita; se la si sa risolvere, si trova esplicitamente l'equazione delle orbite (almeno in $\mathbb{R}^2 - \gamma_1$).

Esempio 10.1 Consideriamo il sistema

$$\begin{cases} x' = xy \\ y' = x^2 \end{cases}$$

le nullicline sono

$$\gamma_1 \quad xy = 0 \text{ (unione dell'asse } x \text{ e dell'asse } y)$$

$\gamma_2 \quad x^2 = 0$ (asse y)

$\gamma_1 \cap \gamma_2 =$ asse y , quindi tutti i punti dell'asse y sono punti di equilibrio. Il segno di x', y' é $++$ nel 1 e 3 quadrante, $-+$ nel secondo e quarto quadrante. Nei punti dell'asse x diversi da $(0, 0)$ si ha $x' = 0, y' > 0$

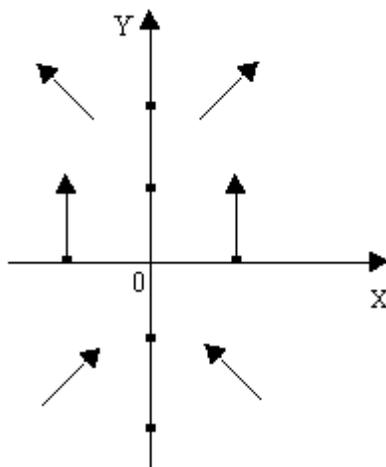


Figura 10.2

. Ciò potrebbe suggerire un andamento delle orbite simile a quello di figura 10.3

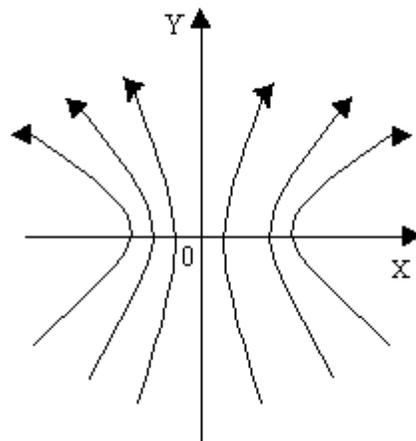


Figura 10.3

, ma invece non é cosí. Infatti cercando di trovare l'equazione delle orbite si ottiene:

$$\frac{dy}{dx} = \frac{y'}{x'} = \frac{x^2}{xy} = \frac{x}{y} \quad ydy = xdx \quad \int ydy = \int xdx + C$$

$$\frac{y^2}{2} = \frac{x^2}{2} + C \quad y^2 - x^2 = 2C$$

Si tr.
figur

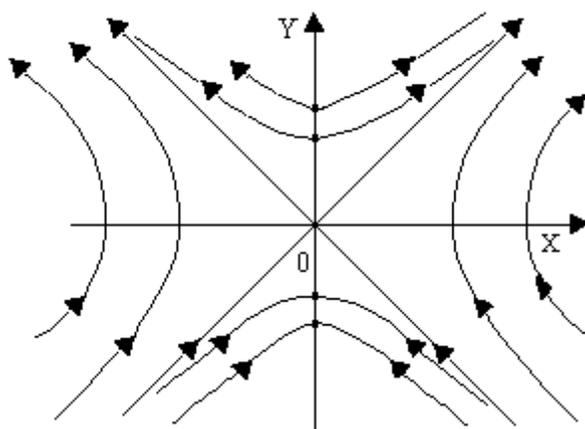


Figura 10.4

. Il generico punto di equilibrio $\bar{P} = (\bar{0}, \bar{y})$ é instabile se $y \geq 0$, stabile, ma non asintoticamente stabile se $\bar{y} < 0$. La matrice jacobiana del sistema é

$$J(x, y) = \begin{pmatrix} \frac{\partial f}{\partial x} & \frac{\partial f}{\partial y} \\ \frac{\partial g}{\partial x} & \frac{\partial g}{\partial y} \end{pmatrix} = \begin{pmatrix} y & x \\ 2x & 0 \end{pmatrix}$$

In \bar{P} si ha $J(0, \bar{y}) = \begin{pmatrix} \bar{y} & 0 \\ 0 & 0 \end{pmatrix}$ che ha i 2 autovalori $0, \bar{y}$, quindi il teorema (10.5) ci permette di concludere che \bar{P} é instabile solo se $\bar{y} > 0$, non ci permette di concludere niente se $\bar{y} \neq 0$. ■

Un particolare comportamento delle orbite si ha quando esse si avvicinano ad un cosiddetto *ciclo limite*, come mostra il seguente

Esempio 10.2

$$\begin{cases} x' = -y + x(1 - x^2 - y^2) \\ y' = x + y(1 - x^2 - y^2) \end{cases}$$

Si vede facilmente che l'unico punto di equilibrio é $(0, 0)$. L'equazione delle orbite non é facile da trovare e neanche il loro studio qualitativo. Però ponendo $z = x^2 + y^2$ (z rappresenta il quadrato della distanza del punto mobile $(x(t), y(t))$ dall'origine) si trova che

$$\begin{aligned} z' &= \frac{dz}{dt} = 2xx' + 2yy' = 2x(-y + x(1 - x^2 - y^2)) + 2y(x + y(1 - x^2 - y^2)) \\ &= 2x^2(1 - z) + 2y^2(1 - z) = 2z(1 - z) \end{aligned}$$

$$z' = 2z(1 - z)$$

Questa é un'equazione differenziale autonoma del 1 ordine in $z(t)$, anzi si tratta dell'equazione $z' = 2z(1 - z)$. È facile studiare qualitativamente le sue soluzioni: al variare della

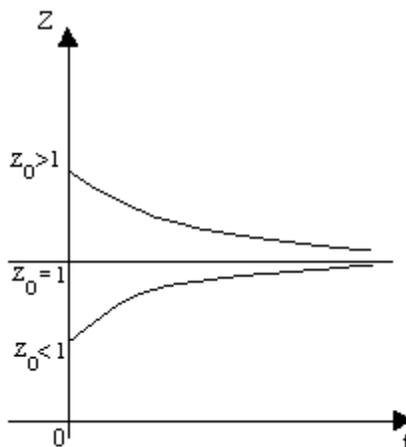


Figura 10.5

Quindi l'orbita di un generico $P_0 = (x_0, y_0)$ diverso dal punto di equilibrio $\bar{P} = (0, 0)$ tende sempre ad avvicinarsi alla circonferenza γ di equazione $x^2 + y^2 = 1$. Però, siccome su γ non ci sono punti di equilibrio, si dimostra che l'orbita deve avvicinarsi a γ "spiraleggiando" infinite volte (figura 10.6)

.Si dice che γ é un *ciclo limite*. Il punto di equilibrio $\bar{P} = (0, 0)$ é un fuoco instabile; infatti la matrice jacobiana $J(0, 0) = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}$ ha come autovalori $1 \pm i$. ■

Abbiamo visto finora che alcuni possibili comportamenti delle orbite sono :

1. orbite ridotte ad un punto (che deve essere di equilibrio)

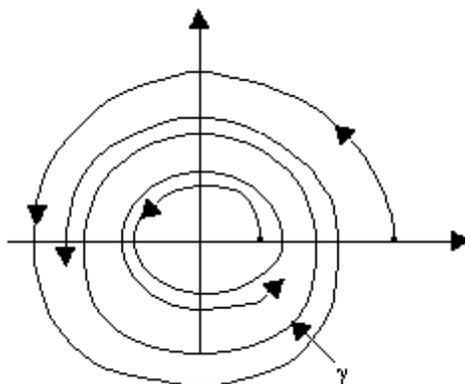


Figura 10.6

2. orbite che tendono asintoticamente ad un punto di equilibrio
3. orbite che tendono all'infinito
4. orbite cicliche
5. orbite che tendono ad un ciclo limite

Questi 5 comportamenti non esauriscono tutti i possibili comportamenti delle orbite, tuttavia vale il seguente

Teorema 10.6 (*Poincaré - Bendixson*)

Dato il sistema $\begin{cases} x' = f(x, y) \\ y' = g(x, y) \end{cases}$ con f, g di classe C^1 in \mathbb{R}^2 , se una soluzione $x(t), y(t)$ è sempre contenuta all'interno di una regione limitata A del piano (x, y) e A non contiene punti di equilibrio, allora la soluzione in questione è ciclica o tende ad un ciclo limite.

Osserviamo che una generica equazione differenziale del secondo ordine in una funzione incognita $x(t)$

$$F(t, x(t), x'(t), x''(t)) = 0$$

si può sempre ricondurre ad un sistema di 2 equazioni differenziali del primo ordine introducendo una funzione incognita ausiliaria $y(t) = x'(t)$. Infatti l'equazione considerata è equivalente al sistema:

$$\begin{cases} x' = y \\ F = (t, x, y, y') = 0 \end{cases}$$

In generale, un'equazione differenziale di ordine n è equivalente ad un sistema di n equazioni differenziali del primo ordine. In fisica s'incontrano spesso

equazioni differenziali del secondo ordine, che si possono ricondurre ad un sistema di 2 equazioni differenziali del primo ordine , come dimostrano i seguenti esempi.:

Esempio 10.3 OSCILLAZIONI DI UN CORPO ATTACCATO AD UNA MOLLA, SENZA ATTRITO L'equazione del moto é

$$mx'' = -kx$$

dove $x(t)$ = posizione del corpo mobile, m = massa del corpo, k = costante elastica della molla. Posto $y(t) = x'(t)$, l'equazione é equivalente al sistema autonomo lineare a coefficienti costanti:

$$\begin{cases} x'(t) = y \\ y' = -\frac{k}{m}x \end{cases}$$

L'unico punto di equilibrio é $(0, 0)$ e la matrice A dei coefficienti é $\begin{pmatrix} 0 & 1 \\ -\frac{k}{m} & 0 \end{pmatrix}$, che ha gli autovalori $\pm\sqrt{\frac{k}{m}}$. Perció $(0, 0)$ é un centro (stabile, ma non asintoticamente stabile) e le orbite sono ellissi aventi come assi gli assi x, y (vedi Figura 10.7).

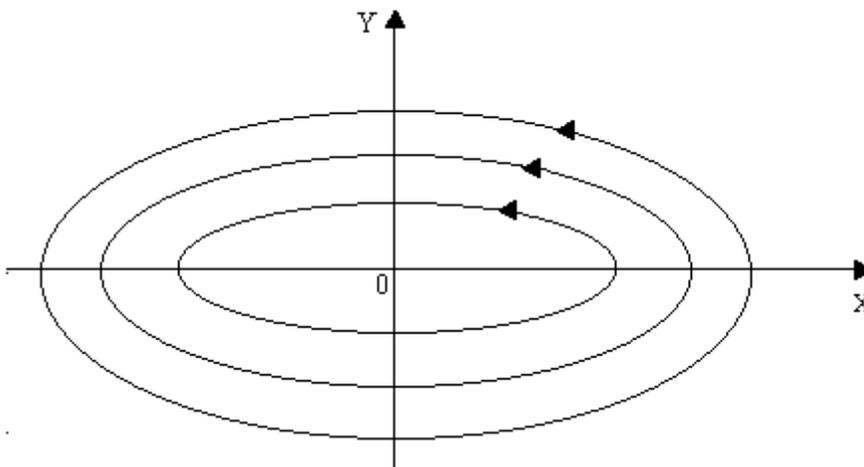


Figura 10.7

Esempio 10.4 OSCILLAZIONI DI UN CORPO ATTACCATO AD UNA MOLLA, CON ATTRITO

L'equazione del moto é

$$mx'' = -kx - rx'$$

dove $x(t)$ = posizione del corpo mobile, m = massa del corpo, k = costante elastica della molla, r = resistenza dell'aria. Posto $y(t) = x'(t)$, l'equazione é equivalente al sistema

$$\begin{cases} x' = y \\ y' = -\frac{k}{m}x - \frac{r}{m}y \end{cases}$$

la matrice A dei coefficienti é $\begin{pmatrix} 0 & 1 \\ -\frac{k}{m} & -\frac{r}{m} \end{pmatrix}$ ha i 2 autovalori $\frac{-r \pm \sqrt{r^2 - 4mk}}{2}$.

L'origine é l'unico punto di equilibrio ed é un nodo asintoticamente stabile se $r^2 > 4mk$ (autovalori reali distinti, entrambi negativi: smorzamento supercritico, non ci sono oscillazioni) oppure $r^2 = 4mk$ (autovalori reali coincidenti $= -\frac{r}{2} < 0$: smorzamento critico); é un fuoco asintoticamente stabile se $r^2 < 4mk$ (autovalori complessi coniugati con parte reale negativa: oscillazioni smorzate). ■

11 Competizione tra specie

Consideriamo due diverse specie animali X e Y che si nutrono dello stesso cibo ed indichiamo con $x(t)$, $y(t)$ il numero di individui delle specie A e B al variare del tempo t . Il tasso di crescita $r(t)$ della specie A é dato da $r(t) = x'(t)/x(t) = r_1(t) - r_2(t)$, dove $r_1(t)$ rappresenta il tasso di natalità (pari alla probabilità che un individuo ha di generarne un altro nell'unità di tempo) ed $r_2(t)$ rappresenta il tasso di mortalità (pari alla probabilità che ha un individuo di morire nell'unità di tempo).

Supponiamo che $r_2(t)$ sia costante ($= r_2$) e che $r_1(t)$ sia proporzionale alle riserve di cibo non utilizzate, cioè:

$$r_1(t) = k(M - ax(t) - by(t))$$

dove M = riserve totali di cibo(che supponiamo rinnovabili), a = quantità di cibo consumata nell'unità di tempo da un individuo della specie X , b = quantità di cibo consumata nell'unità di tempo da un individuo della specie B . Allora, si ha:

$$r(t) = x'(t)/x(t) = k(M - ax(t) - by(t)) - r_2$$

ció abbiamo la seguente equazione differenziale:

$$x'(t) = x(t) (A - Bx(t) - Cy(t)) \quad \text{dove} \quad A = kM - r_2, B = ka, C = kb$$

In assenza di individui della specie Y , $x(t)$ seguirebbe l'equazione differenziale $x'(t) = x(t) (A - Bx(t))$ (crescita logistica), dove A rappresenta il tasso di crescita di $x(t)$ quando $x(t)$ é molto piccolo (per essere precisi, A é il limite del tasso di crescita di $x(t)$ quando $x(t)$ tende a 0).

Supponiamo che A sia positivo, altrimenti la specie X si estinguerebbe sicuramente. Le costanti B e C , per come sono state definite, sono sicuramente positive. Considerazioni analoghe valgono per la specie Y e quindi, possiamo rappresentare il problema della competizione tra 2 specie X e Y per il cibo con il seguente sistema differenziale autonomo:

$$x' = f(x, y) = x(A - Bx - Cy)$$

$$y' = g(x, y) = y(D - Ex - Fy)$$

Questo sistema non può essere risolto esplicitamente, ma solo numericamente con il calcolatore (determinando $x(t)$ e $y(t)$ con la precisione desiderata). Però si possono determinare i punti di equilibrio, studiare la loro stabilità, eseguire lo studio qualitativo delle soluzioni, stabilire se $x(t)$ e $y(t)$ sono crescenti o decrescenti e studiare la forma delle orbite.

Osserviamo che $x(t)$ e $y(t)$, in quanto rappresentano le numerosità di 2 popolazioni di esseri viventi, devono esseri numeri positivi. Perciò, se rappresentiamo le soluzioni nel piano cartesiano xy ($x(t)$ in ascisse e $y(t)$ in ordinate), ci interessa solo il primo quadrante (semiassi compresi). Cerchiamo i punti di equilibrio compresi nel primo quadrante uguagliando a 0 $f(x, y)$ e $g(x, y)$:

$$x(A - Bx - Cy) = 0$$

$$y(D - Ex - Fy) = 0$$

Queste 2 equazioni rappresentano 2 curve C_1 e C_2 (le nullicline) tali che su C_1 si ha $x'(t) = 0$ e su C_2 si ha $y'(t) = 0$. C_1 é formata dall'unione di 2 rette: l'asse y ($x = 0$) e la retta r di equazione $A - Bx - Cy = 0$, che incontra gli assi x, y nei 2 punti $P = (A/B, 0)$ e $Q = (0, A/C)$.

Invece C_2 é formata dall'unione dell'asse x ($y = 0$) con la retta s di equazione $D - Ex - Fy = 0$, che incontra gli assi x, y nei 2 punti $R = (D/E, 0)$ e $S = (0, D/F)$.

L'insieme dei punti di equilibrio non é altro che l'intersezione di C_1 con C_2 e contiene 4 punti: l'origine $O = (0, 0)$, il punto $P = (A/B, 0)$, il punto $S = (0, D/F)$ ed il punto $T = ((CD - AF)/(CE - BF), (AE - BD)/(CE - BF))$ intersezione delle 2 rette r, s . Il punto T esiste solo se $CE - BF$ é diverso da 0; se $CE - BF$ uguale a 0, le 2 rette r, s sono parallele ed i punti di equilibrio sono soltanto 3. Però il punto T puó trovarsi o meno nel primo quadrante, a seconda dei valori delle 6 costanti A, B, C, D, E, F .

Possono presentarsi vari casi, di cui i principali sono 4, a seconda che il punto P sia a sinistra o a destra di R (cioé $A/B <=> D/E$) e che il punto Q stia sotto o sopra S (cioé $A/C <=> D/F$). Facciamo 4 esempi numerici, uno per ciascuno dei 4 casi.

Esempio 11.1

$$\begin{aligned} x' &= x(1 - x - 2y) && \text{figura 11.1} \\ y' &= y(1 + -2x - y) \end{aligned}$$

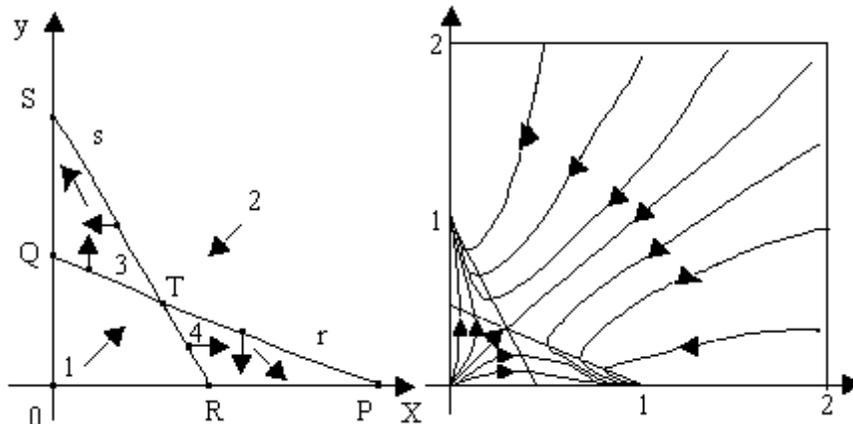


Figura 11.1

Figura 11.2

Le 2 rette r, s s'incontrano nel punto $T = (1/3, 1/3)$. Ci sono 4 punti di equilibrio: $O = (0, 0), P = (1, 0), S = (0, 1), T = (1/3, 1/3)$. La loro stabilità si può stabilire mediante la matrice jacobiana $J(x, y)$:

$$f(x, y) = x - x^2 - 2xy \quad g(x, y) = y - y^2 - 2xy$$

$$J(x, y) = \begin{pmatrix} \frac{\partial f}{\partial x} & \frac{\partial f}{\partial y} \\ \frac{\partial g}{\partial x} & \frac{\partial g}{\partial y} \end{pmatrix} = \begin{pmatrix} 1 - 2x - 2y & -2x \\ -2y & 1 - 2x - 2y \end{pmatrix}$$

$$J(0, 0) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \rightarrow \text{gli autovalori sono } 1, 1$$

$$J(1, 0) = \begin{pmatrix} -1 & -2 \\ 0 & -1 \end{pmatrix} \rightarrow \text{gli autovalori sono } -1, -1$$

$$J(0, 1) = \begin{pmatrix} -1 & 0 \\ -2 & -1 \end{pmatrix} \rightarrow \text{gli autovalori sono } -1, -1$$

$$J(1/3, 1/3) = \begin{pmatrix} -1/3 & -2/3 \\ -2/3 & -1/3 \end{pmatrix} \rightarrow \text{gli autovalori sono } -1, 1/3$$

Quindi, P ed S sono 2 punti di equilibrio asintoticamente stabili, mentre O e T sono instabili (O è un nodo instabile, T è una sella).

Le 2 rette r, s , dividono il primo quadrante in 4 zone 1,2,3,4 (fig.11.1), in ciascuna delle quali $x'(t) = f(x(t), y(t))$ e $y'(t) = g(x(t), y(t))$ hanno segno costante, quindi si può studiare il comportamento qualitativo delle orbite.

Precisamente:

1. Nella zona 1 si ha $x'(t) > 0, y'(t) > 0$ e quindi $x(t), y(t)$ sono funzioni crescenti di t e l'orbita di un punto qualsiasi si dirige verso l'alto e verso destra, come indica la freccia in fig.11.1

2. Nella zona 2 si ha $x'(t) < 0, y'(t) < 0$ e quindi, le orbite si dirigono in basso a sinistra.
3. Nella zona 3 si ha $x'(t) < 0, y'(t) > 0$ e quindi, le orbite si dirigono in alto a sinistra.
4. Nella zona 4 si ha $x'(t) > 0, y'(t) < 0$ e quindi, le orbite si dirigono in basso a destra.
5. In un punto del segmento QT si ha $x'(t) = 0, y'(t) > 0$ e quindi, l'orbita ha tangente verticale, proviene dalla zona 1 ed entra nella zona 3.
6. In un punto del segmento TP si ha $x'(t) = 0, y'(t) < 0$ e quindi, l'orbita ha tangente verticale, proviene dalla zona 2 ed entra nella zona 4.
7. In un punto del segmento ST si ha $x'(t) < 0, y'(t) = 0$ e quindi, l'orbita ha tangente orizzontale, proviene dalla zona 2 ed entra nella zona 3.
8. In un punto del segmento TR si ha $x'(t) > 0, y'(t) = 0$ e quindi, l'orbita ha tangente orizzontale, proviene dalla zona 1 ed entra nella zona 4.

