

MODELLI E INTERPOLAZIONI

Senza alcuna pretesa di completezza, vogliamo mostrare come gli strumenti del calcolo che abbiamo sviluppato e in particolare le funzioni elementari introdotte nelle dispense [EC] e [ED] intervengano per costruire *modelli* di “sistemi” della realtà (fisica, economica, biologica, ...). In modo un po’ grossolano ma abbastanza pertinente possiamo dire che tutti questi sistemi hanno delle caratteristiche in comune: un sistema reale consiste in diverse grandezze, X, Y, \dots , che interagiscono tra loro. Per ogni grandezza è fissata un’unità di misura. Per ogni “stato” del sistema si possono misurare (almeno in modo approssimato) le grandezze in gioco, ottenendo così un certo insieme di numeri reali x, y, \dots . In generale si tratta di grandezze vettoriali, ma ai fini di questa discussione ci restringiamo a grandezze scalari. Si cerca allora di determinare le “leggi” che governano le relazioni esistenti nella realtà tra queste grandezze in termini delle rispettive misure. Di solito ci sono due tipi di “legge”: *dinamico* in cui si tende a descrivere come, a partire da un certo stato iniziale, il sistema occupi in seguito altri stati; *statico* i cui si determinano le relazioni tra le grandezze in un dato stato, specialmente in uno *stato di equilibrio* nel quale il sistema tende a permanere (in assenza di agenti esterni). Almeno nei casi più semplici, tali leggi hanno una forma del tipo $y = f(x)$ cioè esprimono il fatto che certe grandezze sono funzione delle altre. E’ importante capire che questo è solo un *modello teorico* del nostro sistema reale, che sarà corroborato (o meno) a seconda dell’accordo tra la previsione teorica e le reali misure in un arbitrario stato del sistema. Un modello teorico $y = f(x)$ comporta due aspetti, uno di natura *qualitativa* che prescrive solo la “forma” della funzione f , uno di natura *effettiva* che definisce esattamente chi è f . Questa distinzione sarà chiarita in seguito. A parte il necessario accordo con i dati sperimentali già noti, in buona misura la scelta di un modello teorico è spesso arbitraria e dettata anche da fattori e “pregiudizi” di ordine psicologico, estetico, ideologico, religioso ... (“le leggi della natura sono dettate da Dio che è perfetto e dunque devono essere espresse in termini di enti perfetti” - è “evidente” per esempio che le circonferenze sono le ellissi “perfette” ...). In ogni caso il modello viene corroborato (o meno) a posteriori come già detto.

1. GERARCHIA DELLE FUNZIONI ELEMENTARI E DEI MODELLI

Un “pregiudizio” o se preferiamo un “criterio” largamente condiviso è che un modello teorico debba essere il più possibile “semplice”. Limitandoci, come faremo, a modelli del tipo $y = f(x)$, questo si traduce in una richiesta di semplicità della funzione f . Allora possiamo intanto convenire che f debba essere individuata tra le funzioni *elementari derivabili*, cioè $f \in \mathcal{E}'$. Osserviamo poi che l’insieme di funzioni \mathcal{E}' può essere organizzato in una gerarchia di sottoclassi di *complessità crescente*. Senza formalizzare troppo questa idea, limitiamoci alle seguenti osservazioni:

- Tra le funzioni “fondamentali” le costanti sono chiaramente le meno complicate; a seguire le inclusioni $x \rightarrow x$ (in particolare l’identità di \mathbb{R}). Seguono poi la funzione esponenziale e il seno. La definizione dell’esponenziale è stata abbastanza più laboriosa ([Reali], [EC], [D-Exp],...) dunque possiamo postulare che il seno viene prima dell’esponenziale.
- Una misura della complessità di una funzione $f \in \mathcal{E}'$ è certamente data dal numero minimo di procedure \mathbf{P}' necessarie per costruire f a partire dalle funzioni fondamentali.

Combinando in una specie di ordinamento *lessicografico* i due punti precedenti, otteniamo quella gerarchia di sottoclassi che abbiamo sopra menzionato. E’ chiaro allora che la prima grossa sottoclasse è quella delle *funzioni polinomiali* che è a sua volta organizzata in sottoclassi secondo il *grado crescente* dei polinomi. Poi seguono le *funzioni razionali*, ad un certo punto della gerarchia troviamo la classe delle *funzioni polinomiali trigonometriche* in cui ogni monomio si ottiene componendo ogni monomio ordinario con funzioni seno o coseno, ect ...

2. MODELLI DI TIPO LINEARE E QUADRATICO

Applicando il criterio di semplicità, i modelli teorici $y = f(x)$ ($f \in \mathcal{E}'$) *qualitativamente* più semplici sono quelli di *tipo lineare*, cioè della forma

$$y = f(x) = mx + q$$

per qualche costante $m, q \in \mathbb{R}$. In tal caso il grafico della funzione è una retta in \mathbb{R}^2 di coefficiente angolare $m = \tan(\alpha)$, passante per il punto $(0, q)$. Una volta che avessimo postulato questa forma qualitativa del nostro modello, per fissarlo effettivamente dovremmo specificare la coppia (m, q) .

Esempio 2.1. La legge di Newton

$$a = F/m$$

che governa il moto di un punto materiale di massa m vincolato a muoversi lungo una retta, è di tipo lineare e esprime l'accelerazione del punto in funzione della forza applicata al punto. La cosa si fa più complicata se interpretiamo $a = a(t)$ come derivata seconda della posizione $x(t)$ del punto in funzione del tempo. Entriamo qui nel campo delle equazioni differenziali, restando comunque in un contesto "lineare". L'equazione di stato dei gas ideali è un altro esempio:

$$p = \frac{nR}{V}T$$

dove, essendo volutamente vaghi sulle unità di misura in gioco, V indica il volume fissato occupato dal gas, n è il numero fissato di *moli* del gas, R è la costante universale dei gas perfetti, T è la temperatura del gas, p è la pressione. Dunque la legge è un modello di tipo lineare che esprime la pressione in funzione della temperatura.

I modelli immediatamente più complicati sono quelli con legge *quadratica*, cioè del tipo:

$$y = f(x) = ax^2 + bx + c$$

per qualche costante $a, b, c \in \mathbb{R}$, $a \neq 0$. Sempre in riferimento al movimento del punto materiale,

$$e = \frac{1}{2}mv^2$$

è un esempio di legge quadratica che esprime l'energia cinetica in funzione della velocità del punto. Come è ben noto il grafico di f è una parabola. Il grafico di $-f$ si ottiene semplicemente per riflessione rispetto all'asse delle ascisse, dunque non è restrittivo assumere $a > 0$. Possiamo analizzare f con gli strumenti del calcolo differenziale. Allora, $f'(x) = 2ax + b$ e si annulla solo quando $x = -b/2a$; la derivata è < 0 se $x < -b/2a$, è > 0 se $x > -b/2a$. Dunque $-b/2a$ è un punto di minimo assoluto per f . Il grafico è simmetrico rispetto alla retta verticale $\{(x, y); x = -b/2a\}$ ed incontra l'asse delle y nel punto $y = c$. Il grafico incontra l'asse delle x solo se $\Delta = b^2 - 4ac \geq 0$. La derivata seconda $f''(x) = 2a > 0$, dunque la funzione è convessa.