Vediamo ora qual é il comportamento qualitativo delle soluzioni. Supponiamo che il punto mobile $P(t) = (x(t), y(t))$ parta da una posizione iniziale $P_0 = (x_0, y_0)$ e vediamo come si comporta l'orbita di P_0 , cioè la traiettoria di $P(t)$ per $t > 0$. Ci sono diversi casi:

1. Se P_0 coincide con uno dei 4 punti di equilibrio O, P, S, T , il punto mobile $P(t)$ rimane fisso in P_0 .
2. Se P_0 sta sull'asse x , allora é chiaro che $y(t) = 0$ per ogni t e quindi $x(t)$ segue l'equazione differenziale logistica $x' = x(1 - x)$. L'orbita di P_0 é un tratto dell'asse x . Se P_0 é a sinistra di $P = (1, 0)$, allora $x(t)$ cresce e tende a 1 per $t \rightarrow +\infty$. Il punto mobile $P(t)$ si sposta verso destra e tende asintoticamente a P . Se P_0 é a destra di P , il punto mobile $P(t)$ si sposta verso sinistra e tende asintoticamente a P .
3. Analogamente, se P_0 sta sull'asse y é chiaro che $x(t) = 0$ per ogni t e quindi, l'orbita di P_0 é un tratto dell'asse y . Se P_0 sta sotto $S = (0, 1)$, allora il punto mobile $P(t)$ si sposta verso l'alto e tende a S per $t \rightarrow +\infty$. Se P_0 sta sopra S , il punto mobile $P(t)$ si sposta verso il basso e tende a S per $t \rightarrow +\infty$.
4. Se P_0 sta nella zona 1, $x(t)$ e $y(t)$ crescono entrambe al crescere di t . Si potrebbe pensare che il punto mobile $P(t)$ per $t \rightarrow +\infty$ tenda a un punto P_1 interno alla zona 1, ma questo non é possibile perché P_1 dovrebbe essere un punto di equilibrio e non ci sono punti di equilibrio all'interno della zona 1. Quindi $P(t)$ arriva sulla frontiera della zona 1 e può intersecarla in un punto del segmento QT (Q escluso), nel

punto di equilibrio T o in un punto del segmento TR (R escluso). Ci sarà una particolare orbita che tende asintoticamente a T ed è facile convincersi che tale orbita coincide col segmento OT della retta $y = x$. Infatti, ponendo $y = x$, entrambe le equazioni del sistema differenziale diventano $x' = x(1 - 3x)$; se $x(t)$ indica una soluzione di questa equazione differenziale logistica, allora $P(t) = (x(t), x(t))$ è una soluzione del sistema. Se ne conclude che se P_0 sta sul segmento OT , $P(t)$ si muove sulla retta $y = x$ e tende asintoticamente a T ; se P_0 sta sopra al segmento OT , la sua orbita interseca il segmento QT verticalmente e poi entra nella zona 3; se infine P_0 sta sotto al segmento OT , la sua orbita interseca orizzontalmente il segmento TR e poi entra nella zona 4.

5. Se P_0 sta nella zona 2, $x(t)$ e $y(t)$ decrescono entrambe al crescere di t . Il punto mobile $P(t)$ si muove in basso a sinistra e arriva sulla frontiera della zona 2, però non può intersecarla in un punto dell'asse x situato a destra di P o in un punto dell'asse y situato sopra S , perché le semirette $(1, +\infty)$ dell'asse x e dell'asse y sono 2 orbite e 2 orbite diverse non possono intersecarsi. Le uniche possibilità sono che l'orbita di P_0 tenda asintoticamente a uno dei 3 punti di equilibrio S, T, P oppure intersechi il segmento ST ed entri nella zona 3, oppure intersechi il segmento TP ed entri nella zona 4. È facile convincersi che se P_0 sta sulla semiretta $x > 1, y > 1$ della retta $y = x$, la sua orbita coincide col segmento P_0T e il punto mobile $P(t)$ tende asintoticamente a T . Se P_0 sta sotto alla retta $y = x$, in teoria $P(t)$ potrebbe tendere asintoticamente a T oppure a P , ma il disegno delle orbite col computer sembra suggerire che in questo caso, ciò non si verifica mai: $P(t)$ incontra verticalmente il segmento TP e poi entra nella zona 4. Infine, se P_0 sta sopra la retta $y = x$, $P(t)$ incontra orizzontalmente il segmento ST e poi entra nella zona 3.
6. Se P_0 sta nella zona 3, $x(t)$ è decrescente e $y(t)$ è crescente. Il punto mobile $P(t)$ si muove in alto a sinistra e arriva sulla frontiera della zona 3, che a priori potrebbe incontrare in un punto qualsiasi (escluso T). Ma in realtà è facile vedere che $P(t)$ deve per forza tendere asintoticamente al punto di equilibrio S . Infatti, l'orbita di P_0 non può intersecare il segmento QT perché altrimenti dovrebbe provenire dalla zona 1, non può intersecare il segmento ST perché altrimenti dovrebbe provenire dalla zona 2, non può intersecare il segmento QS in un punto diverso da S perché QS è un'orbita e 2 orbite diverse non possono intersecarsi.
7. Se P_0 sta nella zona 4, $x(t)$ è crescente e $y(t)$ è decrescente. Il punto mobile $P(t)$ si muove in basso a destra e arriva sulla frontiera della zona 4, che a priori potrebbe incontrare in un punto qualsiasi (escluso

T). Ma in realtà é facile vedere che $P(t)$ deve tendere per forza asintoticamente a al punto di equilibrio P . Infatti, l'orbita di P_0 non puó intersecare il segmento TR perché altrimenti dovrebbe provenire dalla zona 1, non puó intersecare il segmento TP perché altrimenti dovrebbe provenire dalla zona 2, non puó intersecare il segmento RP in un punto diverso da P , perché RP é un'orbita e 2 orbite diverse non possono intersecarsi.

L'aspetto complessivo delle orbite (che si possono facilmente tracciare col computer) é illustrato in fig.11.2 É notevole il fatto che in questa situazione, quasi certamente una delle 2 specie X, Y si estingue. Quale delle 2 dipende dalle condizioni iniziali; in questo caso particolare si vede facilmente che la specie che inizialmente, all'istante $t = 0$, é meno numerosa, é destinata all'estinzione. Infatti, se il punto iniziale P_0 si trova sopra la retta $y = x$, la sua orbita tende asintoticamente al punto di equilibrio $Q = (0, 1)$ e quindi, la specie X si estingue. Si dice che il bacino di attrazione del punto di equilibrio asintoticamente stabile Q é la zona del primo quadrante che sta sopra la retta $y = x$. Se P_0 sta sotto la retta $y = x$, la sua orbita tende asintoticamente a $R = (1, 0)$ e quindi, Y si estingue (il bacino di attrazione di R é la zona del primo quadrante che sta sotto la retta). Se P_0 si trova esattamente sulla retta $y = x$, in teoria la sua orbita tende asintoticamente al punto di equilibrio instabile T quindi, le 2 specie potrebbero coesistere, ma in pratica questo non si verifica mai perché delle variazioni casuali di pochi individui in piú o in meno possono deviare facilmente il punto mobile $P(t)$ dalla sua orbita, facendolo penetrare nel bacino di attrazione di uno dei 2 equilibri stabili Q, R . Perció, ciascuna delle 2 specie corre il rischio di estinguersi, con una probabilitá del 50%. ■

Esempio 11.2

$$\begin{aligned} x' &= x(1 - 2x - y) && \text{figura 11.3} \\ y' &= y(1 - x - 2y) \end{aligned}$$

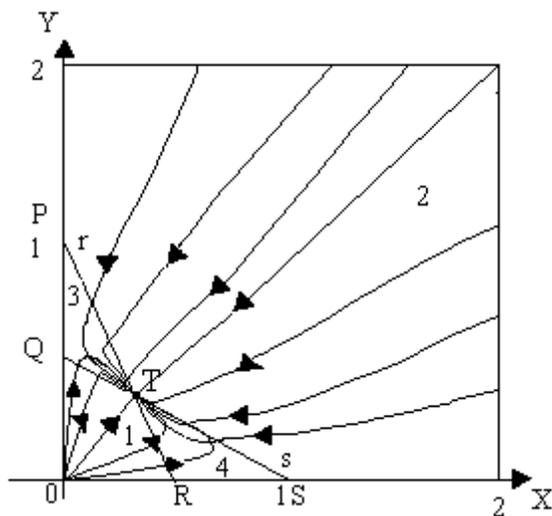


Figura 11.3

Le 2 rette r, s si incontrano nel punto $T = (1/3, 1/3)$. Ci sono 4 punti di equilibrio: $O = (0, 0), Q = (0, 1/2), R = (1/2, 0), T = (1/3, 1/3)$. La loro stabilità si può stabilire mediante la matrice jacobiana $J(x, y)$:

$$f(x, y) = x - 2x^2 - xy \quad g(x, y) = y - 2y^2 - xy$$

$$J(x, y) = \begin{pmatrix} 1 - 4x - y & -x \\ -y & 1 - x - 4y \end{pmatrix}$$

$$J(0, 0) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \rightarrow \text{gli autovalori sono } 1, 1$$

$$J(1/2, 0) = \begin{pmatrix} -1 & -1/2 \\ 0 & 1/2 \end{pmatrix} \rightarrow \text{gli autovalori sono } -1, 1/2$$

$$J(0, 1/2) = \begin{pmatrix} 1/2 & 0 \\ -1/2 & -1 \end{pmatrix} \rightarrow \text{gli autovalori sono } -1, 1/2$$

$$J(1/3, 1/3) = \begin{pmatrix} -2/3 & -1/3 \\ -1/3 & -2/3 \end{pmatrix} \rightarrow \text{gli autovalori sono } -1/3, -1$$

Quindi, T è un punto di equilibrio asintoticamente stabile, mentre Q, R, O sono instabili (O è un nodo instabile, Q ed R sono selle). Le 2 rette r, s dividono il primo quadrante in 4 zone 1,2,3,4 (fig. 11.3), in ciascuna delle quali $x'(t) = f(x(t), y(t))$ e $y'(t) = g(x(t), y(t))$ hanno segno costante, quindi si può studiare il comportamento qualitativo delle orbite. Precisamente:

1. Nella zona 1 si ha $x'(t) > 0, y'(t) > 0$ e quindi, $x(t), y(t)$ sono funzioni crescenti di t e l'orbita di un punto qualsiasi si dirige in alto a destra.

2. Nella zona 2 si ha $x'(t) < 0, y'(t) < 0$ e quindi, le orbite si dirigono in basso a sinistra.
3. Nella zona 3 si ha $x'(t) > 0, y'(t) < 0$ e quindi, le orbite si dirigono in basso a destra.
4. Nella zona 4 si ha $x'(t) < 0, y'(t) > 0$ e quindi, le orbite si dirigono in alto a sinistra.

L'aspetto complessivo delle orbite é illustrato in fig.11.3 É notevole il fatto che, in questa situazione, qualunque sia il punto di partenza $P_0 = (x(0), y(0))$, il punto mobile $P(t)$ tende asintoticamente verso il punto di equilibrio stabile T e quindi, le 2 specie X, Y possono coesistere pacificamente (a meno che naturalmente, P_0 non si trovi sull'asse x o sull'asse y , perché se é presente una sola delle 2 specie, l'altra non può comparire per magia). Il bacino di attrazione di T é tutto il primo quadrante, assi esclusi. ■

Esempio 11.3

$$\begin{aligned} x' &= x(1 - 2x - 3y) && \text{figura 11.4} \\ y' &= y(1 - x - y) \end{aligned}$$

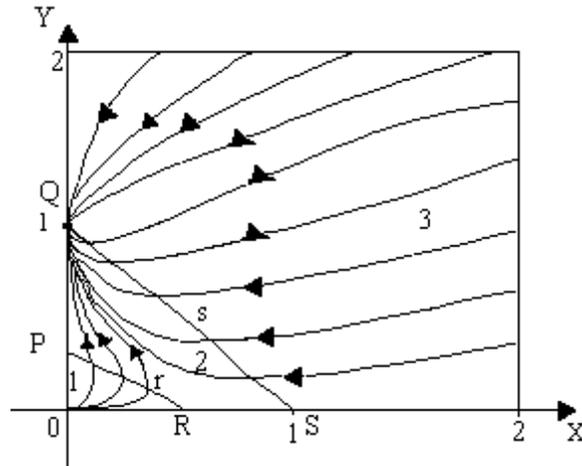


Figura 11.4

Le 2 rette r, s non si incontrano nel primo quadrante, ma r sta sotto s . Perciò ci sono solo 3 punti di equilibrio: $O = (0, 0), Q = (0, 1), R = (1/2, 0)$. La

loro stabilità si può stabilire mediante la matrice jacobiana $J(x, y)$:

$$f(x, y) = x - 2x^2 - 3xy \quad g(x, y) = y - y^2 - xy$$

$$J(x, y) = \begin{pmatrix} 1 - 4x - 3y & -3x \\ -y & 1 - x - 2y \end{pmatrix}$$

$$J(0, 0) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \rightarrow \text{gli autovalori sono } 1, 1$$

$$J(0, 1) = \begin{pmatrix} -2 & 0 \\ -1 & -1 \end{pmatrix} \rightarrow \text{gli autovalori sono } -2, -1$$

$$J(1/2, 0) = \begin{pmatrix} -1 & -3/2 \\ 0 & 1/2 \end{pmatrix} \rightarrow \text{gli autovalori sono } -1, 1/2$$

Quindi, Q è un punto di equilibrio asintoticamente stabile mentre R e O sono instabili (O è un nodo instabile, R è una sella). Le 2 rette r , s , dividono il primo quadrante in 3 zone 1,2,3 (fig. 11.4), in ciascuna delle quali $x'(t) = f(x(t), y(t))$ e $y'(t) = g(x(t), y(t))$ hanno segno costante, quindi si può studiare il comportamento qualitativo delle orbite. Precisamente:

1. Nella zona 1 si ha $x'(t) > 0, y'(t) > 0$ e quindi, $x(t), y(t)$ sono funzioni crescenti di t e l'orbita di un punto qualsiasi si dirige in alto a destra.
2. Nella zona 2 si ha $x'(t) < 0, y'(t) > 0$ e quindi, le orbite si dirigono in alto a sinistra.
3. Nella zona 3 si ha $x'(t) < 0, y'(t) < 0$ e quindi, le orbite si dirigono in basso a sinistra.

L'aspetto complessivo delle orbite è illustrato in fig.11.4. È notevole il fatto che, in questa situazione, qualunque sia il punto di partenza $P_0 = (x(0), y(0))$, il punto mobile $P(t)$ tende asintoticamente verso il punto di equilibrio stabile Q e quindi, la specie X è destinata ad estinguersi (a meno che, naturalmente, P_0 non si trovi sull'asse x , perché se è presente solo la specie X , la specie Y non può comparire per magia). Il bacino di attrazione di Q è tutto il primo quadrante, escluso l'asse x . ■

Esempio 11.4

$$\begin{aligned} x' &= x(1 - x - y) && \text{figura 11.5} \\ y' &= y(1 - 2x - 3y) \end{aligned}$$

Le 2 rette r, s non si incontrano nel primo quadrante, ma r sta sopra s .

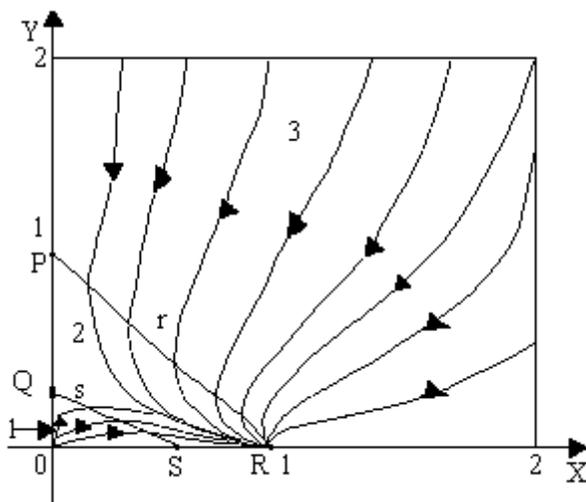


Figura 11.5

Peri3o ci sono solo 3 punti di equilibrio: $O = (0, 0)$, $Q = (0, 1/3)$, $R = (1, 0)$.
La loro stabilit3a si pu3o stabilire mediante la matrice jacobiana $J(x, y)$:

$$f(x, y) = x - x^2 - xy \quad g(x, y) = y - 3y^2 - 2xy$$

$$J(x, y) = \begin{pmatrix} 1 - 2x - y & -x \\ -2y & 1 - 2x - 6y \end{pmatrix}$$

$$J(0, 0) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \rightarrow \text{gli autovalori sono } 1, 1$$

$$J(1, 0) = \begin{pmatrix} -1 & -1 \\ 0 & -1 \end{pmatrix} \rightarrow \text{gli autovalori sono } -1, -1$$

$$J(0, 1/3) = \begin{pmatrix} 2/3 & 0 \\ -2/3 & -1 \end{pmatrix} \rightarrow \text{gli autovalori sono } -1, 2/3$$

Quindi, R 3 un punto di equilibrio asintoticamente stabile, mentre Q e O sono instabili (O 3 un nodo instabile, Q 3 una sella).

Le 2 rette r, s dividono il primo quadrante in 3 zone 1,2,3 (fig. 11.5), in ciascuna delle quali $x'(t) = f(x(t), y(t))$ e $y'(t) = g(x(t), y(t))$ hanno segno costante, quindi si pu3o studiare il segno qualitativo delle orbite. Precisamente:

1. Nella zona 1 si ha $x'(t) > 0$, $y'(t) > 0$ e quindi, $x(t), y(t)$ sono funzioni crescenti di t e l'orbita di un punto qualsiasi si dirige in alto a destra.
2. Nella zona 2 si ha $x'(t) > 0$, $y'(t) < 0$ e quindi, le orbite si dirigono in basso a destra.

3. Nella zona 3 si ha $x'(t) < 0$, $y'(t) < 0$ e quindi, le orbite si dirigono in basso a sinistra.

L'aspetto complessivo delle orbite é illustrato in fig.11.5. É notevole il fatto che, in questa situazione, qualunque sia il punto di partenza $P_0 = (x(0), y(0))$, il punto mobile $P(t)$ tende asintoticamente verso il punto di equilibrio stabile R e quindi, la specie Y é destinata ad estinguersi (a meno che naturalmente, P_0 non si trovi sull'asse y perché se é presente solo la specie Y , la specie X non può comparire per magia). Il bacino di attrazione di R é tutto il primo quadrante, escluso l'asse y . ■

12 Diffusione di epidemie

Consideriamo il problema della diffusione di una malattia infettiva in una popolazione chiusa, cioè nella quale non possono entrare individui, né possono uscirne (la cosa non é molto realistica, ma si tratta di un'ipotesi semplificativa per analizzare il problema). Supponiamo che la malattia conferisca immunità e quindi una persona non possa ammalarsi 2 volte (perché guarisce e diventa immune, oppure nella peggiore delle ipotesi muore). Supponiamo anche che la malattia abbia un tempo di incubazione trascurabile.

Sia N il numero totale di individui della popolazione. Ad ogni istante t di tempo indichiamo con

- $x(t)$ il numero di individui sani che non hanno ancora avuto la malattia e quindi possono essere contagiati
- $y(t)$ il numero di individui malati (e quindi contagiosi)
- $z(t)$ il numero di individui che hanno già avuto la malattia e quindi non possono più ammalarsi

É chiaro che si ha $x(t) + y(t) + z(t) = N$ per ogni valore di t . Supponiamo che all'istante iniziale $t = 0$ si abbia un certo numero y_0 (di solito molto piccolo) di malati e quindi sia $x_0 = N - y_0, y(0) = y_0, z(0) = 0$. Vogliamo vedere come variano nel tempo $x(t), y(t), z(t)$ cioè come si diffonde l'epidemia. Poiché $z(t) = N - x(t) - y(t)$, consideriamo come funzioni incognite solo $x(t)$ e $y(t)$. Un sistema di 2 equazioni differenziali per $x(t), y(t)$ é il seguente:

$$\begin{cases} x'(t) = -Ax(t)y(t) \\ y'(t) = Ax(t)y(t) - By(t) \end{cases} \quad (1)$$

Dove A, B sono 2 costanti positive dette rispettivamente *tasso di infezione* e *tasso di rimozione* (cioé tasso di cessazione della malattia, senza distinguere tra guarigione e morte).

Per dedurre il sistema (1) consideriamo un intervallo di tempo $I = (t, t + \Delta t)$ molto piccolo ed indichiamo con n, n' il numero di individui che si ammalano e il numero di individui che guariscono (o muoiono) durante l'intervallo I .