2.1. Interpolazione lineare - metodo dei minimi quadrati. *Questo paragrafo è piuttosto rivolto ad un lettore particolarmente interessato.*

Supponiamo di avere a che fare con un dato sistema reale per il quale abbia senso cercare un modello teorico della forma $y = f(x)$. Supponiamo di avere qualche motivo (tra cui il solito pregiudizio/criterio di semplicità) per congetturare che qualitativamente il modello sia di tipo lineare:

$$y = mx + q .$$

Vogliamo allora corroborare (o meno) sperimentalmente questa ipotesi qualitativa e (nel caso fosse confermata) determinare con buona approssimazione l'effettiva legge (cioè la coppia (m, q) di "costanti di struttura" del sistema). Supponiamo inoltre di sapere preparare l'esperienza in modo tale che le misure x siano determinate con un errore trascurabile, mentre non possiamo trascurare la presenza di errori sperimentali (e sistematici) nella misure y . Quello che possiamo fare è allora produrre sperimentalmente un insieme di coppie di misure (x_j, y_j) , $j = 1, \dots, n$. E' chiaro che per essere significativo n deve essere "abbastanza grande", anche se per adesso non è chiaro cosa significhi "abbastanza"; d'altra parte ci sono dei vincoli pratici per cui n non può essere davvero grande a piacere.

Se il modello qualitativo fosse corretto, se conoscessimo effettivamente la legge $y = f(x) = mx + q$ e fossimo in grado di effettuare misure esatte, allora si avrebbe che le coppie (x_j, y_j) sarebbero allineate, appartenendo al grafico rettilineo di f . Poiché non possiamo escludere errori nelle misure y_j , anche conoscendo effettivamente f , l'errore che misura la distanza tra il dato sperimentale e il dato teorico è per ogni j

$$d_j = mx_j + q - y_j .$$

Prendiamo come *misura globale dell'errore*:

$$S(a, b) = \frac{1}{n} \sum_j d_j^2 .$$

Si noti abbiamo preso la *media aritmetica dei quadrati degli errori individuali* piuttosto che la media aritmetica degli errori stessi. Ci sono varie ragioni per fare questo: intanto errori di segno opposto (anche di valore assoluto grande) possono cancellarsi nella media aritmetica, mentre questo non succede prendendo i quadrati; inoltre con i quadrati si minimizza il contributo degli errori sperimentali (che possiamo cercare di rendere piccoli) mentre amplifica gli eventuali errori "grandi" dovuti ad una possibile inadeguatezza *sistematica* non eliminabile del nostro modello. Ricordiamoci ora che in effetti la retta teorica è *incognita*, mentre il numero $S(m, q)$ è definito per ogni retta. E' ragionevole pensare che la retta teorica che cerchiamo (ammesso che esista) sia approssimata bene dalla retta che *minimizza* $S(m, q)$. Si tratta quindi di un problema di ricerca di punto di minimo assoluto per una funzione di due variabili. Vediamo come questo specifico problema può essere ricondotto alla ricerca dei punti di minimo di due determinate funzioni quadratiche di una sola variabile, cosa che sappiamo fare facilmente come abbiamo visto sopra. Per enunciare i prossimi risultati, conviene introdurre alcune notazioni:

Notazioni Date due sequenze finite di numeri reali $x = (x_1, \dots, x_n)$, $y = (y_1, \dots, y_n)$, indichiamo con \bar{x} la media aritmetica degli x_j , \bar{xy} la media aritmetica dei prodotti $x_j y_j$, $\bar{x^2}$, la media aritmetica dei punti x_j^2 .

Procediamo allora nel modo seguente:

- Per ogni m fissato, calcoliamo $S(m, q)$ pensata come funzione della sola variabile q . Con calcoli semplici che non riportiamo, si verifica allora che è una funzione quadratica della forma:

$$S(m, q) = q^2 + bq + c$$

dove i coefficienti b, c possono essere calcolati esplicitamente, dipendono dagli x_j, y_j ed anche dal parametro m . Dunque, per ogni m , esiste un unico punto di minimo $q_0 = q_0(m) = -b/2$, che alla fine risulta essere

$$q_0(m) = \bar{y} - m\bar{x} .$$

- Per ogni m , calcoliamo $S(m, q_0(m))$. Omettendo ancora una volta i dettagli, concludiamo che anche questa è una funzione quadratica di m , della forma:

$$S(m, q_0(m)) = am^2 + hm + k$$

dove i coefficienti a, h, k possono essere calcolati esplicitamente in funzione dei soli x_j, y_j , ed inoltre $a > 0$. Allora anche $S(m, q_0(m))$ ha un unico punto di minimo locale $m_0 = -h/2a$ che alla fine risulta essere:

$$m_0 = \frac{\bar{xy} - \bar{x}\bar{y}}{\bar{x^2} - \bar{x}^2} .$$

Infine il punto di minimo assoluto per $S(m, q)$ cercato è dato dalla coppia

$$(m_0, q_0(m_0)) .$$

Abbiamo trovato così questa retta $y = m_0 x + q_0(m_0)$ che viene detta *la retta di regressione* dei nostri dati sperimentali (x_j, y_j) . Si dice anche che tale retta è ottenuta per *interpolazione lineare* di quei dati.

Resta il problema di valutare quanto questa sia una buona approssimazione della retta teorica incognita (ammesso che esista) e, in ultima analisi, se il modello teorico lineare stesso sia adeguato al nostro sistema oppure no. Non è qui il luogo di sviluppare i dettagli di questo argomento (probabilmente questi temi saranno ripresi in altri corsi del corso di laurea). Ci limitiamo a dare il risultato finale. Viene derivato il *coefficiente di Pearson*

$$CP = \frac{\bar{x}\bar{y} - \bar{x}\bar{y}}{\sqrt{(\bar{x}^2 - \bar{x}^2)(\bar{y}^2 - \bar{y}^2)}}$$

che ha il seguente significato: se $|CP|$ è “abbastanza” vicino a 1 (per esempio dell’ordine di 0,995) e n è “abbastanza” grande (per esempio $n \geq 5$ meglio $n \geq 6$) allora si può ritenere che il modello teorico lineare è adeguato e che la retta di regressione è una buona approssimazione della retta teorica. Se CP è vicino allo zero, allora il modello teorico “lineare” va rigettato. Si noti che il modo in cui abbiamo specificato i due “abbastanza” ha carattere prevalentemente *empirico*, viene abitualmente adottato (seguendo protocolli internazionali) per esempio per molte analisi chimiche, ma non ha carattere universale. Inoltre ci sono molti altri accorgimenti per migliorare l’attendibilità del risultato. Per esempio, spesso per ogni x_j si fanno almeno 3 misurazioni diverse per y e si prende poi come y_j la media dei valori ottenuti. Un altro punto delicato è la distribuzione dei valori degli x_j . Si potrebbe per esempio prendere una distribuzione uniforme in cui questi si susseguono uno dopo l’altro a distanza d fissa (e ci sarebbe comunque da “giustificare” la scelta di d). In altri casi (questo è soprattutto vero per l’*interpolazione esponenziale* di cui parleremo poi) conviene scegliere la posizione degli x_j in modo non uniforme, con opportuni “addensamenti” dettati dalle caratteristiche proprie del sistema reale preso in considerazione.

3. MODELLI DI TIPO SINUSOIALE O PERIODICO

Un modello di *tipo sinusoidale* è della forma

$$y = f(x) = A \cos(\omega(x - x_0)) + y_0$$

dove i parametri $A, y_0, x_0 \in \mathbb{R}$, $A > 0$. A meno di comporre con una opportuna traslazione $x \rightarrow x + r$, può essere riscritto in forma analoga, con la funzione sin al posto di cos. I i parametri determinano completamente f , ed hanno ciascuno un significato preciso. L’immagine della funzione f è l’intervallo $[y_0 - A, y_0 + A]$. Quindi l’*ampiezza* delle oscillazioni (positive o negative) del grafico di f è A , mentre y_0 è il *valor medio*. La funzione f è periodica di *periodo*

$$P = \frac{2\pi}{\omega},$$

$$\nu := 1/P$$

è detta la *frequenza* di f , mentre

$$\omega = 2\pi\nu = \frac{2\pi}{P}$$

è la *frequenza angolare*. L’effetto di x_0 è quello di traslare le ascisse. Ne segue che i punti di massimo di f sono della forma $x_0 + kP$, $k \in \mathbb{Z}$. Grazie alla periodicità di f , la funzione non cambia se sostituiamo x_0 prendendolo uguale al *primo punto di massimo non negativo*; scelto con questa convenzione, x_0 è detto la *fase* di f .