Si ha allora

$$x(t + \Delta t) = x(t) - n \qquad y(t + \Delta t) = y(t) + n - n'$$

ma n chiaramente é proporzionale sia al numero di individui sani $x(t)$, che al numero di malati $y(t)$, che al tempo Δt , cioè $n = Ax(t)y(t)\Delta t$, dove il tasso di infezione A ha il significato di "probabilità che un individuo sano incontri un individuo malato e ne sia contagiato, nell'unità di tempo". Inoltre é chiaro che $n' = By(t)\Delta t$, dove il tasso di rimozione B ha il significato di "probabilità che un individuo malato esca dalla malattia (guarisca o

muoia) nell'unit  di tempo". Si ha allora:

$$x(t + \Delta t) = x(t) - Ax(t)y(t)\Delta t \quad y(t + \Delta t) = y(t) + Ax(t)y(t)\Delta t - By(t)\Delta t$$

$$\frac{x(t + \Delta t) - x(t)}{\Delta t} = -Ax(t)y(t) \quad \frac{y(t + \Delta t) - y(t)}{\Delta t} = Ax(t)y(t) - By(t)$$

e passando al limite per $\Delta t \rightarrow 0$ si ottiene il sistema (1) che riscriviamo brevemente nella forma

$$\begin{cases} x' = f(x, y) = -Axy & x(0) = x_0 = N - y_0, y(0) = y_0 \\ y' = g(x, y) = Axy - By & A, B > 0 \end{cases} \quad (2)$$

Si tratta di un sistema differenziale autonomo, le cui soluzioni $(x(t), y(t))$ si possono rappresentare nel piano cartesiano x, y , studiando l'orbita del punto iniziale $P_0 = (x_0, y_0)$, cio  la traiettoria del punto mobile $P(t) = (x(t), y(t))$ a partire da P_0 . Si possono anche cercare i punti di equilibrio, ma il loro studio in questo caso particolare non dice molto. Infatti i punti di equilibrio si trovano uguagliando a zero $f(x, y)$ e $g(x, y)$:

$$Axy = 0 \quad Axy - By = 0 \quad \text{da cui} \quad y = 0$$

Se $y = 0$, qualunque valore di x soddisfa queste 2 equazioni, quindi i punti di equilibrio sono tutti e soli i punti dell'asse x . Ma questo era gi  ovvio a priori, perch  se $y_0 = 0$ non ci sono malati e quindi non ci pu  essere nessuna epidemia, mentre se $y_0 > 0$ qualche individuo sar  contagiato, quindi $x(t)$ diminuir  al crescere di t e P_0 non pu  essere un punto di equilibrio.

Pi  interessante   lo studio qualitativo delle orbite, che si pu  fare studiando il segno di $x'(t), y'(t)$. Si vede subito che $x'(t) = Axy$   positiva in tutto il primo quadrante (e vale 0 sugli assi) mentre $y'(t) = Axy - By = y(Ax - B)$   positiva se $x > \frac{B}{A}$,   negativa se $x < \frac{B}{A}$ e vale 0 se $x = \frac{B}{A}$. Quindi la retta verticale r di equazione $x = \frac{B}{A}$ divide il primo quadrante in due zone 1 e 2 (figura 12.1). Nella zona 1 si ha $x'(t) < 0, y'(t) < 0$ e quindi le orbite si dirigono in basso a sinistra. Nella zona 2 si ha $x'(t) < 0, y'(t) > 0$ e quindi le orbite si dirigono in alto a sinistra. Nei punti della retta r si ha $x'(t) < 0, y'(t) = 0$ e quindi le orbite hanno tangente orizzontale e si dirigono verso sinistra.

I punti dell'asse x sono tutti punti di equilibrio, mentre se il punto iniziale P_0 sta sull'asse y , cio  significa che per $t = 0$ sono tutti malati e quindi $x(t) = 0$ per ogni t , quindi il sistema differenziale si riduce all'unica equazione differenziale $y'(t) = -By(t)$, che   l'equazione del decadimento esponenziale, la cui soluzione   $y(t) = y_0 e^{-Bt}$. Il numero di malati $y(t)$ decresce esponenzialmente e tende a zero per $t \rightarrow +\infty$. Il periodo di dimezzamento di questo decadimento si trova imponendo che $y(t) = y_0/2$ e vale $\tau = \log 2/B$.

Il significato di τ é quello di “durata media della malattia”, nel senso che all’istante $t = \tau$, metà dei malati iniziali sono guariti (o morti).

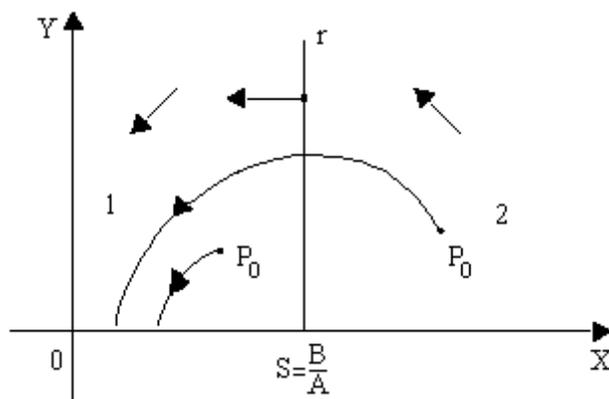


Figura 12.1

Se il punto iniziale P_0 si trova nella zona 1, la sua orbita si dirige in basso a sinistra e raggiunge la frontiera della zona 1, però non può intersecare l’asse y se non nell’origine (perché l’asse y é un’orbita e due orbite diverse non possono intersecarsi). Quindi l’orbita deve dirigersi asintoticamente verso un punto dell’asse x (vedi figura 12.1).

se P_0 sta nella zona 2, la sua orbita si dirige in alto a sinistra, ma $y(t)$ non può crescere illimitatamente (perché $y(t) \leq N$) mentre $x(t)$ deve decrescere fino a raggiungere il valore $\frac{B}{A}$ (perché nella zona 2 non ci sono punti di equilibrio). Quindi l’orbita taglia orizzontalmente la retta r , entra nella zona 1 e poi si dirige asintoticamente verso un punto dell’asse x (figura 11.1)

Osserviamo che tutte le orbite (escluso l’asse y) si possono esprimere anche come il grafico di una funzione $y = y(x)$, perché lungo un’orbita $x(t)$ decresce sempre e quindi non ci sono 2 punti diversi dell’orbita aventi la stessa ascissa. La funzione $y(x)$ si può anche trovare esplicitamente, dividendo la seconda equazione del sistema (2) per la prima:

$$\frac{dy}{dx} = \frac{\left(\frac{dy}{dt}\right)}{\left(\frac{dx}{dt}\right)} = \frac{x'(t)}{y'(t)} = \frac{(Axy - By)}{(-Axy)} = \frac{B}{Ax} - 1$$

questa é un’equazione differenziale molto semplice per la funzione $y(x)$, che si risolve semplicemente integrando ambo i membri:

$$y(x) = \int \left(\frac{B}{Ax - 1}\right) dx + C = \frac{B \log(x)}{A} - x + C \quad (3)$$

La costante C si determina ponendo $t = 0, y = y_0, x = x_0$:

$$C = y_0 - \frac{B \log(x_0)}{A} + x_0 \quad \text{da cui}$$

$$y(t) = y_0 + \frac{B}{A} \log\left(\frac{x}{x_0}\right) + x_0 - x$$

Dalla (2) si deduce che tutte le orbite si ottengono dalla curva $y = \frac{B \log(x)}{A} - x$ traslandola verticalmente verso l'alto o verso il basso. È facile studiare la funzione $y(x)$ e si trova che per $x \rightarrow 0$ oppure per $x \rightarrow +\infty$, $y(x)$ tende a $-\infty$; inoltre $y'(x) = \frac{B}{Ax} - 1 > 0$ se $x < \frac{B}{A}$, $y'(x) < 0$ se $x > \frac{B}{A}$, $y'(x) = 0$ se $x = \frac{B}{A}$. Quindi $y(x)$ è crescente se $x < \frac{B}{A}$, è decrescente se $x > \frac{B}{A}$, ha un massimo se $x = \frac{B}{A}$. Ne deriva che le orbite sono del tipo di figura 12.2.

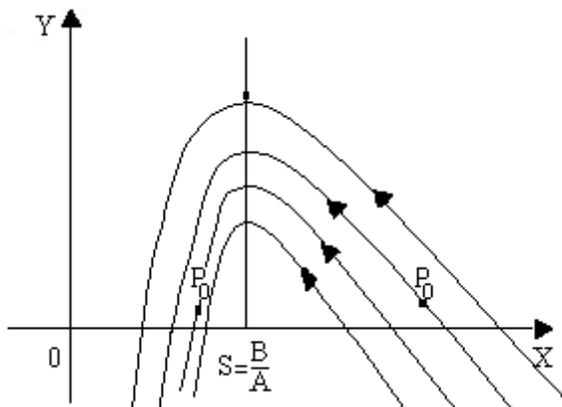


Figura 12.2

Se noi consideriamo un punto iniziale $P_0 = (x_0, y_0)$ con y_0 molto piccolo, vediamo che il comportamento dell'epidemia è molto diverso a seconda che x_0 superi o no il valore $\frac{B}{A}$, che indichiamo con S e chiamiamo *valore soglia*.

Se $x_0 < S$, $y(t)$ decresce e l'epidemia si esaurisce in breve tempo. Se invece $x_0 > S$, $y(t)$ inizialmente cresce, raggiunge un valore massimo $y_1 = y_0 + S \log\left(\frac{S}{x_0}\right) + x_0 - S$ quando $x(t) = S$ e poi decresce. Per $t \rightarrow +\infty$, $y(t)$ tende a zero e $x(t)$ tende ad un certo valore x_2 , che rappresenta il numero di individui che riescono a sfuggire al contagio e non si ammalano mai. Il numero totale di individui che vengono contagiati è $x_0 - x_2$, mentre y_1 rappresenta il numero massimo di individui malati nello stesso istante.

Si può dire che l'epidemia esplode solo se $x_0 > S$, non esplode se $x_0 < S$. Cioè l'effetto dei pochi malati iniziali è molto più grave in una popolazione grande che in una popolazione piccola.

Se x_0 supera di poco S , y_1 è di poco superiore a y_0 e x_2 è di poco inferiore a S , ma se x_0 supera di molto S , allora y_1 è molto grande e x_2 è molto piccolo, vicino a zero. Però x_2 non è mai zero, il che significa che anche nel peggiore dei casi, qualcuno riesce sempre a sfuggire al contagio.

L'effettivo andamento temporale dell'epidemia (dato dalle funzioni $x(t), y(t)$, non si può determinare esplicitamente (mediante formule), ma solo approssimativamente con metodi numerici, mediante il computer.

Osserviamo che l'equazione della generica orbita, cioè la (2), si può scrivere semplicemente così:

$$y = S \log(x) - x + C$$

e quindi non dipende separatamente dai 2 parametri A, B , ma solo dal valore soglia $S = \frac{B}{A}$.

Dividendo ambo i membri per S si trova:

$$\frac{y}{S} = \log(x) - \frac{x}{S} + \frac{C}{S} = \log\left(\frac{x}{S}\right) - \frac{x}{S} + \log(S) + \frac{C}{S}$$

Usando al posto delle 2 variabili $x(t), y(t)$ le 2 nuove variabili $u(t) = \frac{x(t)}{S}, v(t) = \frac{y(t)}{S}$ (il che equivale ad esprimere le numerosità delle popolazioni usando come unità di misura il valore soglia S), l'equazione della generica orbita diventa:

$$v = \log(u) - u + C_1 \quad (\text{dove } C_1 = \log(S) + \frac{C}{S})$$

il che significa che, riportando $u(t)$ in ascisse e $v(t)$, le orbite non dipendono da nessun parametro (tranne che, ovviamente, dal punto iniziale P_0) e quindi il loro andamento è esattamente uguale per tutte le epidemie.

Il sistema differenziale (1), dividendo ambo i membri per $S = \frac{B}{A}$, si può riscrivere così:

$$\frac{x'}{S} = -\frac{Axy}{S} = -AS\left(\frac{x}{S}\right)\left(\frac{y}{S}\right) = -B\left(\frac{x}{S}\right)\left(\frac{y}{S}\right)$$

$$\frac{y'}{S} = \frac{Axy}{S} - \frac{By}{S} = B\left(\frac{x}{S}\right)\left(\frac{y}{S}\right) - \frac{By}{S}$$

cioè

$$u' = -Buv$$

$$v' = Buv - Bv$$

Ciò significa che usando le variabili $u(t), v(t)$, il vettore velocità $(u'(t), v'(t))$ del punto mobile $(u(t), v(t))$ lungo l'orbita è proporzionale a B . Raddoppiando B , raddoppiano $u'(t), v'(t)$ e quindi la velocità di diffusione dell'epidemia, quindi si dimezza la durata T dell'epidemia, cioè il tempo che il punto mobile impiega a percorrere l'orbita, da un punto iniziale $P_0 = (u_0, v_0)$ ad un punto finale $P_2 = (u_2, v_2)$ scelti in modo opportuno. D'altronde si dimezza anche la durata media $\tau = \frac{\log(2)}{B}$ della malattia, quindi il rapporto T/τ (durata dell'epidemia /durata della malattia) rimane invariato. Ne segue che T/τ non dipende da A , né da B , ma è lo stesso per tutte le epidemie (dipende però, naturalmente, da P_0 e P_2). Per definire la "durata media" dell'epidemia" T prenderò arbitrariamente v_0 molto piccolo (per esempio $v_0 = .01$) e $P_2 =$ punto dell'orbita in cui $v(t) = \frac{v_0}{2}$. In tal modo T/τ dipende solo dall'ascissa u_0 del punto iniziale P_0 .

La tabella 12.1 riporta nella prima colonna i valori di u_0 (cioè della popolazione inizialmente a rischio di contagio, posto uguale a 1 il valore soglia S). Facciamo variare u_0 da 0.1 a 5.0 con passo 0.1. Nella seconda colonna riportiamo i valori dell'ordinata u_1 del punto più alto dell'orbita ($u_1 =$ numero massimo di malati nello stesso istante). Nella terza colonna riportiamo i valori di $1 - \frac{u_2}{u_0}$ (= percentuale di individui contagiati alla fine dell'epidemia). Nella quarta colonna riportiamo i valori di T/τ (rapporto tra durata dell'epidemia e durata della malattia).

La figura 12.3 illustra la variazione del tempo di $v(t) =$ rapporto tra numero di malati e valore soglia S , riportando sull'asse delle ascisse T/τ , cioè il tempo misurato prendendo come unità τ (durata media della malattia). Sono disegnate 8 curve (una per ciascuno dei seguenti 8 valori di $u_0 = \frac{x_0}{S}$: 0.1 - 0.5 - 1 -1,5 - 2 - 3 -4 -5. Queste curve danno un'idea della rapidità di diffusione dell'epidemia nel tempo.

Se $u_0 < 1$, l'epidemia si esaurisce rapidamente. Se $u_0 = 1$ (valore soglia), l'epidemia non coinvolge molte persone, ma si diffonde lentamente e dura molto a lungo ($T/\tau = 18.324$). Se $u_0 = 1.5$, l'epidemia dura ancora più a lungo ($T/\tau = 26.082$) e nella fase iniziale si diffonde in modo esplosivo. Se u_0 cresce ancora (assumendo i valori 2, 3, 4, 5 e oltre) l'epidemia all'inizio si diffonde in modo sempre più rapido, contagiando un numero sempre maggiore di persone, ma la sua durata totale tende a diminuire. Queste curve sono state confrontate in vari casi con dati epidemiologici reali, ottenendo un discreto accordo, nonostante le ipotesi del modello semplifichino molto la realtà.

La figura 12.4 (dedotta direttamente dalla tabella 1) illustra il grafico di $1 - \frac{u_2}{u_0}$ (percentuale di individui contagiati alla fine dell'epidemia) in funzione di u_0 (popolazione a rischio iniziale). Se $u_0 = 0.1$, solo l'1.1% delle persone a rischio si ammala; se $u_0 = 1$, si ammala il 13.4%; se $u_0 = 5$, si ammala il 99,3%. Quando u_0 cresce da 0.5 a 2, la percentuale di persone

contagiate si impenna rapidamente dall'1.9% all' 80%, mentre se $u_0 < 0.5$ é molto piccola e se $u_0 > 2$ é molto vicina al 100%.

La figura 12.5 (anch'essa dedotta direttamente dalla tabella 1) illustra il grafici di T/τ (rapporto tra la durata dell'epidemia e la durata della malattia) in funzione di u_0 . Se $u_0 < 0.5$ la durata T dell'epidemia é molto breve ($< 2\tau$), poi si impenna se u_0 cresce da 0.5 a 1.2, raggiungendo un massimo di circa 32τ . Se u_0 cresce ancora, T comincia a diminuire, dapprima rapidamente poi piú lentamente. Se $u_0 = 5$, T vale circa 12.7τ .

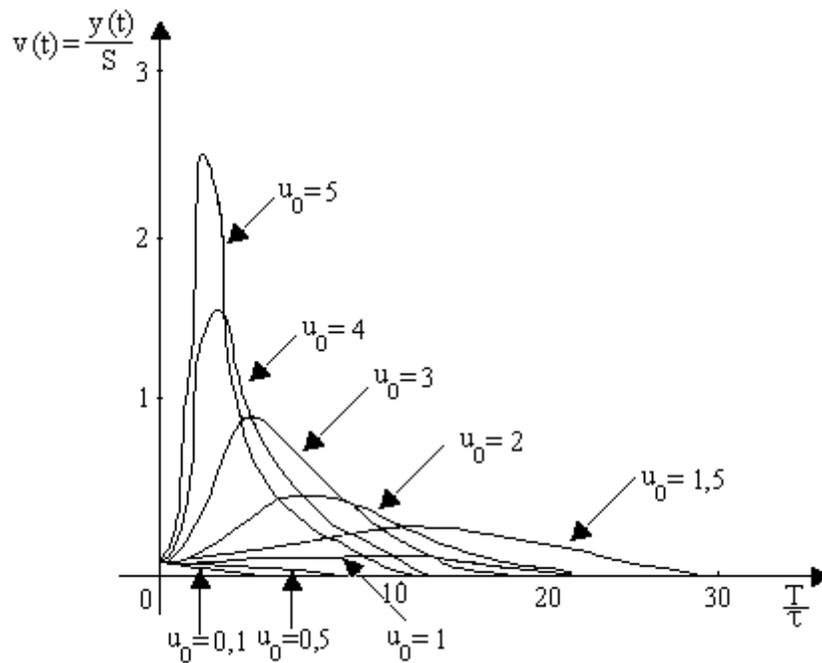
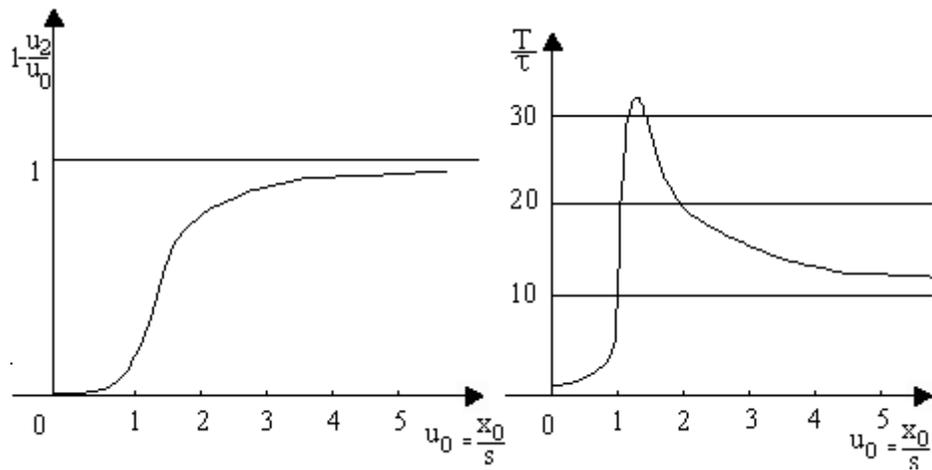


Figura 12.3



u_0	u_1	$1-u_2/u_0$	T/τ	u_0	u_1	$1-u_2/u_0$	T/τ
0.1	0.010	0.011	1.111	2.6	0.654	0.906	16.343
0.2	0.010	0.012	1.249	2.7	0.717	0.917	15.982
0.3	0.010	0.014	1.427	2.8	0.780	0.926	15.656
0.4	0.010	0.016	1.662	2.9	0.845	0.934	15.363
0.5	0.010	0.019	1.989	3.0	0.911	0.941	15.097
0.6	0.010	0.024	2.474	3.1	0.979	0.948	14.856
0.7	0.010	0.031	3.265	3.2	1.047	0.953	14.637
0.8	0.010	0.044	4.749	3.3	1.116	0.958	14.436
0.9	0.010	0.072	8.278	3.4	1.186	0.963	14.253
1.0	0.010	0.134	18.324	3.5	1.257	0.966	14.085
1.1	0.015	0.238	30.206	3.6	1.329	0.970	13.932
1.2	0.028	0.347	30.227	3.7	1.402	0.973	13.791
1.3	0.048	0.444	30.437	3.8	1.475	0.976	13.661
1.4	0.074	0.526	28.148	3.9	1.549	0.978	13.541
1.5	0.105	0.594	26.082	4.0	1.624	0.980	13.430
1.6	0.140	0.650	24.336	4.1	1.699	0.982	13.328
1.7	0.179	0.698	22.875	4.2	1.775	0.984	13.234
1.8	0.222	0.737	21.648	4.3	1.851	0.986	13.146
1.9	0.268	0.771	20.611	4.4	1.928	0.987	13.065
2.0	0.317	0.800	19.723	4.5	2.006	0.988	12.991
2.1	0.368	0.825	18.959	4.6	2.084	0.990	12.922
2.2	0.422	0.846	18.296	4.7	2.162	0.991	12.857
2.3	0.477	0.864	17.714	4.8	2.241	0.992	12.798
2.4	0.535	0.880	17.202	4.9	2.321	0.992	12.743
2.5	0.594	0.894	16.748	5.0	2.401	0.993	12.691

Tabella 1

Tabella 12.1

Figura 12.5

13 Equazioni alle differenze e transizione al caos

In una generica equazione differenziale del primo ordine in forma normale nella funzione incognita $x(t)$:

$$x'(t) = F(t, x(t)) \tag{1}$$

sia la variabile indipendente t che la variabile dipendente x sono variabili continue che variano in \mathbb{R}_+ in un opportuno intervallo contenuto in \mathbb{R} .