Molti sistemi reali che presentano comportamenti “ondulatori”, “oscillatori”, “vibranti”, “radianti” (si pensi all’emissione del suono di praticamente qualsiasi strumento musicale, alle onde elettromagnetiche . . .) o anche semplicemente periodici, si prestano ad essere trattati con modelli sinusoidali o opportune combinazioni di questi. A questo proposito facciamo un paio di osservazioni:

- Si osserva che sommando per esempio funzioni sinusoidali di uguale ampiezza e periodo ma con fasi differenti, si possono avere fenomeni di *interferenza* con associati effetti di *risonanza* o di *cancellazione* che sono tipici dei sistemi di tipo ondulatorio. Per esempio usando note formule di trigonometria si calcola che:

$$A \cos(\omega x) + A \cos(\omega(x - x_0)) = 2A \cos\left(\frac{x_0}{2}\omega\right) \cos\left[\omega\left(x - \frac{x_0}{2}\right)\right].$$

Se poniamo $x_0 = \frac{\pi}{\omega} = \frac{P}{2}$ le due funzioni si cancellano e la somma è nulla. Questo si può effettivamente realizzare in laboratorio illuminando un oggetto con due raggi di luce (che hanno “anche” un carattere ondulatorio) della stessa “intensità” (ampiezza), “colore” (frequenza) ma fasi che differiscono per $P/2$; il risultato è che l’oggetto resta al buio. E’ curioso notare come questa fenomenologia scientifica è per una volta passata nel linguaggio comune in modo pertinente: si usa dire infatti “sono fuori fase” per dire che non riusciamo ad accordarci con il ritmo della realtà circostante e che i nostri sforzi anche intensi hanno effetti mediocri.

- Sviluppando una funzione sinusoidale f usando ancora una volta note formule di trigonometria, si verifica che f può essere riscritta nella forma

$$f(x) = y_0 + \sum_{j=1}^n a_j \cos(j\omega x) + \sum_{j=1}^n b_j \sin(j\omega x)$$

per un certo $n \in \mathbb{N}$ e opportuni coefficienti reali a_j, b_j . Chiameremo una tale scrittura un *polinomio trigonometrico di frequenza angolare* ω . Dunque ogni funzione sinusoidale è un caso particolare di polinomio trigonometrico; ogni polinomio trigonometrico di data frequenza angolare è periodico di periodo $P = \frac{2\pi}{\omega}$. Sia ora $y = g(x)$ una qualsiasi funzione continua su \mathbb{R} , periodica di periodo P . Una parte dell’analisi matematica chiamata *analisi di Fourier* permette, per ogni $\epsilon > 0$, di determinare un polinomio trigonometrico p , di frequenza ω , tale che per ogni $x \in \mathbb{R}$, $|g(x) - p(x)| < \epsilon$. Dunque qualsiasi sistema reale che ammette un modello teorico adeguato $y = g(x)$ con g periodica, può essere ai fini pratici, in buona approssimazione, trattato con modelli della forma $y = p(x)$, dove p è un polinomio trigonometrico.

4. MODELLI DI TIPO ESPONENZIALE

Un modello di tipo esponenziale è della forma

$$y = Ae^{ax}$$

dove A e a sono due costanti reali.