In alcuni casi però, quando si cerca di fare un modello matematico di un fenomeno naturale, può essere opportuno discretizzare la variabile dipendente x (vedi paragrafo seguente: processi stocastici) oppure la variabile indipendente t (in tal caso si ottiene un'equazione alle differenze, che definirò tra poco). Le equazioni alle differenze sono importanti anche perché nella maggior parte dei casi le equazioni differenziali non si sanno risolvere analiticamente (con una formula), ma solo con metodi numerici approssimati (col computer), i quali in realtà risolvono un'opportuna equazione alle differenze associata all'equazione differenziale data.

Un esempio tipico di fenomeno in cui conviene considerare discretizzata la variabile indipendente (il tempo t) potrebbe essere la crescita di popolazioni con generazioni che non si sovrappongono, come ad esempio certi insetti che depongono uova che si schiederanno l'anno dopo, o certe piante che fanno semi che germoglieranno l'anno dopo, o anche certi vertebrati la cui riproduzione presenta un andamento stagionale.

In questo esempio conviene discretizzare il tempo in anni, facendogli assumere solo i valori interi $0, 1, 2, 3, \dots, n, \dots$ (anni) e indicando con x_n la numerosità della popolazione al tempo n .

Supponendo che il tasso di crescita da un anno al successivo sia una certa costante $a > 0$, (a = numero medio di discendenti generati da un individuo ogni anno) si ha allora il modello semplicissimo di *crescita esponenziale discreta*:

$$x_{n+1} = x_n + ax_n = (1 + a)x_n \tag{2}$$

Se invece supponiamo che ci sia un valore di saturazione M , corrispondente all'esaurimento delle risorse, e che il tasso di crescita sia proporzionale alla percentuale di risorse non utilizzate ($1 - \frac{x_n}{M}$), si ha il seguente modello di *crescita logistica discreta*:

$$x_{n+1} = x_n + ax_n \left(1 - \frac{x_n}{M}\right) = x_n + ax_n - \frac{ax_n^2}{M}$$

e ponendo $\frac{a}{M} = b$:

$$x_{n+1} = x_n + ax_n - bx_n^2 \tag{3}$$

La (2) e la (3) sono esempi particolari di *equazioni alle differenze* nella funzione incognita x_n della variabile discreta n . La piú generica *equazione alle differenze del primo ordine in forma normale* é un'equazione del tipo :

$$x_{n+1} = G(n, x_n) \quad (4)$$

Il nome *equazione alle differenze* deriva dal fatto che esse esprime la differenza $x_{n+1} - x_n$ (incremento della funzione x_n tra 2 valori successivi della variabile indipendente) mediante una funzione $H(n, x_n) = G(n, x_n) - x_n$ delle 2 variabili n, x .

Se G non dipende esplicitamente da n , ma solo da x_n , l'equazione alle differenze si dice *autonoma*. Le equazioni della crescita esponenziale discreta e della crescita logistica discreta sono esempi di equazioni alle differenze autonome.

É ovvio che la generica equazione alle differenze (4) si può risolvere in modo banale, e ammette una e una sola soluzione una volta fissato il valore iniziale x_0 che x_n assume per $n = 0$. Infatti, noto x_0 , si possono calcolare successivamente $x_1 = G(0, x_0), x_2 = G(1, x_1), \dots x_k = G(k-1, x_{k-1}) \dots$ per ogni valore intero di k .

Si può anche generalizzare la definizione di equazione alle differenze e considerare una generica *equazione alle differenze di ordine p in forma normale*, che é del tipo :

$$x_n = G(n, x_{n-1}, x_{n-2} \dots x_{n-p})$$

In questa equazione il valore di x nel punto n dipende da n e dai valori di x nei p punti immediatamente precedenti.

É ovvio che esiste ed é unica la soluzione di quest'equazione se noi fissiamo p condizioni iniziali (i valori $x_0, x_1 \dots x_{p-1}$).

In alcuni casi la soluzione di un'equazione alle differenze si sa anche esprimere in forma esplicita, con una formula che dá direttamente x_n in funzione di n . É il caso ad esempio della crescita esponenziale discreta, espressa dall'equazione:

$$x_{n+1} = (a + 1)x_n \quad (5)$$

Essa si risolve considerando le n uguaglianze :

$$x_1 = (a + 1)x_0$$

$$x_2 = (a + 1)x_1$$

...

$$x_{n-1} = (a + 1)x_{n-2}$$

$$x_n = (a + 1)x_{n-1}$$

moltiplicandole membro a membro e semplificando x_1, x_2, \dots, x_{n-1} si ottiene la soluzione:

$$x_n = (a + 1)^n x_0$$

al crescere di n , x_n cresce esponenzialmente e tende a $+\infty$ per $n \rightarrow +\infty$. Il comportamento della soluzione é molto simile al caso dell'equazione differenziale della crescita esponenziale ($x'(t) = ax(t)$, $x(0) = x_0$) che ha come soluzione $x(t) = x_0 e^{at}$.

Diverso é il caso della crescita logistica discreta, espressa dall'equazione:

$$x_{n+1} = x_n + ax_n - bx_n^2$$

essa non si può risolvere in forma esplicita e la soluzione dá luogo a dei fenomeni strani, del tutto diversi da quel che succede nel caso della crescita logistica continua, espressa dall'equazione differenziale $x'(t) = ax(t) - bx^2(t)$. Per semplicitá, consideriamo il caso particolare $a = b$. Ció equivale a porre il livello di saturazione $M = 1$ (il che non é restrittivo perché equivale a cambiare l'unitá di misura sull'asse x) e allora l'equazione della logistica discreta diventa:

$$x_{n+1} = (a + 1)x_n - ax_n^2 \tag{6}$$

Fissato un valore iniziale $x_0 \in (0, 1)$, cioè positivo e minore del valore di saturazione, la (5) può essere risolta iterativamente, ma la soluzione presenta andamenti molto diversi al variare del parametro $a > 0$, che rappresenta il tasso di crescita quando x_n é molto minore di 1, cioè quando siamo lontani dalla saturazione. Si possono avere vari casi:

- CASO 1: $a \in [0, 1]$, *crescita monotóna*. In questo caso la soluzione x_n é sempre crescente e per $n \rightarrow \infty$ tende al valore di saturazione 1. Si ha cioè un comportamento molto simile a quello della soluzione dell'equazione differenziale $x'(t) = ax(t)(1 - x(t))$ della logistica continua. Questo risultato si può dimostrare abbastanza facilmente e può anche essere illustrato graficamente (fig.13.1).

Consideriamo la funzione

$$y = \varphi(x) = (a + 1)x - ax^2$$

il cui grafico é una parabola γ passante per i 2 punti $(0, 0)$ e $(1, 1)$, avente concavitá verso il basso, asse parallelo all'asse y e vertice nel

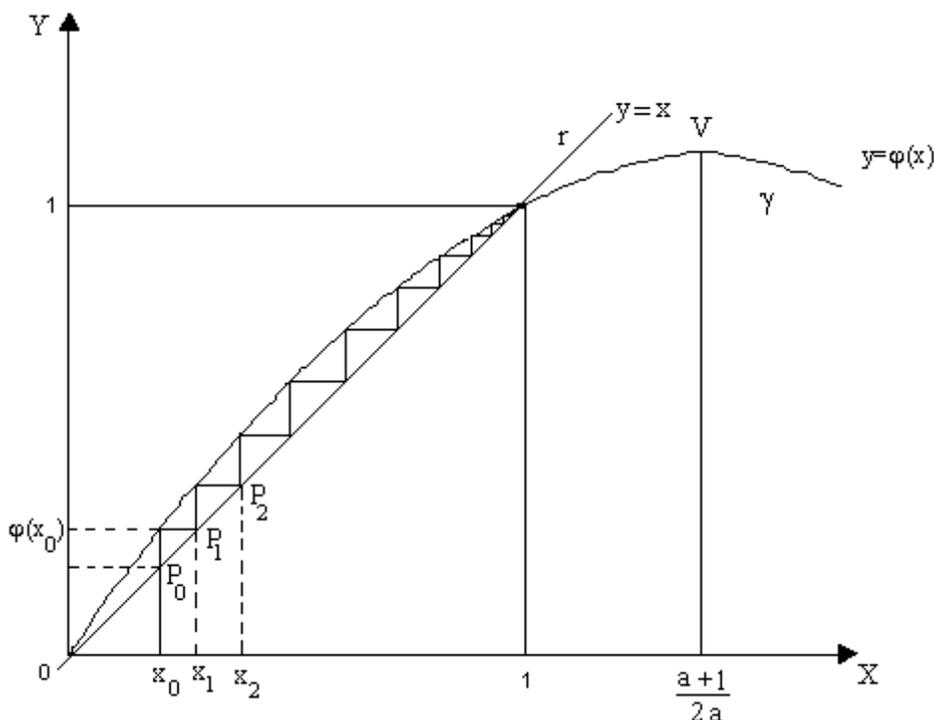


Figura 13.1

punto $V = \left(\frac{a+1}{2a}, \frac{(a+1)^2}{4a} \right)$. L'equazione alle differenze (5) si può allora, scrivere così

$$x_{n+1} = \varphi(x_n)$$

Se noi consideriamo il punto P_0 di coordinate (x_0, x_0) , tracciamo da P_0 la parallela all'asse y e dal punto in cui essa incontra la parabola γ tracciamo la parallela all'asse x e chiamiamo P_1 il punto in cui essa incontra la retta $y = x$ (fig.13.1), è chiaro che P_1 ha coordinate $(x_1, x_1) = (\varphi(x_0), \varphi(x_0))$. In modo analogo si può costruire $P_2 = (x_2, x_2)$ a partire da P_1 , e in generale, $P_{n+1} = (x_{n+1}, x_{n+1})$ a partire da $P_n = (x_n, x_n)$.

Poiché nell'intervallo $[0, 1]$ la parabola γ e la retta $y = x$ sono entrambe crescenti e la parabola sta sempre sopra la retta, è chiaro dalla fig.13.1 (e si può dimostrare facilmente) che la successione $\{x_n\}$ è strettamente crescente (cioè $x_n < x_{n+1} \forall n \in \mathbb{N}$ e che $\lim_{n \rightarrow +\infty} x_n = 1$).

- CASO 2: $a \in [1, 2]$, *oscillazioni smorzate*. In questo caso la soluzione x_n tende ancora al valore di saturazione 1 per $n \rightarrow \infty$, ma compie delle oscillazioni smorzate intorno a 1. Per n abbastanza grande i valori di x_n sono alternativamente maggiori e minori di 1.

Un comportamento del genere non era possibile nel caso della logistica continua. Esso é illustrato graficamente dalla figura 13.2.

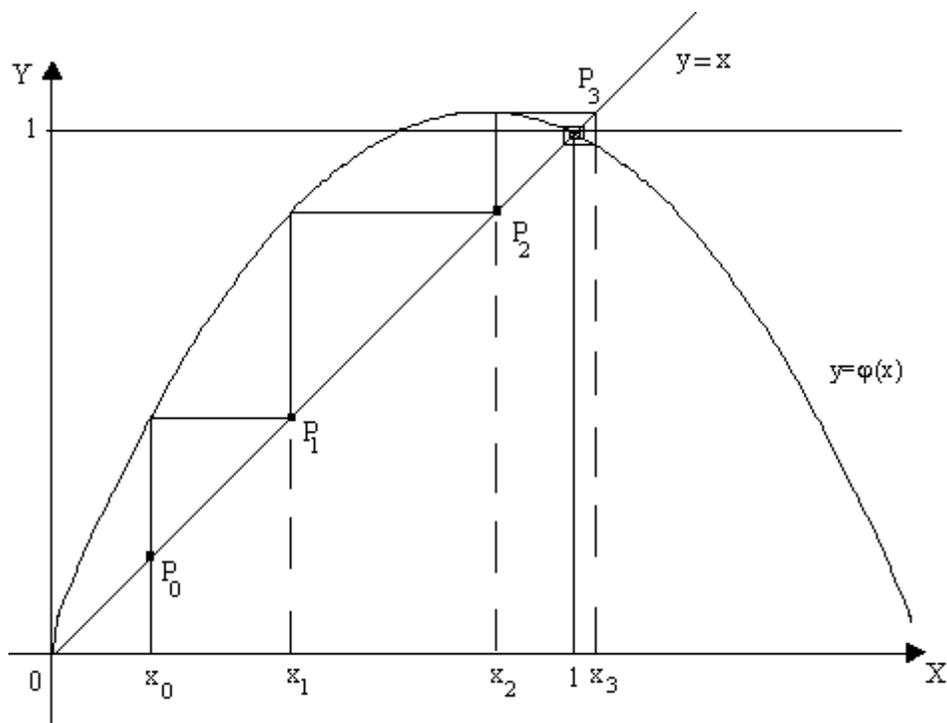
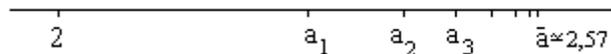


Figura 13.2

In questo caso, la parabola di equazione $y = \varphi(x) = (a + 1)x - ax^2$ ha il vertice nel punto di ascissa $\frac{a+1}{2a} < 1$ ed é decrescente nel punto $(1, 1)$, con pendenza minore di 45° (infatti, $\varphi'(1) = 1 - a \in [-1, 0]$). Allora, si può dimostrare facilmente che la successione di punti $P_0, P_1, P_2, P_3, \dots, P_n, \dots$ costruita nello stesso modo del caso precedente converge al punto $(1, 1)$ e la successione $x_0, x_1, x_2, \dots, x_n, \dots$ delle ascisse di questi punti converge a 1 oscillando (in fig. 13.2 $x_n < 1$ se n é pari ≥ 2 , $x_n > 1$ se n é dispari ≥ 3).

- CASO 3: $a \in [2, 2.57]$: *soluzione quasi periodica*. Quando il tasso di crescita a supera il valore 2, la situazione si complica rapidamente. Se a varia nell'intervallo $[2, 2.57]$ abbiamo una successione di valori critici $a_1, a_2, a_3, \dots, a_n, \dots$ che si addensano verso il valore limite $\bar{a} = \lim_{n \rightarrow \infty} a_n \simeq 2.57$:



(i primi valori sono $a_1 = 2.449, a_2 = 2.544, a_3 = 2.564$) e si ha che:

- Se $a \in [2, a_1]$ la soluzione é quasi periodica di periodo 2, cioè i termini pari (x_{2n}) tendono ad un valore l_1 ed i termini dispari (x_{2n+1}) tendono ad un valore diverso l_2 , perciò per n grande, la successione $\{x_n\}$ diventa indistinguibile da una successione periodica di periodo 2, che alterna i 2 valori l_1, l_2 .
- Se $a \in [a_1, a_2]$ la soluzione é quasi periodica di periodo 4 cioè per n grande diventa indistinguibile da una successione periodica di periodo 4.
- Se $a \in [a_2, a_3]$ la soluzione é quasi periodica di periodo 8 e in generale se $a \in [a_n, a_{n+1}]$ la soluzione é quasi periodica di periodo 2^{n+1} .

Per alcuni valori particolari del tasso di crescita a e del valore iniziale x_0 può darsi che la soluzione sia esattamente periodica, come mostra la fig. 13.3 nel caso del periodo 2.

Nella figura la parabola di equazione $y = \varphi(x) = (a + 1)x - ax^2$ é decrescente con pendenza maggiore di 45° nel punto $(1, 1)$ ed il quadrato AP_2BP_3 ha 2 vertici sulla parabola e gli altri 2 sulla retta $y = x$. Perció la successione $x_1, x_2, x_3, x_4 \dots, x_n, \dots$ costruita come nei casi precedenti ripete indefinitamente i 2 valori x_2, x_3 (cioé si ha $x_2 = x_4 = x_6 \dots = x_{2n}$ e $x_3 = x_5 = x_7 \dots = x_{2n+1}$ per ogni $n \geq 1$).

- CASO 4: $a > 2.57$, *comportamento caotico*. Se $a > \bar{a} = 2.57$ il comportamento della soluzione $\{x_n\}$ diventa praticamente imprevedibile: essa presenta delle oscillazioni molto irregolari di ampiezza variabile, che sembrano quasi una sequenza di numeri casuali.

Si tratta in realtà di numeri pseudo - casuali perché sono generati con la formula deterministica $x_{n+1} = \varphi(x_n)$. Inoltre, piccole variazioni del tasso di crescita a o del valore iniziale x_0 possono provocare, per n abbastanza grande, grandi variazioni di x_n . Questo comportamento é noto come *comportamento caotico* e si dice che il valore $\bar{a} \simeq 2.57$ segna la *transizione al caos*

Le figure 13.4 A,B,C,D,E,F (in cui é sempre $x_0 = 0.1$) mostrano un esempio di crescita monotóna ($a = 0.9$, fig. A), un esempio di oscillazione smorzata

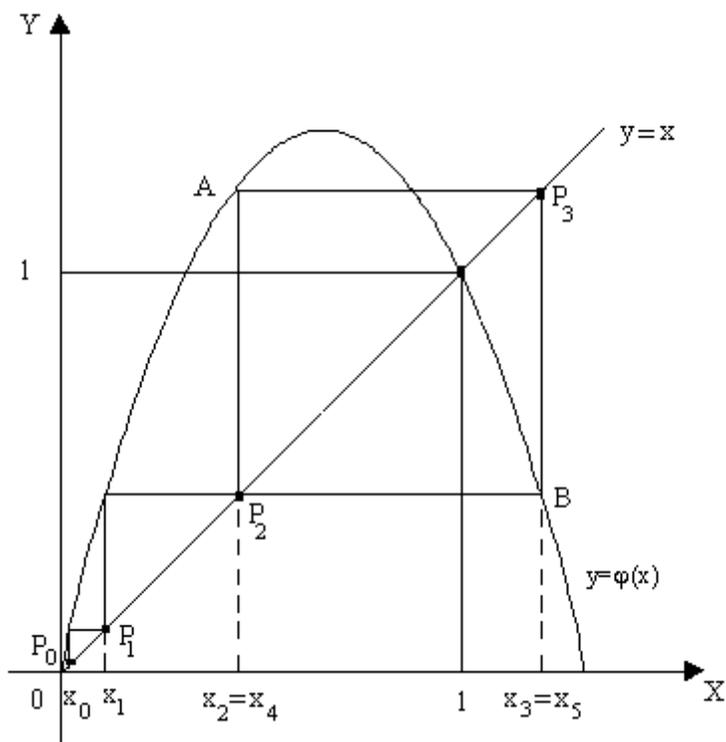


Figura 13.3

($a = 1.98$, fig. B), un esempio di periodo 2 ($a = 2.3$, fig. C), un esempio di periodo 4 ($a = 2.5$, fig. D) e due esempi di caos ($a = 2.8$, fig. E - $a = 3$, fig. F).

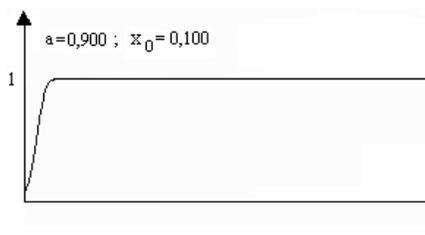


Figura 13.4 - A

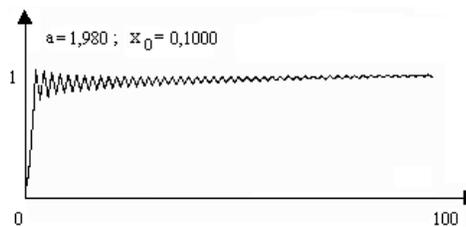


Figura 13.4 - B

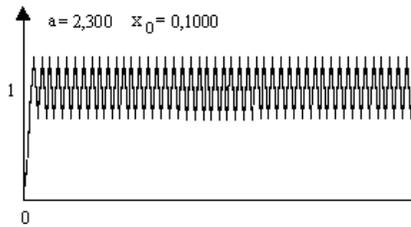


Figura 13.4 - C

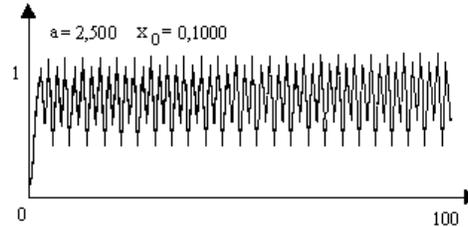


Figura 13.4 - D

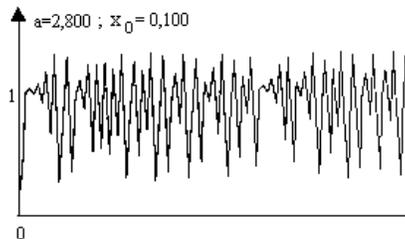


Figura 13.4 - E

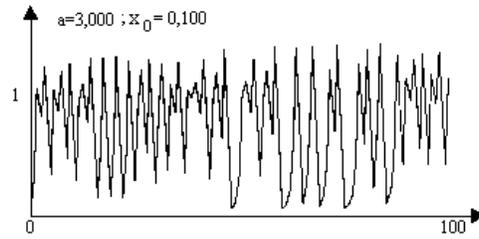


Figura 13.4 - F

Notiamo come al crescere di a tende ad aumentare l'ampiezza delle oscillazioni. Nel caso $a = 3, x_0 = 0.1$, le oscillazioni arrivano quasi a "toccare" l'asse delle ascisse. Se questo succede, cioè se esiste un numero naturale \bar{n} per cui $x_{\bar{n}} \leq 0$, allora è chiaro che $x_n \leq 0$ per ogni $n > \bar{n}$. Ciò significa che la popolazione si estingue dopo un numero finito \bar{n} di generazioni.

Aumentando a questo succede sempre, però in modo del tutto imprevedibile, che cambia completamente per piccole variazioni del valore iniziale x_0 . Ad esempio, per $x_0 = 0.1$, tabulando al calcolatore i primi 100 valori di x_n e facendo variare a da 3 a 4 con passo 0.01, si vede che per tutti i valori di a da 3.01 a 3.54 la popolazione si estingue in un numero finito $\bar{n} < 100$ di generazioni, però \bar{n} varia in modo del tutto irregolare al crescere di a . Per $a \geq 3.55$ invece si ha sempre $\bar{n} = 2$, cioè la popolazione si estingue già alla seconda generazione.

Questo accade in generale: fissato x_0 , se il tasso di crescita (a) è abbastanza grande si ha $x_1 > 1 + \frac{1}{a}$ e allora, $x_2 = \varphi(x_1) = x_1(a + 1 - ax_1) \leq (a + 1 - a - 1) = 0$ e quindi $x_n \leq 0$ per ogni $n \geq 2$.

Quindi, nel caso della logistica discreta, all'aumentare del tasso di crescita a si verificano successivamente 6 situazioni diverse:

1. crescita monotona;
2. oscillazioni smorzate;
3. periodicità di periodo 2, 4, 8, 16, \dots , $2^n, \dots$;
4. caos;

5. estinzione in un tempo finito;
6. estinzione alla seconda generazione.

Niente di tutto ciò evidentemente, può succedere per la logistica continua.

13.1 RAPPORTI TRA EQUAZIONI DIFFERENZIALI ED EQUAZIONI ALLE DIFFERENZE

Uno dei metodi piú semplici di soluzione numerica approssimata di una generica equazione differenziale del primo ordine in forma normale nella funzione incognita $x(t)$:

$$x'(t) = F(t, x(t)) \quad (1)$$

con la condizione iniziale $x(0) = x_0$ é il *metodo di Eulero*. Supponiamo per semplicitá di voler risolvere la in un certo intervallo $[0, T]$ e discretizziamo il tempo t , suddividendo l'intervallo $[0, T]$ in n intervallini uguali di ampiezza $h = T/n$. supponiamo che n sia abbastanza grande (h abbastanza piccolo) in modo che nell' i -esimo intervallino $(ih, (i+1)h)$ la derivata $x'(t)$ si possa considerare praticamente costante e uguale al rapporto incrementale, cioè:

$$x'(t) \simeq \frac{x((i+1)h) - x(ih)}{h} \quad \forall t \in (ih, (i+1)h)$$

allora il grafico della soluzione $x(t)$ della nell'intervallino considerato si può approssimare con la sua retta tangente nel punto iniziale. Prendendo $t = ih$, dall'uguaglianza precedente si ricava:

$$\frac{x((i+1)h) - x(ih)}{h} \simeq x'(ih) = F(ih, x(ih))$$

A questo punto consideriamo solo i valori di x nei punti ih ($i = 0, 1, 2, 3, \dots, n$) e indichiamo $x(ih)$ con x_i . Sostituendo nell'uguaglianza precedente il segno \simeq (circa uguale) col segno $=$ (uguale) si ottiene:

$$x_{i+1} - x_i = hF(ih, x_i) \quad (2)$$

Questa é un'equazione alle differenze associata all'equazione differenziale (1) mediante il processo di discretizzazione di t con passo h (ad essere precisi non si tratta di una sola equazione alle differenze, bensí di una famiglia di equazioni alle differenze diverse, una per ogni valore di h).

Se noi risolviamo l'equazione alle differenze (2) e poi uniamo con una spezzata i successivi punti (ih, x_i) , é intuitivo che otterremo una soluzione approssimata dell'equazione differenziale e che l'approssimazione sará tanto migliore quanto piú é piccolo il passo h .

In opportune ipotesi (verificate nei casi piú comuni) si può dimostrare che per $h \rightarrow 0$ la soluzione approssimata dell'equazione differenziale converge alla soluzione esatta $x(t)$ in ogni intervallo limitato $[a, b]$. In questo consiste il metodo di Eulero.

Esempio 13.1 CRESCITA ESPONENZIALE.

L'equazione differenziale della crescita esponenziale é $x'(t) = ax(t)$. Se noi la risolviamo col metodo di Eulero con passo h , ad essa corrisponde la seguente equazione alle differenze:

$$x_{i+1} - x_i = hax_i \quad x_{i+1} = (ha + 1)x_i$$

che coincide con l'equazione alle differenze della crescita esponenziale discreta con tasso di crescita ha (equazione 5).

Dato il valore iniziale x_0 , la sua soluzione é:

$$x_i = (ha + 1)^i x_0$$

Supponiamo ad esempio, di voler risolvere l'equazione differenziale $x'(t) = ax(t)$ con $x(0) = x_0$ nell'intervallo $[0, T]$, col metodo di Eulero, suddividendo l'intervallo $[0, T]$ in n intervallini uguali di ampiezza $h = \frac{T}{n}$. Nel punto T il metodo di Eulero ci dá:

$$x(T) = x_n = (ha + 1)^n x_0 = x_0 \left(1 + \frac{Ta}{n}\right)^n$$

Invece la soluzione esatta dell'equazione differenziale é $x(t) = x_0 e^{at}$ ed é facile vedere che nel punto T per $n \rightarrow \infty$ la soluzione approssimata tende alla soluzione esatta. Infatti:

$$\lim_{n \rightarrow \infty} x_n = \lim_{n \rightarrow \infty} x_0 \left(1 + \frac{Ta}{n}\right)^n x_0 e^{Ta}$$

Analogamente si puó far vedere che la soluzione approssimata tende alla soluzione esatta in ogni punto dell'intervallo $[0, T]$. ■

Esempio 13.2 CRESCITA LOGISTICA.

L'equazione differenziale della crescita logistica é $x'(t) = ax(t) - bx^2(t)$. Se noi la risolviamo col metodo di Eulero con passo h , ad essa corrisponde la seguente equazione alle differenze:

$$x_{i+1} - x_i = hax_i - hb x_i^2$$

Nel caso particolare $a = b$ (che equivale a porre uguale a 1 il livello di saturazione $M = \frac{a}{b}$ e non é restrittivo, perché equivale a cambiare l'unitá di misura sull'asse x) quest'equazione diventa:

$$x_{i+1} = (ha + 1)x_i - hax_i^2$$

e coincide con l'equazione alle differenze della crescita logistica discreta con tasso di crescita ha (vedi equazione 6).

Risolviendo quest'equazione col calcolatore, il che equivale a risolvere l'equazione differenziale della crescita logistica col metodo di Eulero, si ha una soluzione approssimata dell'equazione differenziale. Fissato un intervallo $[0, T]$ e una condizione iniziale $x(0) = x_0$ e posto $h = \frac{T}{n}$, si può dimostrare che per $n \rightarrow \infty$ ($h \rightarrow 0$), la soluzione approssimata tende alla soluzione esatta dell'equazione differenziale in ogni punto dell'intervallo $[0, T]$.

Peró nel caso della crescita logistica bisogna stare molto attenti alla scelta del passo h . Se h é troppo grande, la soluzione approssimata può presentare delle oscillazioni smorzate (questo succede se $\frac{1}{a} < h < \frac{2}{a}$), oppure un comportamento periodico (se $\frac{2}{a} < h < \frac{2.57}{a}$) o un comportamento caotico (se $h > \frac{2.57}{a}$) oppure può estinguersi (cioé diventare negativa) dopo un numero finito di passi. Invece, scegliendo h abbastanza piccolo, (e precisamente $h < \frac{1}{a}$) la soluzione approssimata ha un andamento crescente che tende asintoticamente al valore di saturazione 1, proprio come succede per la soluzione esatta $x(t)$. ■

14 Processi stocastici di nascita e morte

Se noi consideriamo una popolazione di esseri viventi che ha un tasso di natalità a ed un tasso di mortalità b costanti nel tempo ed indichiamo con $x(t)$ il numero di individui della popolazione al variare del tempo e con $x_0 = x(0)$ il numero iniziale di individui (al tempo $t = 0$), sappiamo che la crescita della popolazione può essere modellizzata con l'equazione differenziale

$$\begin{cases} x'(t) = (a - b)x(t) \\ x(0) = x_0 \end{cases}$$

che ha soluzione $x(t) = x_0 e^{(a-b)t}$: crescita esponenziale se $a > b$, decadimento esponenziale se $a < b$, popolazione costante se $a = b$.

A voler essere pignoli, però questo modello presenta 2 difetti. Il primo è che la variabile dipendente $x(t)$ è una variabile discreta, non continua: $x(t)$ è un numero intero e non un numero reale e la funzione $x(t)$ è una funzione costante a tratti, che presenta delle discontinuità a salto in tutti gli istanti in cui si verifica una nascita o una morte. La derivata $x'(t)$, a voler essere precisi non esiste in tutti questi istanti ed è uguale a zero in tutti gli altri.

Il secondo difetto è che in natura le nascite e le morti non avvengono con precisione cronometrica, avvengono probabilisticamente. Indicando con $n(t)$ il numero di individui al tempo t e supponendo fissato il numero iniziale di individui $n_0 = n(0)$, il numero $n(t)$ non è affatto determinato e potrà essere diverso di volta in volta. Piuttosto che una funzione $n(t)$ della sola variabile t converrà allora descrivere il fenomeno con una funzione di 2 variabili $p(n, t)$ che esprime la probabilità che al tempo t il numero di individui della popolazione sia n . Per ogni valore di t abbiamo allora una variabile aleatoria con distribuzione di probabilità $p(n, t)$, cioè l'andamento del fenomeno è descritto da una variabile aleatoria che cambia col tempo, cioè da quello che si chiama un *processo stocastico*

La variabile t è continua, la variabile n è discreta e per ogni valore di t si ha $\sum_{n=0}^{\infty} p(n, t) = 1$. Supponiamo per semplicità che la popolazione sia composta da organismi unicellulari che possono morire oppure riprodursi dividendosi in due e facciamo inoltre le seguenti ipotesi.

1. Tutti gli individui sono biologicamente identici ed hanno la stessa probabilità di riprodursi e la stessa probabilità di morire, che non dipendono né dal tempo, né dalla numerosità della popolazione.
2. La probabilità di riprodursi e la probabilità di morire non dipendono dalla storia passata dell'organismo (ipotesi poco realistica).
In particolare non teniamo conto di fattori come l'età ed il sesso.
3. La popolazione è chiusa, cioè non è soggetta né ad immigrazione, né a emigrazione.

Inoltre siano a, b i tassi di nascita e morte dal punto di vista probabilistico, cioè poniamo :

- $a\Delta t$ = probabilità che un individuo si divida in due in un intervallo di tempo Δt molto piccolo.
- $b\Delta t$ = probabilità che un individuo muoia in un intervallo di tempo Δt molto piccolo.

La condizione che al tempo $t = 0$ siano presenti esattamente n_0 individui si esprime così:

$$p(n, 0) = \begin{cases} 1 & \text{sen } n = n_0 \\ 0 & \text{sen } n \neq n_0 \end{cases}$$

Consideriamo ora un intervallo di tempo molto piccolo Δt e cerchiamo di esprimere $p(n, t + \Delta t)$. Osserviamo che ci sono soltanto 3 situazioni mutualmente esclusive che possono condurre ad una popolazione di numerosità n al tempo $t + \Delta t$:

1. Al tempo t la popolazione aveva $n - 1$ individui ed uno di essi si è diviso in due nell'intervallo di tempo $(t, t + \Delta t)$. La probabilità che ciò è $p(n - 1, t)(n - 1)a\Delta t$.
2. Al tempo t la popolazione aveva $n + 1$ individui ed uno di essi è morto nell'intervallo di tempo $(t, t + \Delta t)$. La probabilità che ciò accada è $p(n + 1, t)(n + 1)b\Delta t$
3. Al tempo t la popolazione aveva n individui e non si sono verificate né nascite, né morti nell'intervallo di tempo $(t, t + \Delta t)$. La probabilità che ciò accada è $p(n, t)(1 - na\Delta t - nb\Delta t)$.

Notiamo che se Δt è molto piccolo, noi possiamo ignorare la possibilità di eventi multipli, cioè trascurare la probabilità che nell'intervallo di tempo $(t, t + \Delta t)$ ci sia più di una nascita o morte. Sommando fra loro le 3 probabilità precedenti si ha:

$$p(n, t + \Delta t) = p(n - 1, t)(n - 1)a\Delta t + p(n + 1, t)(n + 1)b\Delta t + p(n, t) - p(n, t)n(a + b)\Delta t$$

Da questa formula, nota la condizione iniziale che dá $p(n, 0)$, si possono calcolare ricorsivamente (e tabulare al computer) i valori di $p(n, t)$ per ogni n e per ogni t . Sotto opportune ipotesi, si può dimostrare che la distribuzione di probabilità $p(n, t)$ per $p \rightarrow +\infty$ tende ad una distribuzione normale di media e scarto opportuni.

Se nella formula precedente portiamo $p(n, t)$ a primo membro e dividiamo per Δt otteniamo:

$$\frac{p(n, t + \Delta t) - p(n, t)}{\Delta t} = a(n-1)p(n-1, t) + b(n+1)p(n+1, t) - n(a+b)p(n, t)$$

e facendo il limite per $\Delta t \rightarrow 0$

$$\frac{dp(n, t)}{dt} = a(n-1)p(n-1, t) + b(n+1)p(n+1, t) - n(a+b)p(n, t)$$

Questa é un'equazione differenziale alle differenze che esprime la derivata della funzione $p(n, t)$ rispetto al tempo mediante i valori di p nei 3 punti (n, t) , $(n-1, t)$, $(n+1, t)$. Risolvere quest'equazione é difficile, mentre é piú facile trovare la media $M(t)$ e la varianza $V(t)$ della variabile aleatoria $p(n, t)$ in funzione di t .

Ricordiamo che la media e la varianza sono definite nel seguente modo:

$$\begin{aligned} M(t) &= \sum_{i=0}^{\infty} i p(i, t) \\ V(t) &= \sum_{i=0}^{\infty} (i - M(t))^2 p(i, t) = \sum_{i=0}^{\infty} (i^2 - 2iM(t) + M^2(t)) p(i, t) = \\ &= \sum_{i=0}^{\infty} i^2 p(i, t) - 2M(t) \sum_{i=0}^{\infty} i p(i, t) + M^2(t) \sum_{i=0}^{\infty} p(i, t) = \\ &= \sum_{i=0}^{\infty} i^2 p(i, t) - M^2(t) \end{aligned}$$

Derivando $M(t)$ rispetto al tempo ed usando la (1) otteniamo:

$$\begin{aligned} \frac{dM(t)}{dt} &= \sum_{i=0}^{\infty} i \frac{dp(i, t)}{dt} = \sum_{i=1}^{\infty} i \frac{dp(i, t)}{dt} \\ &\text{(in quanto il primo termine della sommatoria precedente é zero)} \\ &= \sum_{i=1}^{\infty} ai(i-1)p(i-1, t) + \sum_{i=1}^{\infty} bi(i+1)p(i+1, t) + \\ &\quad - \sum_{i=1}^{\infty} i^2(a+b)p(i, t) \end{aligned}$$

Adesso nella prima sommatoria poniamo $i-1 = k$ e nella seconda poniamo $i+1 = k$ e otteniamo

$$\frac{dM(t)}{dt} = \sum_{k=0}^{\infty} ak(k+1)p(k, t) + \sum_{k=2}^{\infty} bk(k-1)p(k, t) - \sum_{k=1}^{\infty} k^2(a+b)p(k, t)$$

osserviamo che le prime due sommatorie si possono far partire entrambe da 1, perché i primi termini sono *zero*. Allora riunendo le 3 sommatorie in una si ha:

$$\begin{aligned} \frac{dM(t)}{dt} &= \sum_{k=1}^{\infty} p(k, t)(ak^2 + ak + bk^2 - 2bk^2 - bk - ak^2 - bk^2) = \\ &= \sum_{k=1}^{\infty} (a - b)kp(k, t) = (a - b)M(t) \end{aligned}$$

Questa é l'equazione differenziale della crescita esponenziale che, unitamente alla condizione iniziale $M(0) = n_0$ fornisce l'espressione di $M(t)$

$$M(t) = n_0 e^{(a-b)t}$$

Osserviamo che $M(t)$ si comporta esattamente come fa $x(t)$ nel modello deterministico.

Con calcoli analoghi, un pó piú complicati, si può trovare anche l'espressione della varianza $V(t)$. A questo scopo poniamo

$$W(t) = \sum_{i=0}^{\infty} i^2 p(i, t) = \sum_{i=1}^{\infty} i^2 p(i, t)$$

(in modo che $V(t) = W(t) - M^2(t)$) e calcoliamo la derivata di $W(t)$ rispetto al tempo:

$$\begin{aligned} \frac{dW(t)}{dt} &= \sum_{i=1}^{\infty} ai^2(i-1)p(i-1, t) + \sum_{i=1}^{\infty} bi^2(i+1)p(i+1, t) + \\ &= - \sum_{i=1}^{\infty} i^3(a+b)p(i, t) \end{aligned}$$

ponendo $i-1 = k$ nella prima sommatoria e $i+1 = k$ nella seconda si ha:

$$\begin{aligned} \frac{dW(t)}{dt} &= \sum_{k=0}^{\infty} ak(k+1)^2 p(k, t) + \sum_{k=2}^{\infty} bk(k-1)^2 p(k, t) + \\ &= - \sum_{k=1}^{\infty} k^3(a+b)p(k, t) \end{aligned}$$

le prime 2 sommatorie si possono far partire entrambe da 1 e allora riunendo le 3 sommatorie in una si ha:

$$\begin{aligned} \frac{dW(t)}{dt} &= \sum_{k=1}^{\infty} p(k, t)(ak^3 + 2ak^2 + ak + bk^3 - 2bk^2 + bk - ak^3 - bk^3) = \\ &= \sum_{k=1}^{\infty} (2a - 2b)k^2 p(k, t) + \sum_{k=1}^{\infty} (a + b)kp(k, t) = 2(a - b)W(t) + (a + b)M(t) \end{aligned}$$

$$\frac{dW(t)}{dt} = 2(a - b)W(t) + n_0(a + b)e^{(a-b)t}$$

Questa é un'equazione differenziale in $W(t)$ che si può risolvere portando $2(a-b)W(t)$ al primo membro e moltiplicando ambo i membri per $e^{-2(a-b)t}$:

$$e^{-2(a-b)t} \frac{dW(t)}{dt} - 2(a-b)W(t)e^{-2(a-b)t} = n_0(a+b)e^{-(a-b)t}$$

$$D\left(W(t)e^{-2(a-b)t}\right) = n_0(a+b)e^{-(a-b)t}$$

$$W(t)e^{-2(a-b)t} = -n_0(a+b) \frac{e^{-(a-b)t}}{a-b} + C$$

La costante C si determina mediante la condizione iniziale $W_0 = n_0^2$, che fornisce

$$n_0^2 = -n_0 \frac{a+b}{a-b} + C$$

$$C = n_0^2 + n_0 \frac{a+b}{a-b}$$

Si ha allora:

$$V(t) = W(t) - M^2(t) = e^{2(a-b)t} \left(n_0^2 + n_0 \frac{a+b}{a-b} - n_0 e^{-(a-b)t} \frac{a+b}{a-b} \right) - n_0^2 e^{2(a-b)t}$$

$$V(t) = n_0 \frac{a+b}{a-b} e^{(a-b)t} \left(e^{(a-b)t} - 1 \right)$$

Questo vale se $a \neq b$. Nel caso piú semplice $a = b$ (tasso di nascita = tasso di morte) si ha $M(T) = n_0$, $\frac{dW(t)}{dt} = n_0 a + b$, $W(t) = n_0^2 + n_0(a+b)t$, $V(t) = W(t) - M^2(t) = n_0(a+b)t = 2an_0t$.

Notiamo che la varianza $V(t)$ dipende non soltanto dalla differenza $a - b$ (il tasso intrinseco di crescita), ma anche dalla loro somma. Questo non é sorprendente, perché se a, b sono piú grandi la popolazione subisce oscillazioni maggiori.

Data la varianza $V(t)$ si può calcolare anche lo scarto quadratico medio $\sigma(t) = \sqrt{V(t)}$ ed il coefficiente di variazione $c(t) = \frac{\sigma(t)}{M(t)}$. Risulta:

$$c(t) = \sqrt{\frac{a+b}{n_0(a-b)} \frac{e^{(a-b)t} - 1}{e^{(a-b)t}}} \quad \text{se } a \neq b; \quad c(t) = \sqrt{\frac{2at}{n_0}} \quad \text{se } a = b$$

Notiamo che per $t \rightarrow \infty$, $c(t) \rightarrow \infty$, se $a = b$, $c(t) \rightarrow \sqrt{\frac{a+b}{n_0|a-b|}}$ se $a \neq b$.

Se n_0 é molto grande, $c(t)$ si mantiene molto piccolo (se $a \neq b$), il che conferma la validità del modello deterministico di crescita esponenziale se n_0 é grande.

Cé un altro punto interessante legato al modello stocastico, che si presenta diverso dal modello deterministico: la probabilità $p_E(t)$ che al tempo t la popolazione si sia estinta. Calcolare $p_E(t)$ é molto difficile, ma il risultato é

espresso dalla seguente formula:

$$p_E(t) = \left(\frac{be^{(a-b)t} - b}{ae^{(a-b)t} - b} \right)^{n_0} \quad \text{se } a \neq b$$
$$p_E(t) = \frac{at}{at + 1} \quad \text{se } a = b$$

Facendo il limite di $p_E(t)$ per $t \rightarrow +\infty$ si ha la probabilità p_E che la popolazione si estingua prima o poi. Si ha:

- se $a > b$, $p_E = \left(\frac{b}{a}\right)^{n_0}$. L'estinzione si può verificare anche se il valore medio $M(t)$ cresce esponenzialmente. Però l'estinzione è tanto meno probabile quanto più è grande il rapporto $\frac{a}{b}$ e quanto più grande è n_0 .
- se $a \leq b$, $p_E = 1$. L'estinzione prima o poi, è certa

15 Diffusione del calore

Una delle piú interessanti equazioni differenziali alle derivate parziali é l'*equazione del calore*.

Consideriamo una funzione di 2 variabili $u(x, t)$ dove x rappresenta una dimensione spaziale (l'ascissa lungo una retta) e t rappresenta il tempo. L'equazione del calore é:

$$\frac{\delta u}{\delta t} = D \frac{\delta^2 u}{\delta x^2} \quad (1)$$

dove D é una costante positiva, detta *coefficiente di diffusione*.

La funzione $u(x, t)$ puó essere interpretata in 2 modi principali:

1. Come temperatura nel punto di ascissa x all'istante t .
2. Come densitá di una certa sostanza A nel punto di ascissa x all'istante t .

Nel primo caso l'equazione differenziale (1) rappresenta il processo di diffusione del calore per conduzione lungo un'ideale sbarretta rettilinea (i cui punti vengono individuati da una sola coordinata x).

Nel secondo caso rappresenta il processo di diffusione della sostanza A (per il semplice moto casuale delle molecole di A , dovuto all'agitazione termica) lungo un mezzo che possa essere schematizzato come una retta (ad esempio un fiume, un canale, un tubo, ...). Non é una rappresentazione molto realistica, perché lo spazio reale ha tre dimensioni e non una, ma é facile generalizzare l'equazione del calore in 3 dimensioni: basta supporre che la funzione incognita u (temperatura o densitá) sia una funzione di 4 variabili $u(x, y, z, t)$, le tre variabili spaziali ed il tempo, ed allora, l'equazione del calore diventa:

$$\frac{\delta u}{\delta t} = D \left(\frac{\delta^2 u}{\delta x^2} + \frac{\delta^2 u}{\delta y^2} + \frac{\delta^2 u}{\delta z^2} \right) \quad (2)$$

Deduzione dell'equazione del calore

Per giustificare l'equazione (1) riferiamoci al caso della diffusione di una sostanza A ed indichiamo con $v(x, t)$ il flusso di A attraverso x all'istante t , cioé la massa di A che passa attraverso il punto di ascissa x al tempo t , nell'unitá di tempo (per convenzione consideriamo $v(x, t) > 0$ se il flusso é da sinistra a destra, $v(x, t) < 0$ se il flusso é da destra a sinistra). La massa di A contenuta in un intervallo $(x, x + \Delta x)$ all'istante t é:

$$M(t) = \int_x^{x+\Delta x} u(x, t) dx$$

la derivata $M'(t)$ cioè la velocità di variazione di $M(t)$ al variare di t per la conservazione della massa, è uguale al flusso entrante a sinistra $v(x, t)$ meno il flusso uscente a destra $v(x + \Delta x, t)$:

$$M'(t) = \frac{d}{dt} \int_x^{x+\Delta x} u(x, t) dx = \int_x^{x+\Delta x} \frac{\delta u}{\delta t}(x, t) dx = v(x, t) - v(x + \Delta x, t)$$

e dividendo ambo i membri per Δx :

$$\frac{1}{\Delta x} \int_x^{x+\Delta x} \frac{\delta u}{\delta t}(x, t) dx = -\frac{v(x + \Delta x, t) - v(x, t)}{\Delta x}$$

ma per il teorema della media sugli integrali, il primo membro è uguale a $\frac{\delta u}{\delta t}(c, t)$, dove c è un punto opportuno dell'intervallo $(x, x + \Delta x)$. Facendo il limite per $\Delta x \rightarrow 0$, il punto c tende a x e quindi, otteniamo:

$$\frac{\delta u}{\delta t}(x, t) = -\frac{\delta v}{\delta x}(x, t) \quad (3)$$

ora facciamo l'ipotesi che il flusso $v(x, t)$ sia proporzionale al “gradiente di densità” cioè alla pendenza del grafico di $u(x, t)$ rispetto a x (fig.15.1). Supponiamo cioè che $v(x, t) = -D \frac{\delta u}{\delta x}(x, t)$, dove D è una costante positiva detta “coefficiente di diffusione” ed il segno $-$ è dovuto al fatto che se $\frac{\delta u}{\delta x} > 0$, cioè $u(x, t)$ è crescente rispetto a x , allora il flusso è negativo, cioè va verso sinistra (fig. 15.1).

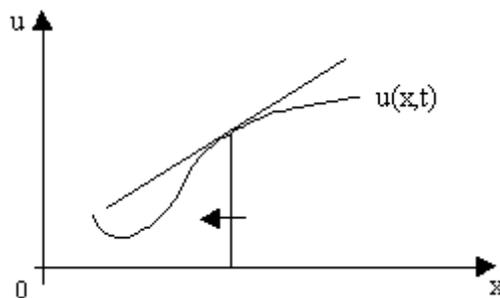


Figura 15.1

Con questa ipotesi, che si può verificare sperimentalmente con facilità, l'equazione 3 si trasforma in:

$$\frac{\delta u}{\delta t}(x, t) = D \frac{\delta^2 u}{\delta x^2}(x, t)$$

che é l'equazione del calore. Il ragionamento vale anche nel caso della diffusione del calore (anziché di una sostanza A) perché in tal caso $u(x, t)$ rappresenta la temperatura e si sa (per verifica sperimentale diretta) che il flusso di calore é proporzionale al gradiente di temperatura.

CONDIZIONI INIZIALI PER L'EQUAZIONE DEL CALORE

Per risolvere l'equazione del calore (1), occorre specificare opportune condizioni iniziali, che sostanzialmente consistono nel conoscere $u(x, 0)$, cioè la distribuzione spaziale della temperatura (o della densità) all'istante iniziale $t = 0$ lungo l'asse x . Distinguiamo però 2 casi diversi:

1. CASO 1 DIFFUSIONE LIBERA.

Supponiamo che $u(x, t)$ sia definita per ogni numero reale x , cioè consideriamo la diffusione del calore (o di una certa sostanza A) lungo un'ipotetica sbarretta rettilinea indefinita. In tal caso, si può dimostrare che la conoscenza di $u(x, 0)$ determina univocamente $u(x, t)$ per ogni $t > 0$.

2. CASO 2 DIFFUSIONE CON CONDIZIONI AL CONTORNO.

Supponiamo che x non sia libero di variare in tutto \mathbf{R} , ma solo in un certo intervallo limitato e chiuso $[a, b]$. Allora, si può dimostrare che per conoscere $u(x, t)$ per ogni x in $[a, b]$ e per ogni $t > 0$ non é sufficiente conoscere $u(x, 0)$ per ogni x in $[a, b]$, ma bisogna anche sapere cosa succede agli estremi ($x = a, x = b$) per ogni $t > 0$. Bisogna cioè conoscere $u(a, t)$ e $u(b, t)$ per ogni $t > 0$ (condizioni di Dirichlet), oppure bisogna conoscere $\frac{\delta u}{\delta x}(a, t)$ e $\frac{\delta u}{\delta x}(b, t)$ per ogni $t > 0$ (condizioni di Neumann).

Questo é abbastanza facile da capire se pensiamo al problema della diffusione del calore in una sbarretta di lunghezza finita: non basta conoscere la temperatura iniziale in ogni punto della sbarretta, ma bisogna sapere anche quanto calore entra ed esce dai 2 estremi della sbarretta per ogni $t > 0$, oppure bisogna sapere la temperatura ai 2 estremi per ogni $t > 0$.

ALCUNE PROPRIETÁ DELLE SOLUZIONI DELL'EQUAZIONE DEL CALORE

Supponiamo che ad un certo istante t il grafico di $u(x, t)$ in funzione di x sia come in fig.15.2. Nei punti in cui $u(x, t)$ é una funzione convessa di x (come il punto a in fig. 2), cioè $\frac{\delta^2 u}{\delta x^2} > 0$, si ha $\frac{\delta u}{\delta t} > 0$, cioè $u(x, t)$ cresce col tempo, ossia il grafico di fig. 2 tende ad alzarsi al passare del tempo (come indica la freccia in figura). Invece nei punti in cui $u(x, t)$ é concava rispetto a x (come il punto b in fig.15.2), cioè $\frac{\delta^2 u}{\delta x^2} < 0$, si ha $\frac{\delta u}{\delta t} < 0$ e quindi, il grafico di fig. 15.2 tende ad abbassarsi col tempo.

É facile capire perché questo succede : se $u(x, t)$ rappresenta la temperatura, il calore passa dai punti caldi ai punti freddi (e non viceversa), quindi i punti freddi si scaldano, i punti caldi si raffreddano (e non viceversa) e

alla fine tutte le differenze di temperatura tendono a sparire. Questo, in sostanza non é altro che il secondo principio della termodinamica (principio della degradazione dell'energia, o del livellamento del calore), visto come conseguenza dell'equazione del calore.

Ci si può chiedere se esistono *soluzioni stazionarie* dell'equazione del calore (1), cioè soluzioni che non dipendono dal tempo, ovvero sono del tipo $u(x, t) = f(x)$. Se ciò succede si ha $\frac{\delta u}{\delta t} = 0$ e quindi, dev'essere $\frac{\delta^2 u}{\delta x^2} = 0$, da cui si ricava $f(x) = Ax + B$, cioè il grafico di $f(x)$ dev'essere una retta.

Vediamo cosa può succedere in 3 diversi casi particolari:

1. CASO 1 DIFFUSIONE DEL CALORE IN UNA SBARRETTA ISOLATA TERMICAMENTE AGLI ESTREMI.

In questo caso agli estremi della sbarretta ($x = a, x = b$) il flusso di calore $v(x, t) = -D \frac{\delta u}{\delta x}(x, t)$ é sempre 0, cioè $\frac{\delta u}{\delta x}(a, t) = \frac{\delta u}{\delta x}(b, t) = 0$ per ogni t . Allora, qualunque sia la distribuzione iniziale della temperatura, le differenze si livellano rapidamente e per $t \rightarrow +\infty$ la temperatura tende ad essere uguale dappertutto (fig.15.3). Però c'è una quantità che rimane sempre uguale al passare del tempo: l'integrale di $u(x, t)$ rispetto a x , nell'intervallo $[a, b]$:

$$Q(t) = \int_a^b u(x, t) dx = \text{costante} = Q$$

Questo integrale rappresenta la quantità totale di calore contenuta nella sbarretta. Il fatto che $Q(T)$ sia costante rappresenta, in sostanza, il principio di conservazione dell'energia (o primo principio della termodinamica). É chiaro che il valore finale di equilibrio a cui tende la temperatura nella sbarretta é $T = Q/(b - a)$.

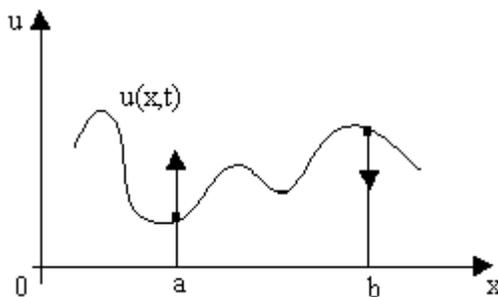


Figura 15.2

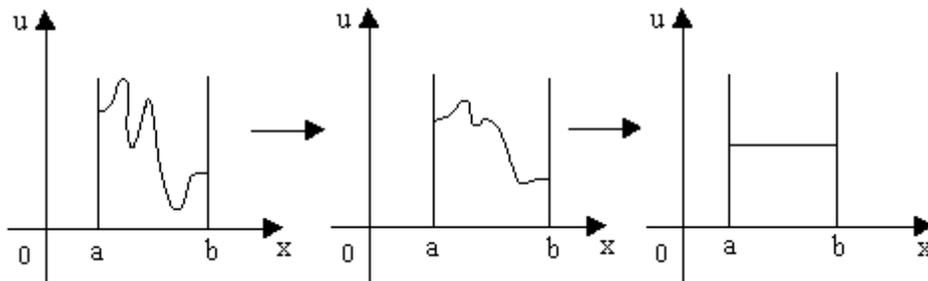


Figura 15.3

2. DIFFUSIONE DEL CALORE IN UNA SBARRETTA I CUI 2 ESTREMI a, b SONO TENUTI A TEMPERATURE FISSE $T(a), T(b)$. Anche in questo caso le differenze di temperatura all'interno della sbarretta si livellano rapidamente e per $t \rightarrow +\infty$ la temperatura tende ad una distribuzione di equilibrio rappresentata da una retta. Nel caso della fig.15.4, in cui $T(a) > T(b)$, la temperatura di equilibrio nella sbarretta decresce linearmente da $T(a)$ a $T(b)$ quando x cresce da a a b . In questo caso però la sbarretta non può essere isolata, anzi il calore fluisce ininterrottamente nella sbarretta, da sinistra verso destra, entrando dall'estremo a ed uscendo dall'estremo b .
3. DIFFUSIONE LIBERA DEL CALORE IN UNA SBARRETTA INDEFINITA IN CUI LA TEMPERATURA INIZIALE È DIVERSA DA 0 SOLO IN UN CERTO INTERVALLO $[a, b]$, MENTRE È 0 FUORI DI $[a, b]$.

In tal caso il calore esce dell'intervallo $[a, b]$ e si "espande" sia verso destra che verso sinistra, quindi la distribuzione iniziale della temperatura si allarga sempre di più e tende asintoticamente al valore $t = 0$, senza però raggiungerlo mai (fig.15.5). Inoltre per ogni tempo t si ha :

$$Q(t) = \int_{-\infty}^{+\infty} u(x, t) dx = \text{costante} = Q$$

cioè l'area complessiva compresa tra la curva e l'asse x rimane costante (principio di conservazione dell'energia).

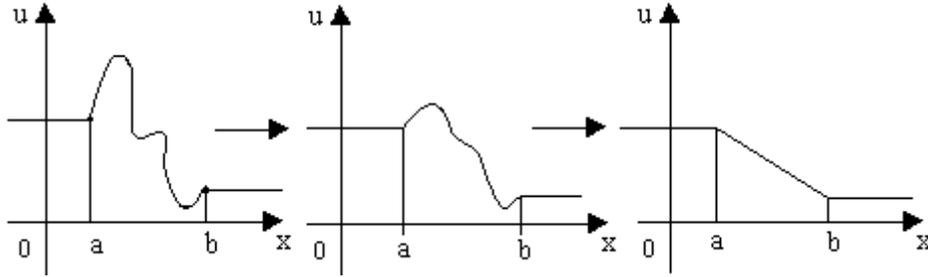


Figura 15.4

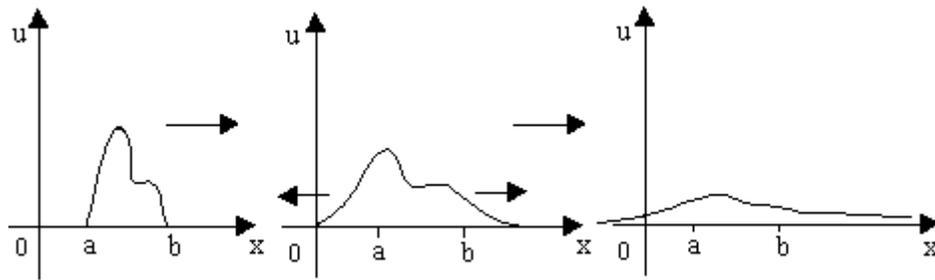


Figura 15.5

RISOLUZIONE NUMERICA APPROSSIMATA DELL'EQUAZIONE DEL CALORE

In generale, non esistono delle formule esplicite semplici per esprimere la soluzione dell'equazione del calore: bisogna risolverla con metodi numerici approssimati, discretizzando sia lo spazio x che il tempo t ed usando delle opportune formule per approssimare le derivate.

Consideriamo una qualsiasi funzione $f(x)$ di una sola variabile e vediamo come si possono approssimare la sua derivata prima $f'(x)$ e la sua derivata seconda $f''(x)$. Scelto un numero positivo h abbastanza piccolo, si può approssimare $f'(x)$ con un opportuno rapporto incrementale, in 3 modi diversi:

$$\begin{aligned}
 f'(x) &\simeq \frac{f(x+h) - f(x)}{h} = \nabla f && \text{differenza in avanti} \\
 f'(x) &\simeq \frac{f(x) - f(x-h)}{h} = \bar{\nabla} f && \text{differenza all'indietro} \\
 f'(x) &\simeq \frac{f(x+h) - f(x-h)}{2h} = \delta f = \frac{\nabla f + \bar{\nabla} f}{2} && \text{differenza centrale}
 \end{aligned}$$

Per quanto riguarda la derivata seconda $f''(x)$ usiamo la seguente approssimazione alle differenze centrali:

$$f''(x) \simeq \frac{f(x+h) - 2f(x) + f(x-h)}{h^2} = \nabla \bar{\nabla} f = \bar{\nabla} \nabla f$$

Ciò premesso, per risolvere l'equazione del calore (1) discretizziamo x con un certo passo h e t con un certo passo k . Per semplicità, indichiamo con u_{ij} il valore $u(ih, jk)$, dove i, j sono numeri interi qualsiasi (anche negativi). Usiamo per la derivata temporale $\frac{\delta u}{\delta t}$ l'approssimazione alle differenze in avanti:

$$\frac{\delta u}{\delta t}(ih, jk) \simeq (u(ih, (j+1)k) - u(ih)) / k = (u_{i,j+1} - u_{i,j}) / k$$

e per la derivata spaziale seconda $\frac{\delta^2 u}{\delta x^2}$ l'approssimazione alle differenze centrali:

$$\frac{\delta^2 u}{\delta x^2}(ih, jk) \simeq (u((i+1)h, jk) - 2u(ih, jk) + u((i-1)h, jk)) / h^2 = (u_{i+1,j} - 2u_{i,j} + u_{i-1,j}) / h^2$$

allora l'equazione del calore $\frac{\delta u}{\delta t} = D \frac{\delta^2 u}{\delta x^2}$ diventa :

$$(u_{i,j+1} - u_{i,j}) / k = D(u_{i+1,j} - 2u_{i,j} + u_{i-1,j}) / h^2$$

da cui, ponendo $p = Dk/h^2$:

$$\begin{aligned} u_{i,j+1} &= u_{i,j} + p(u_{i+1,j} - 2u_{i,j} + u_{i-1,j}) \\ u_{i,j+1} &= pu_{i-1,j} + (1 - 2p)u_{i,j} + pu_{i+1,j} \end{aligned} \quad (4)$$

Questa formula permette di trovare i valori di $u(x, t)$ all'istante $(j+1)k$ nel punto ih conoscendo quelli all'istante precedente jk nel punto ih , nel punto precedente $(i-1)h$ e nel punto successivo $(i+1)h$.

È chiaro che nel caso della diffusione libera basta conoscere i valori u_{i0} all'istante $t = 0$ per ogni valore intero di i per calcolare tutti i valori u_{i1} all'istante successivo $t = k$, poi da questi si calcolano tutti i valori u_{i2} all'istante $t = 2k$ e così via si calcolano tutti i valori u_{ij} per ogni j intero > 0 .

Però nei casi pratici non si possono considerare infiniti valori u_{i0} . Nel caso della diffusione con condizioni al contorno, non ci sono problemi: se ad esempio x può variare in un certo intervallo $[-a, a]$, si può fissare un numero intero positivo n (numero di passi in cui si vuole suddividere ciascuno dei 2 intervalli $[-a, 0]$ e $[0, a]$) e porre $h = a/n$. Per poter applicare la formula (4) bisogna conoscere non solo i valori u_{i0} (valori di $u(x, t)$ all'istante iniziale $t = 0$) per ogni i compreso tra $-n$ ed n , ma anche i valori degli estremi $u_{-n,k}$ e $u_{n,k}$ per ogni k intero > 0 (condizioni al contorno). In tal modo si

possono calcolare tutti i valori u_{ij} per ogni i compreso tra $-n$ ed n e per ogni $j > 0$.

La formula (4) ha un'interessante interpretazione in chiave probabilistica. Riferiamoci al caso della diffusione di una certa sostanza A e supponiamo che u_{ij} indichi il numero di molecole di A presenti all'istante $t = jk$ nel punto $x = ih$ (o meglio nell'intervallino $(ih - h/2, ih + h/2)$). Supponiamo anche di scegliere i passi spaziale e temporale h, k così piccoli che in un intervallo di tempo k una molecola che si trova nel punto ih abbia solo 3 possibilità:

1. Spostarsi a destra, andando nel punto $(i + 1)h$, con una certa probabilità p
2. Spostarsi a sinistra, andando nel punto $(i - 1)h$, con la stessa probabilità p
3. Restare dov' é, nel punto ih , con probabilità $1 - 2p$.

Cioé in ogni istante ogni molecola può fare un passo a destra o a sinistra o restare ferma, in modo puramente casuale, come un ubriaco che cammina avanti ed indietro a caso. Questo modo di muoversi si chiama *passeggiata aleatoria*. Naturalmente, questa é una rappresentazione molto schematica e semplificata del processo di diffusione, che però conduce alla soluzione esatta (come ora vedremo).

Ciò premesso, qual é il numero $u_{i,j+1}$ di molecole presenti nel punto $x = ih$ all'istante $t = (j + 1)k$? Si ha $u_{i,j+1} = n_1 + n_2 + n_3$ dove:

- $n_1 = pu_{i-1,j}$ = numero di molecole che all'istante jk si trovavano nel punto $(i - 1)h$ e hanno fatto un passo a destra (con probabilità p)
- $n_2 = (1 - 2p)u_{ij}$ = numero di molecole che all'istante jk si trovavano nel punto ih e sono rimaste lí (con probabilità $1 - 2p$)
- $n_3 = pu_{i+1,j}$ = numero di molecole che all'istante jk si trovavano nel punto $(i + 1)h$ e hanno fatto un passo a sinistra (con probabilità p).

Quindi, $u_{i,j+1} = pu_{i-1,j} + (1 - 2p)u_{ij} + pu_{i+1,j}$ cioè la (4).

Ma la (4) é proprio la formula che risolve in modo approssimato l'equazione del calore. Ciò significa che la rappresentazione del processo di diffusione come passeggiata aleatoria conduce proprio alla giusta equazione differenziale.

Osserviamo che p e $1 - 2p$ hanno il significato di probabilità, devono essere dei numeri compresi tra 0 e 1, quindi, p deve stare nell'intervallo $[0, 1/2]$. Ciò significa che i passi spaziale e temporale h, k non possono essere scelti in modo arbitrario, ma solo in modo tale che

$$p = Dk/h^2 \leq 1/2 \quad \text{da cui} \quad k \leq h^2/2D \quad (5)$$

Intuitivamente, sembra che facendo tendere h, k a zero l'approssimazione diventi sempre migliore, quindi, si presume che la soluzione approssimata data dalla (4) converga alla soluzione esatta dell'equazione del calore. Ma si può dimostrare che questo non è vero: dipende dalla velocità con cui h e k tendono a zero.

Una condizione necessaria per la *stabilità numerica* delle approssimazioni (cioè perché la soluzione approssimata converga alla soluzione esatta) è la (5). Se invece $k > h^2/2D$, allora, si hanno dei fenomeni di instabilità numerica: al tendere a zero di h, k la soluzione approssimata presenta delle oscillazioni sempre più forti, la cui ampiezza tende addirittura all'infinito (lo si vede facilmente al computer).

Siccome dev'essere $k \leq h^2/2D$, k deve tendere a zero almeno con la stessa rapidità di h^2 . Per ottenere buone approssimazioni conviene ad esempio, porre $p = Dk/h^2 = 1/4$, da cui $k = h^2/4D$. Ciò significa che prendendo il passo spaziale h 10 volte più piccolo, il passo temporale k deve essere preso ben 100 volte più piccolo (e naturalmente il tempo di calcolo aumenta a dismisura).

Esempio 15.1 DIFFUSIONE LIBERA DI MOLECOLE INIZIALMENTE CONCENTRATE IN UN PUNTO.

Consideriamo N molecole di una certa sostanza A (N numero intero positivo molto grande) e supponiamo che all'istante $t = 0$ esse siano tutte concentrate all'origine ($x = 0$) cioè:

$$u_{00} = N \quad u_{i0} = 0 \quad \text{per ogni } i \text{ diverso da } 0$$

Cosa succede all'istante $t = k$? pN molecole avranno fatto un passo a sinistra e quindi, si troveranno nel punto $x = -h$, pN molecole avranno fatto un passo a destra e quindi, si troveranno nel punto $x = h$ e le altre $(1 - 2p)N$ molecole saranno ancora in $x = 0$. Cioè:

$$u_{-1,1} = Np \quad u_{01} = N(1 - 2p) \quad u_{11} = Np$$

È facile capire che all'istante $t = 2k$ si avrà :

$$u_{-2,2} = u_{22} = Np^2 \quad u_{-1,2} = u_{12} = 2Np(1 - 2p) \quad u_{02} = N(2p^2 + (1 - 2p)^2)$$

(vedi fig. 15.6 nel caso $p = 1/2$). Al passare del tempo, le molecole si disperdono lungo l'asse x e la forma della loro distribuzione si avvicina sempre di più ad una curva gaussiana. Per capire meglio questo fatto consideriamo la variabile aleatoria $X_j =$ posizione di una generica molecola all'istante $t = jk$.

- Per $j = 0$ si ha che X_0 assume solo valore 0 con probabilità 1.

- Per $j = 1$ si ha che X_1 assume i 3 valori $-h, 0, h$ con probabilità $p, 1 - 2p, p$.
- Per $j = 2$ si ha che X_2 assume i 5 valori $-2h, -h, 0, h, 2h$ con probabilità $p^2, 2p(1 - 2p), 2p^2 + (1 - 2p)^2, 2p(1 - 2p), p^2$.

É facile verificare che $X_2 = 2X_1 = X_1 + X_1$ cioè X_2 é la somma di 2 variabili aleatorie indipendenti uguali a X_1 . Analogamente: $X_j = jX_1 = X_1 + X_1 + \dots + X_1$ (j volte).

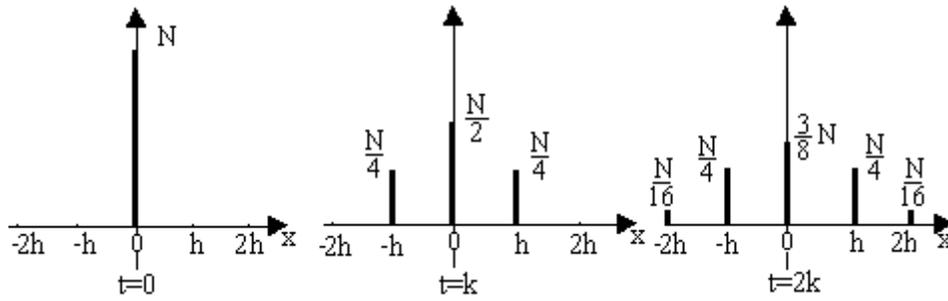


Figura 15.6

Per il teorema del limite centrale, la distribuzione di probabilità di X_j tende a quella di una variabile aleatoria con distribuzione normale (o gaussiana) di media 0 e varianza jV dove V indica la varianza di X_1 (si vede facilmente che $V = 2h^2p$). Quindi, X_j ha media 0, varianza $V_j = 2jh^2p$ e quindi, scarto quadratico medio $\sigma_j = h\sqrt{2pj}$.

Ricordiamo che la curva gaussiana di media m e scarto σ ha equazione:

$$\varphi(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}}$$

quindi, la distribuzione delle molecole al tempo $t = jk$ si ottiene ponendo $\sigma = h\sqrt{2pj}$, $m = 0$ ed é data da :

$$u(x, jk) = N\varphi(x) = \frac{N}{h\sqrt{4\pi pj}} e^{-\frac{x^2}{4pjh^2}}$$

Poiché $t = jk, p = Dk/h^2 = Dt/h^2j$, si trova per la funzione $u(x, t)$, l'espressione:

$$u(x, t) = \frac{N}{\sqrt{4\pi Dt}} e^{-\frac{x^2}{4Dt}} \quad (6)$$

Questa funzione rappresenta una gaussiana di media 0 e scarto $\sqrt{2Dt}$, che si allarga con velocità decrescente (proporzionale a $\sqrt{\frac{1}{t}}$). La curva gaussiana viene chiamata anche “curva della probabilità”, ma si potrebbe anche chiamare “curva della diffusione libera” o “curva della passeggiata aleatoria”. Il ragionamento fatto ci fa supporre che la (6) sia effettivamente una soluzione dell’equazione del calore, ma non ce ne dá la certezza. Per esserne certi basta semplicemente fare la verifica:

$$\frac{\delta u}{\delta x} = \frac{N}{\sqrt{4\pi Dt}} e^{-\frac{x^2}{4Dt}} \frac{-x}{2Dt}$$

$$\frac{\delta^2 u}{\delta x^2} = \frac{N}{\sqrt{4\pi Dt}} \left(e^{-\frac{x^2}{4Dt}} \frac{x^2}{4D^2 t^2} + e^{-\frac{x^2}{4Dt}} \frac{-1}{2Dt} \right) = \frac{N}{\sqrt{4\pi Dt}} e^{-\frac{x^2}{4Dt}} \frac{x^2 - 2Dt}{4D^2 t^2}$$

$$\frac{\delta u}{\delta t} = \frac{N}{\sqrt{4\pi D}} \left(-\frac{1}{2t\sqrt{t}} e^{-\frac{x^2}{4Dt}} + \frac{1}{\sqrt{t}} e^{-\frac{x^2}{4Dt}} \frac{x^2}{4Dt^2} \right) = \frac{N}{\sqrt{4\pi Dt}} e^{-\frac{x^2}{4Dt}} \left(-\frac{1}{2t} + \frac{x^2}{4Dt^2} \right)$$

$$\frac{\delta u}{\delta t} = \frac{N}{\sqrt{4\pi Dt}} e^{-\frac{x^2}{4Dt}} \frac{x^2 - 2Dt}{4Dt^2} = D \frac{\delta^2 u}{\delta x^2} \quad \text{come volevasi dimostrare}$$

■

SOLUZIONE GENERALE DELL’EQUAZIONE DEL CALORE NEL CASO DELLA DIFFUSIONE LIBERA

Le considerazioni fatte sulla passeggiata aleatoria ci permettono anche di trovare una formula generale che esprime la soluzione $u(x, t)$ dell’equazione del calore nel caso della diffusione libera, ponendo la condizione iniziale $u(x, 0) = f(x)$ per ogni numero reale x , essendo $f(x)$ una funzione assegnata. La formula é la seguente:

$$u(x, t) = \frac{1}{\sqrt{4\pi Dt}} \int_{-\infty}^{+\infty} e^{-\frac{(x-r)^2}{4Dt}} f(r) dr \quad (7)$$

Si può dimostrare che questa é la soluzione dell’equazione del calore semplicemente facendo la verifica, ma c’è anche un altro modo piú facile e piú interessante.

Anzitutto osserviamo che l’equazione del calore é lineare, quindi se u_1, u_2 sono soluzioni, anche $u_1 + u_2$ lo é. Poi disegniamo il grafico della funzione $y = f(x)$ (fig.15.7) e consideriamo sull’asse x un intervallino infinitesimo $[r, r + dr]$. L’area compresa tra il grafico e l’asse x in questo intervallino é $f(r)dr$ e rappresenta il numero di molecole che all’istante $t = 0$ si trovano

nell'intervallo. Queste molecole, al crescere di t , si diffondono lungo l'asse x e quindi, la loro distribuzione al tempo t che indichiamo con $u_r(x, t)$, é data dalla (6) con $x - r$ al posto di x e $f(r)dr$ al posto di N :

$$u_r(x, t)dr = \frac{f(r)dr}{\sqrt{4\pi Dt}} e^{-\frac{(x-r)^2}{4Dt}}$$

Ma se indichiamo con $f_r(x)$ la funzione che vale $f(x)$ nell'intervallo $[r, r + dr]$ e 0 fuori, possiamo dire che $f(x)$ é la somma di infinite funzioni del tipo $f_r(x)$ al variare di r da $-\infty$ a $+\infty$. Per linearitá, la soluzione $u(x, t)$ dell'equazione del calore che corrisponde alla condizione iniziale $u(x, 0) = f(x)$ sará la somma di tutte le $u_r(x, t)$ cioé:

$$u(x, t) = \int_{-\infty}^{+\infty} u_r(x, t)dr = \frac{1}{\sqrt{4\pi Dt}} \int_{-\infty}^{+\infty} e^{-\frac{(x-r)^2}{4Dt}} f(r)dr \quad \text{come volevasi dimostrare}$$

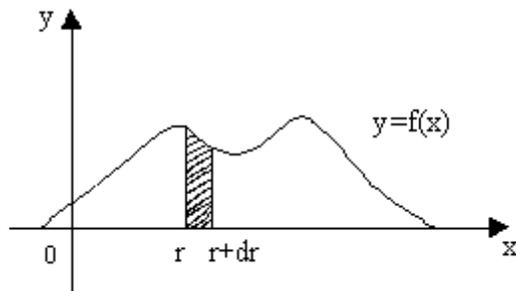


Figura 15.7

Osserviamo infine, che le formule (6) e (7) si riferiscono al problema della diffusione libera in una dimensione: supponiamo cioé che ci sia una sola dimensione spaziale (x), anziché $2(x, y)$ o 3 come in realtà (x, y, z). Ma delle formule analoghe si possono dare anche in 2 o in 3 dimensioni.

In 2 dimensioni l'equazione del calore é $\frac{\delta u}{\delta t} = D(\frac{\delta^2 u}{\delta x^2} + \frac{\delta^2 u}{\delta y^2})$, dove $u(x, y, t)$ é la funzione incognita. Il problema della diffusione libera di N molecole che al tempo $t = 0$ sono tutte concentrate nell'origine ha come soluzione la seguente funzione, analoga alla (6):

$$u(x, y, t) = N\varphi(x, t)\varphi(y, t) = \frac{N}{4\pi Dt} e^{-\frac{x^2+y^2}{4Dt}} \quad (8)$$

dove abbiamo posto

$$\varphi(x, t) = \frac{1}{\sqrt{4\pi Dt}} e^{-\frac{x^2}{4Dt}}$$

cioé $\Phi(x, t)$ é una gaussiana di media 0 e scarto $\sqrt{2Dt}$. Il secondo membro della (8) é una distribuzione gaussiana bidimensionale. Si suppone che le molecole possano muoversi indipendentemente nelle direzioni dei 2 assi x, y . Quindi, la probabilità che una molecola all'istante t si trovi nel punto (x, y) é uguale al prodotto $\Phi(x, t)\Phi(y, t)$ dove $\Phi(x, t)$ é la probabilità che l'ascissa della molecola si trovi in x e $\Phi(y, t)$ é la probabilità che la sua ordinata si trovi in y .

Nel caso piú generale di una condizione iniziale generica $u(x, y, 0) = f(x, y)$ la soluzione dell'equazione del calore é data dalla seguente formula, analoga alla (7):

$$u(x, y, t) = \frac{1}{4\pi Dt} \int \int_{\mathbb{R}^2} e^{-\frac{(x-r)^2 - (y-s)^2}{4Dt}} f(r, s) dr ds \quad (9)$$

Nel caso piú realistico delle 3 dimensioni spaziali x, y, z l'equazione del calore é $\frac{\delta u}{\delta t} = D(\frac{\delta^2 u}{\delta x^2} + \frac{\delta^2 u}{\delta y^2} + \frac{\delta^2 u}{\delta z^2})$ e la soluzione $u(x, y, z, t)$ é data nel caso delle N molecole che all'istante $t = 0$ sono tutte concentrate nell'origine, da:

$$u(x, y, z, t) = \frac{N}{\sqrt{(4\pi Dt)^3}} e^{-\frac{x^2 - y^2 - z^2}{4Dt}} = N\phi(x, t)\phi(y, t)\phi(z, t)$$

e nel caso generale della condizione iniziale $u(x, y, z, 0) = f(x, y, z)$ la soluzione é data da :

$$u(x, y, z, t) = \frac{N}{\sqrt{(4\pi Dt)^3}} \int \int \int_{\mathbb{R}^3} e^{-\frac{(x-r)^2 - (y-s)^2 - (z-w)^2}{4Dt}} f(r, s, w) dr ds dw$$

16 Zone di divieto di pesca

Consideriamo una popolazione di pesci in un ambiente ricco di cibo, che in assenza di pesca si riproduce con un tasso di crescita fisso $K > 0$. Indicando con $p(t)$ il numero totale di pesci al tempo t , la funzione $p(t)$ segue l'equazione differenziale $p'(t) = Kp(t)$ della crescita esponenziale, la cui soluzione é $p(t) = p_0 e^{Kt}$, dove $p_0 = p(0)$ rappresenta il numero di pesci all'istante iniziale $t = 0$.

Supponiamo ora che questi pesci vengano sottoposti ad una pesca molto intensa, che non dá loro il tempo di riprodursi e rischia di condannarli all'estinzione. Per evitare l'estinzione, viene stabilita una certa zona A di divieto di pesca, dove i pesci possono starsene tranquilli e proliferare indisturbati. Però i pesci, nuotando liberamente a caso in tutte le direzioni, possono uscire da A , e allora vengono immediatamente pescati. É un fenomeno analogo a quello della diffusione casuale di una certa sostanza X nell'ambiente dove al posto delle molecole di X si considerano i pesci.

Vogliamo risolvere il seguente problema: quanto grande deve essere A perché i pesci non si estinguano e come varia la "densità di pesci" $u(x, y, z, t)$ all'interno di A ? x, y, z rappresentano le 3 coordinate spaziali, t il tempo e $u(x, y, z, t)$ ha il significato di "numero di pesci per unità di volume all'istante t in un intorno molto piccolo del punto (x, y, z) ."

L'intuizione suggerisce che affinché i pesci non si estinguano le dimensioni di A devono essere superiori ad un certo valore critico, perché se A é troppo piccola ogni pesce nella sua "nuotata aleatoria" finirá per uscire da A molto presto, prima di riprodursi. Inoltre, si intuisce che la densità di pesci sará maggiore nelle zone centrali di A e minore vicino al bordo, perché in un piccolo volumetto $\Delta V \subset A$ che si trova sul bordo di A i pesci possono uscire subito da A ed essere pescati, mentre nuovi pesci possono entrare in ΔV solo dalla parte interna.

Il problema può essere risolto scrivendo la seguente equazione differenziale alle derivate parziali per la funzione incognita $u(x, y, z, t)$:

$$\frac{\partial u}{\partial t} = D \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \right) + Ku \quad (1)$$

con la condizione al contorno $u = 0$ in tutti i punti del bordo di A . Infatti, la velocità di variazione $\frac{\delta u}{\delta t}$ di u rispetto al tempo é la somma di un termine di diffusione $D \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \right)$ analogo a quello dell'equazione del calore (il coefficiente di diffusione D é proporzionale alla velocità con cui nuotano i pesci) e di un termine di riproduzione Ku analogo a quello della crescita esponenziale. Se $D = 0$ (cioé i pesci non nuotano, o comunque non si allontanano mai dal luogo in cui sono nati) si ottiene una pura e semplice crescita esponenziale senza diffusione, se $K = 0$ si ottiene un semplice processo di diffusione senza crescita della popolazione ed alla fine, i pesci si estinguono.

La (1) può essere risolta numericamente in modo approssimato, ma per poterla risolvere analiticamente (in formule) facciamo delle opportune ipotesi sulla geometria di A e sulla distribuzione iniziale dei pesci, che esemplificano molto la situazione.

Supponiamo che la zona di divieto di pesca A si trovi lungo la costa, all'interno di un golfo di forma esattamente rettangolare con lati di lunghezza l_1, l_2 e che la profondità del golfo sia h in tutti i punti (quindi, in 3 dimensioni il golfo è un parallelepipedo di lati l_1, l_2, h). Supponiamo anche che la zona di divieto di pesca A sia contenuta tutta nel golfo e si estenda fino ad una certa distanza a dal suo lato interno.

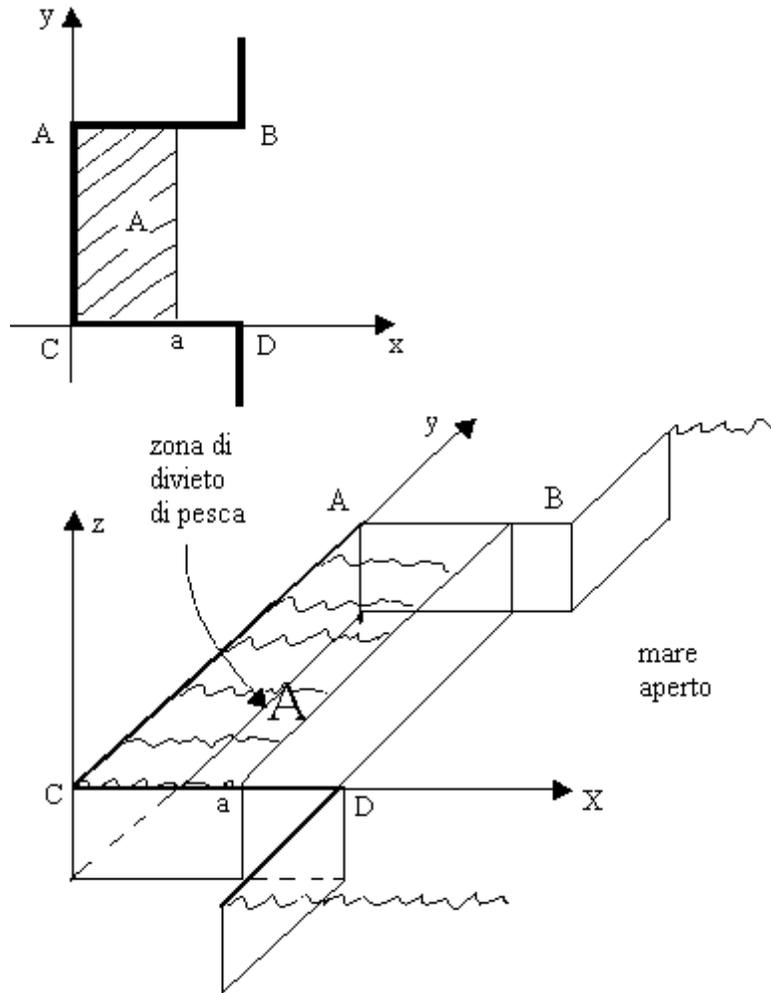


Figura 16.1 Figura 16.2

La figura 16.1 mostra il golfo visto in pianta, mentre la figura 16.2 é un disegno in 3 dimensioni. Poniamo l'origine delle coordinate nel punto C della costa, l'asse x coincidente col lato CD del golfo l'asse y coincidente col lato CA e l'asse z verticale (perpendicolare alla superficie del mare).

Facciamo inoltre un'ipotesi molto particolare sulla condizione iniziale al tempo $t = 0$: che $u(x, y, z, 0)$ dipenda solo da x e sia uguale ad una certa funzione $f(x)$ assegnata. Cioé la densità di pesci dev'essere la stessa per ogni profondità e per ogni posizione lungo l'asse y . Supponiamo cioè che

$$u(x, y, z, 0) = f(x)$$

con $f(x)$ definita nell'intervallo $[0, a]$, $f(a) = 0$ e $f(x) > 0$ per ogni $x \in [0, a)$. Questa condizione iniziale fa sí che all'istante $t = 0$ il flusso di pesci sia nullo nelle direzioni y, z perché in tali direzioni é nullo il "gradiente di densità" dei pesci. É vero che i pesci possono nuotare anche nelle direzioni y, z , ma in queste 2 direzioni ci saranno in media tanti pesci che nuotano in un senso quanti nel senso opposto, quindi, la densità di pesci non cambierà. Invece nella direzione dell'asse x c'è un flusso netto di pesci che va dai punti in cui $f(x)$ é piú grande ai punti in cui $f(x)$ é piú piccolo; inoltre molti pesci sfuggono dall'estremità libera di A ($x = a$). Queste considerazioni ci fanno capire che anche negli istanti successivi la funzione densità $u(x, y, z, t)$ in realtà dipenderá solo dalle 2 variabili x, t ; per semplicità scriveró $u(x, t)$ anziché $u(x, y, z, t)$. Il problema, che sembra un problema di diffusione in 3 dimensioni (x, y, z) , in realtà é un problema di diffusione in una sola dimensione (x) . Allora, l'equazione differenziale (1) diventa semplicemente:

$$\frac{\partial u}{\partial t} = D \left(\frac{\partial^2}{\partial x^2} + \right) + Ku \quad (2)$$

dove $x \in [0, a]$, $t \in \mathbb{R}$ e poniamo le seguenti condizioni iniziali:

- $u(x, 0) = f(x)$ per ogni $x \in [0, a]$
- $u(a, t) = 0$ per ogni t e $\frac{\delta u}{\delta x}(0, t) = 0$ per ogni t

(si può dimostrare che queste condizioni implicano l'esistenza e unicità della soluzione).

Nelle condizioni iniziali, non basta specificare $u(x, 0)$, bisogna anche stabilire che i pesci che escono in $x = a$ vengono subito pescati (questo si traduce in $u(a, t) = 0$) e che i pesci non possono attraversare l'estremo $x = 0$ perché c'è la costa (questo si traduce in $\frac{\delta u}{\delta x}(0, t) = 0$). Queste condizioni sono analoghe alle condizioni al contorno per il problema della diffusione del calore in una sbarretta di lunghezza a con l'estremo sinistro termicamente isolato e l'estremo destro tenuto a temperatura fissa $T = 0$. Anche l'equazione (2) é simile all'equazione del calore, però ha un termine in piú: il termine Ku che rappresenta la riproduzione dei pesci.

Ciò premesso, chiediamoci anzitutto delle soluzioni stazionarie dell'equazione (2), cioè delle soluzioni dove $u(x, t)$ non dipende dal tempo. Se tali soluzioni esistono, la popolazione dei pesci si mantiene costante nel tempo nonostante la pesca. Se $u(x, t)$ non dipende dal tempo deve essere $u(x, t) = f(x)$ per ogni t , quindi $\frac{\delta u}{\delta x}(0, t) = 0$ e la (2) diventa:

$$Df''(x) + Kf(x) = 0 \quad f''(x) + \frac{K}{D}f(x) = 0$$

questa è un'equazione differenziale lineare a coefficienti costanti nella funzione incognita $f(x)$ e la sua soluzione generale è :

$$f(x) = C_1 \cos \sqrt{\frac{K}{D}}x + C_2 \sin \sqrt{\frac{K}{D}}x$$

però la funzione $f(x)$ deve soddisfare anche le condizioni al contorno $f(a) = 0$, $f'(0) = 0$. Essendo:

$$f'(x) = -C_1 \sqrt{\frac{K}{D}} \sin \sqrt{\frac{K}{D}}x + C_2 \sqrt{\frac{K}{D}} \cos \sqrt{\frac{K}{D}}x$$

La condizione $f'(0) = 0$ diventa $C_2 \sqrt{\frac{k}{D}} = 0 \implies C_2 = 0$ e quindi:

$$f(x) = C_1 \cos \sqrt{\frac{K}{D}}x$$

ma allora, imponendo la condizione $f(a) = 0$ si trova:

$$c_1 \cos \sqrt{\frac{K}{D}}a = 0$$

Se $\cos \sqrt{\frac{K}{D}}a \neq 0$, questo implica $C_1 = 0$ e quindi, $f(x) = 0$, che non c'interessa perché $u(x, t) = f(x)$ non può essere la funzione identicamente nulla. Allora, dev'essere $\cos \sqrt{\frac{K}{D}}a = 0$, il che si verifica solo se $\sqrt{\frac{K}{D}}a$ è un multiplo dispari di $\frac{\pi}{2}$, cioè $\sqrt{\frac{K}{D}}a = (2n + 1)\frac{\pi}{2}$, con n intero ≥ 0 . Ne segue che:

$$a = (2n + 1)\frac{\pi}{2}\sqrt{\frac{D}{K}} \quad f(x) = C_1 \cos \sqrt{\frac{K}{D}}x \quad \text{con } C_1 \text{ qualsiasi}$$

Però $f(x)$ si annulla nei punti $\frac{\pi}{2}\sqrt{\frac{D}{K}}, \frac{3\pi}{2}\sqrt{\frac{D}{K}}, \frac{5\pi}{2}\sqrt{\frac{D}{K}}, \dots$ e ha segno opposto nei 2 intervalli $[0, \frac{\pi}{2}\sqrt{\frac{D}{K}}]$ e $[\frac{\pi}{2}\sqrt{\frac{D}{K}}, \frac{3\pi}{2}\sqrt{\frac{D}{K}}]$, quindi, in realtà, a dev'essere

uguale a $\frac{\pi}{2}\sqrt{\frac{D}{K}}$ (cioé $n = 0$), perché altrimenti, nell'intervallo $[0, a]$, $f(x)$ potrebbe essere negativa, assurdo perché $f(x)$ rappresenta la densità di pesci. Se ne conclude che possono esistere delle soluzioni stazionarie della (2) solo se a assume un preciso valore critico:

$$a = \frac{\pi}{2}\sqrt{\frac{D}{K}} \quad (3)$$

in tal caso la soluzione stazionaria é data da $u(x, t) = C_1 \cos \sqrt{\frac{K}{D}}x$, con $C_1 > 0$ qualsiasi e quindi nell'intervallo $[0, a]$, $u(x, t)$ ha la forma di un "quarto d'onda" sinusoidale (fig.16.3).

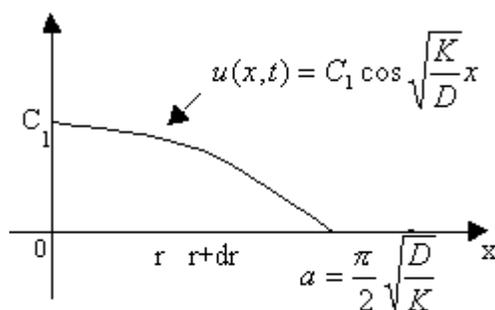


Figura 16.3

Il valore di a dato dalla (3) rappresenta la minima larghezza che deve avere la zona di divieto di pesca affinché i pesci non si estinguano. Infatti, si può dimostrare il seguente:

Teorema 16.1 1. Se $a < \frac{\pi}{2}\sqrt{\frac{D}{K}}$, allora $\lim_{t \rightarrow +\infty} u(x, t) = 0$ per ogni $x \in [0, a]$ e quindi, i pesci si estinguono sempre.

2. Se $a = \frac{\pi}{2}\sqrt{\frac{D}{K}}$ e $u(x, 0)$ non é della forma $C_1 \cos \sqrt{\frac{K}{D}}x$, allora esiste $C_1 > 0$ tale che $\lim_{t \rightarrow +\infty} u(x, t) = C_1 \cos \sqrt{\frac{K}{D}}x$ per ogni $x \in [0, a]$, cioè la soluzione della (2) per $t \rightarrow +\infty$ tende asintoticamente a una soluzione stazionaria che ha la forma di un quarto d'onda sinusoidale.

3. Se $a > \frac{\pi}{2}\sqrt{\frac{D}{K}}$, allora $\lim_{t \rightarrow +\infty} u(x, t) = +\infty$ per ogni $x \in [0, a]$, cioè la popolazione dei pesci cresce senza limiti.

Osservazione 16.1 1. In realtà se $a > \frac{\pi}{2}\sqrt{\frac{D}{K}}$ la popolazione dei pesci non può crescere oltre ogni limite perché le risorse di cibo sono limitate:

essa tenderá ad un valore limite di saturazione $S(x)$ che dipende da X , ma per studiare la forma della funzione $S(x)$ bisognerebbe modificare il modello sostituendo il termine Ku (crescita esponenziale) con un termine del tipo $Ku(M - u)$ (crescita logistica).

2. Considerazioni analoghe a quelle che abbiamo fatto valgono anche se la zona di divieto di pesca A ha una forma qualsiasi. In tal caso bisogna imporre la condizione al contorno $u = 0$ in tutti i punti in cui A confina col mare aperto, mentre nei punti della costa bisogna imporre che la derivata spaziale di u sia nulla nella direzione perpendicolare alla costa.
3. Un problema analogo a quello della zona di divieto di pesca é quello della “eutrofizzazione” di organismi come batteri, alghe, plancton. Se in una zona A di mare esistono delle condizioni favorevoli (temperatura, sostanze nutritive) per un rapido sviluppo di certe alghe, all’interno di A le alghe si riproducono rapidamente, ma quando escono da A per un moto casuale di diffusione (dovuto ad esempio al vento, alle onde . . .) non si riproducono piú e vanno perse. Si puó dimostrare però che la zona A deve avere dimensioni superiori ad un certo valore critico, altrimenti il fenomeno non puó avere luogo.

■

Dimostrazione. La dimostrazione é piuttosto complessa. Il punto essenziale consiste nel cercare una soluzione $u(x, t)$ della (2) che sia del tipo $u(x, t) = y(x)g(t)$, cioè sia il prodotto di una funzione della sola x per una funzione della sola t (metodo di separazione delle variabili). Si ha allora:

$$\frac{\partial u}{\partial t} = y(x)g'(t) \quad \frac{\partial u}{\partial x} = y'(x)g(t) \quad \frac{\partial^2 u}{\partial x^2} = y''(x)g(t)$$

e quindi, la (2) diventa:

$$y(x)g'(t) = Dy''(x)g(t) + Ky(x)g(t)$$

e dividendo ambo i membri per $u(x, t) = y(x)g(t)$:

$$\frac{g'(t)}{g(t)} = D \frac{y''(x)}{y(x)} + K$$

ma il primo membro é funzione della sola t e il secondo membro é funzione della sola x , e questo é possibile solo se entrambi sono uguali ad una certa costante C (per ora sconosciuta). Si ottengono quindi, 2 diverse equazioni differenziali, una per $g(t)$ e una per $y(x)$:

$$g'(t) = Cg(t) \tag{4}$$

$$D \frac{y''(x)}{y(x)} + K = C \quad y''(x) = \frac{C - K}{D} y(x) \quad (5)$$

La soluzione generale della (4) é :

$$g(t) = C_1 e^{Ct} \quad \text{con } C_1 \text{ costante qualsiasi}$$

mentre la(5) é un'equazione differenziale lineare a coefficienti costanti del secondo ordine, la cui soluzione é diversa a seconda che sia $C > K, C = K, C < K$.

1. PRIMO CASO: $C > K$.

Allora, la soluzione generale della (5) é:

$$y(x) = C_2 e^{\sqrt{\frac{C-K}{D}}x} + C_3 e^{-\sqrt{\frac{C-K}{D}}x}$$

peró le condizioni iniziali $u(a, t) = 0$ e $\frac{\delta u}{\delta x}(0, t) = 0$ implicano:

$$y(a)g(t) = 0 \text{ per ogni } t \longrightarrow y(a) = 0$$

$$y'(0)g(t) = 0 \text{ per ogni } t \longrightarrow y'(0) = 0$$

essendo $y'(x) = C_2 \sqrt{\frac{C-K}{D}} e^{\sqrt{\frac{C-K}{D}}x} - C_3 \sqrt{\frac{C-K}{D}} e^{-\sqrt{\frac{C-K}{D}}x}$ la condizione $y'(0) = 0$ implica $C_2 = C_3$, da cui:

$$y(x) = C_2 \left(e^{\sqrt{\frac{C-K}{D}}x} + e^{-\sqrt{\frac{C-K}{D}}x} \right)$$

essendo la quantità in parentesi > 0 per ogni x , può essere $y(a) = 0$ solo se $C_2 = C_3 = 0$, quindi $y(x)$ e $u(x, t)$ sarebbero identicamente nulle, perciò questo caso non ci interessa.

2. SECONDO CASO: $C = K$.

Allora la (5) diventa $y''(x) = 0$ e la sua soluzione generale é $y(x) = C_2 x + C_3$, da cui $y'(x) = C_2$. Dovendo essere $y'(0) = 0$, si ha $C_2 = 0$ e quindi, $y(x) = C_3$, ma allora, dovendo essere $y(a) = 0$ si ha $C_3 = 0$ e quindi, $y(x)$ e $u(x, t)$ sono identicamente nulle. Perció anche questo caso non ci interessa.

3. TERZO CASO: $C < K$.

Allora, la soluzione generale della (5)é:

$$y(x) = C_2 \cos \sqrt{\frac{K-C}{D}}x + C_3 \sin \sqrt{\frac{K-C}{D}}x$$

da cui

$$y'(x) = -C_2 \sqrt{\frac{K-C}{D}} \sin \sqrt{\frac{K-C}{D}} x + C_3 \sqrt{\frac{K-C}{D}} \cos \sqrt{\frac{K-C}{D}} x$$

imponendo $y'(0) = 0$, si trova $C_3 = 0 \rightarrow y(x) = C_2 \cos \sqrt{\frac{K-C}{D}} x$ e

quindi, imponendo $y(a) = 0$ si trova $C_2 \cos \sqrt{\frac{K-C}{D}} a = 0$.

Se $\cos \sqrt{\frac{K-C}{D}} a \neq 0$, questo implica $C_2 = 0 \rightarrow u(x, t)$ identicamente nulla e questo caso non ci interessa, quindi l'unico caso interessante è $\cos \sqrt{\frac{K-C}{D}} a = 0$, cioè $\sqrt{\frac{K-C}{D}} a = (2n+1)\frac{\pi}{2}$ (n intero ≥ 0). Da qui si ricava il valore della costante sconosciuta C :

$$\frac{K-C}{D} = \frac{(2n+1)^2 \pi^2}{4a^2} \quad K-C = \frac{D(2n+1)^2 \pi^2}{4a^2}$$

$$C = K - \frac{D(2n+1)^2 \pi^2}{4a^2}$$

Se C ha questo valore, la (5) ha soluzione generale:

$$y(x) = C_2 \cos \sqrt{\frac{K-C}{D}} x = C_2 \cos \frac{(2n+1)\pi}{2a} x \quad \text{con } C_2 \text{ qualsiasi}$$

e la (4) ha soluzione generale

$$g(t) = C_1 e^{Ct} = C_1 e^{\left(K - \frac{D(2n+1)^2 \pi^2}{4a^2}\right)t} \quad \text{con } C_1 \text{ qualsiasi}$$

quindi, la soluzione non identicamente nulla $u(x, t)$ dell'equazione originaria (2) del tipo $u(x, t) = y(x)g(t)$ è data da

$$u(x, t) = C_3 e^{\left(K - \frac{D(2n+1)^2 \pi^2}{4a^2}\right)t} \cos \frac{(2n+1)\pi}{2a} x \quad (6)$$

con C_3 qualsiasi ($C_3 = C_1 C_2$). Questa soluzione della (2) però non verifica la condizione iniziale $u(x, 0) = f(x)$ con $f(x)$ assegnata. Infatti, essa verifica la condizione iniziale $u(x, 0) = C_3 \cos \frac{(2n+1)\pi}{2a} x$, cioè $u(x, 0)$ è del tipo coseno di un multiplo dispari di $\frac{\pi x}{2a}$.

A questo punto osserviamo che l'equazione differenziale originaria (2) è lineare, quindi se $u_1(x, t), u_2(x, t)$ sono soluzioni e λ, μ sono numeri reali qualsiasi, anche $\lambda u_1 + \mu u_2$ è soluzione.

Ora se noi riusciamo a sviluppare la funzione $f(x)$ nell'intervallo $[0, a]$

in una serie di Fourier di coseni di un multiplo dispari di $\frac{\pi x}{2a}$, cioè in una serie del tipo

$$f(x) = A_1 \cos \frac{\pi x}{2a} + A_3 \cos \frac{3\pi x}{2a} + A_5 \cos \frac{5\pi x}{2a} + \dots$$

allora, per linearità la soluzione della (2) tale che $u(x, 0) = f(x)$ sarà data dalla seguente serie:

$$u(x, t) = A_1 e^{\left(K - \frac{D\pi^2}{4a^2}\right)t} \cos \frac{\pi x}{2a} + A_3 e^{\left(K - \frac{9D\pi^2}{4a^2}\right)t} \cos \frac{3\pi x}{2a} + A_5 e^{\left(K - \frac{25D\pi^2}{4a^2}\right)t} \cos \frac{5\pi x}{2a} \dots \quad (7)$$

Resta da vedere se è sempre possibile sviluppare $f(x)$ in $[0, a]$ in una serie di Fourier di soli “coseni dispari”. Questo è possibile con un trucco: prolungare $f(x)$ dall’intervallo $[0, a]$ all’intervallo $[-2a, 2a]$ costruendo una funzione $g(x)$ così definita:

$$g(x) = \begin{cases} f(x) & \text{in } [0, a] \\ -f(2a - x) & \text{in } [a, 2a] \\ f(-x) & \text{in } [-a, 0] \\ -f(2a + x) & \text{in } [-2a, -a] \end{cases}$$

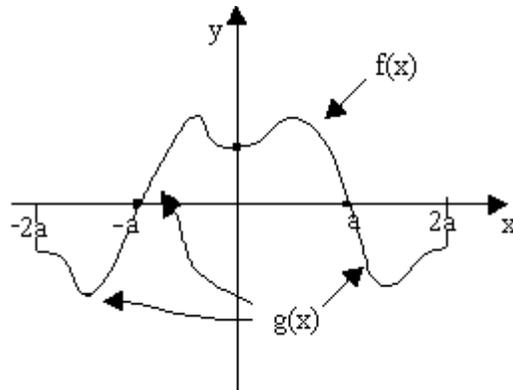


Figura 16.4

(il grafico di $g(x)$ è simmetrico rispetto all’asse y ed inoltre nell’intervallo $[0, 2a]$ è simmetrico rispetto al punto $[0, a]$ = fig.16.4.)

Quando si sviluppa $g(x)$ in serie di Fourier si trova una serie che contiene termini del tipo $A_n \cos \frac{n\pi x}{2a}$ e $B_n \sin \frac{n\pi x}{2a}$, ma è facile vedere che $B_n = 0$ e che $A_n = 0$ se n è pari.

Infatti (a meno di una costante moltiplicativa) B_n é dato da $\int_{-2a}^{2a} g(x) \sin \frac{n\pi x}{2a} dx$ e l'integrale vale 0 essendo $g(-x) = g(x)$.

Inoltre (a meno di una costante moltiplicativa) A_n é dato da $\int_{-2a}^{2a} g(x) \cos \frac{n\pi x}{2a} dx$ e se n é pari, l'integrale vale 0 perché la funzione integranda assume valori opposti in $[0, a]$ e in $[0, 2a]$, come pure in $[-a, 0]$ e in $[-2a, -a]$.

Concludendo, abbiamo trovato che la soluzione dell'equazione differenziale originaria (2) con le condizioni iniziali specificate é data dalla formula (??) dove A_1, A_3, A_5, \dots sono i coefficienti di Fourier dello sviluppo di $g(x)$ in $[-2a, 2a]$. A questo punto possiamo distinguere 3 casi, a seconda che a superi o meno il valore critico $\frac{\pi}{2} \sqrt{\frac{D}{K}}$:

1. $a < \frac{\pi}{2} \sqrt{\frac{D}{K}}$

Allora si ha $a^2 < \frac{\pi^2 D}{4K}$ cioè $K < \frac{\pi^2 D}{4a^2}$ e quindi, gli esponenti degli esponentiali nella formula (??) sono tutti < 0 . Perciò $\lim_{t \rightarrow +\infty} u(x, t) = 0$ per ogni $x \in [0, a]$. Quindi, i pesci si estinguono.

2. $a = \frac{\pi}{2} \sqrt{\frac{D}{K}}$, cioè $K = \frac{\pi^2 D}{4a^2}$.

In tal caso il primo termine della formula (7) diventa $A_1 \cos \frac{\pi x}{2a}$ e negli altri termini l'esponente tende a 0 per $t \rightarrow +\infty$. Quindi $\lim_{t \rightarrow +\infty} u(x, t) = A_1 \cos \frac{\pi x}{2a}$ per ogni $x \in [0, a]$.

Allora per $t \rightarrow +\infty$, $u(x, t)$ tende ad una soluzione stazionaria a forma di quarto d'onda sinusoidale, che non é altro che il primo termine dello sviluppo di Fourier di $g(x)$.

3. $a > \frac{\pi}{2} \sqrt{\frac{D}{K}}$ cioè $K > \frac{\pi^2 D}{4a^2}$.

In tal caso c'è un numero finito di termini della serie (??) in cui l'esponente dell'esponenziale é positivo, mentre nei termini successivi é negativo. Comunque per $t \rightarrow +\infty$ il termine dominante (che tende all'infinito piú rapidamente) é sempre il primo.

Quindi, $\lim_{t \rightarrow +\infty} u(x, t) = +\infty$ per ogni $x \in [0, a]$, perciò i pesci aumentano di numero indefinitamente. Inoltre per t molto grande la forma del grafico di $u(x, t)$ rispetto a x é molto simile a quella del primo termine della (??), cioè un quarto d'onda sinusoidale.

Ciò completa la dimostrazione.

■