In molte situazioni concrete si hanno esempi di crescita esponenziale di una data “popolazione”. Un caso estremo è dato da certe ameba che se collocate in un ambiente ideale (per disponibilità di cibo, temperatura, ...), sono virtualmente immortali, nel senso che ogni individuo sparisce in quanto tale duplicandosi in due individui a lui identici, in un intervallo di tempo fisso a partire dal suo momento di esistenza in quanto individuo. Prendendo tale intervallo come unità di misura del tempo, e assumendo che al tempo iniziale $t = 0$ la popolazione y è formata da un solo individuo, al tempo n avremo 2^n individui, e questo può essere esattamente interpolato con il modello esponenziale continuo

$$y = 2^t .$$

Un altro modello di tipo esponenziale compare per trattare fenomeni di *decadimento radioattivo*. Supponiamo di avere una regione dello spazio occupata da un campione di isotopo radioattivo di un certo atomo. Gli atomi radioattivi decadono (cessando di essere radioattivi) emettendo particelle α che tendono a rendere radioattivi altri atomi. Supponiamo che abbia senso definire e si possa misurare una grandezza “macroscopica” $V(t)$ che riflette la quantità di atomi radioattivi presenti al tempo t , a partire da un tempo iniziale $t = 0$. Il modello predice che se $[t_0, t_1]$ è un intervallo di tempo “piccolo”, allora:

$$V(t_1) - V(t_0) = -\lambda V(t_0)(t_1 - t_0)$$

dove $\lambda > 0$ è la *costante di decadimento* propria di quell’atomo. Il senso del modello è che la quantità di atomi che decade è proporzionale alla quantità di atomi radioattivi presenti, e che quest’ultima diminuisce con il passare del tempo. Fissiamo $t_1 > 0$ (piccolo) e calcoliamo per induzione $V(nt_1)$ per ogni $n \geq 1$ (notare che $(n + 1)t_1 - nt_1 = t_1$ è piccolo per ogni n); ponendo $t = nt_1$ si ottiene

$$V(t) = \left(1 - \frac{\lambda t}{n}\right)^n V(0) .$$

In questo modo abbiamo un po' arbitrariamente discretizzato il decadimento con "passo" t_1 . Ma in realtà il decadimento avviene continuamente, dunque ha senso considerare un arbitrario t_1 e studiare il comportamento della nostra funzione $V(t)$ quanto $t_1 = t/n \rightarrow 0^+$, cioè $n \rightarrow +\infty$. Riconosciamo uno dei limiti notevoli e concludiamo con il seguente modello esponenziale:

$$V(t) = V(0)e^{-\lambda t} .$$

Un modello dello stesso tipo è adeguato per analizzare una situazione molto più familiare, cioè il comportamento di un *libretto di risparmio a tasso fisso*. Disponendo di un capitale c_0 che prevediamo di non utilizzare per un tempo abbastanza lungo, possiamo aprire un libretto di risparmio con capitale iniziale c_0 , e farlo crescere grazie agli *interessi composti* maturati nel tempo. Il meccanismo è il seguente:

- E' fissato un intervallo di tempo (un anno, sei mesi,...) che possiamo prendere come unità di misura del tempo; si conviene che per maturare l'interesse alla fine di un tale intervallo, il capitale presente nel libretto all'inizio dell'intervallo è *vincolato*, cioè non può essere toccato per tutta la durata dell'intervallo.
- E' fissato un tasso di interesse fisso del p_0 per cento (per esempio del 2 per cento). Poniamo $r_0 = p_0/100$.
- Allora al tempo iniziale $t = 0$, abbiamo un capitale $c(0) = c_0$. Al tempo $t = 1$ si maturano gli interessi e si determina il nuovo capitale $c(1) = (1 + r_0)c(0)$; procedendo per induzione, al tempo $t = n$, abbiamo un capitale

$$c(n) = (1 + r_0)c(n-1) = (1 + r_0)^n c_0 .$$

La scelta di r_0 e dell'ampiezza dell'intervallo di tempo su cui il capitale è vincolato sono state abbastanza arbitrarie. E' chiaro che in linea di principio il tasso di interesse deve essere una funzione crescente dell'ampiezza di quell'intervallo. In accordo con il solito criterio di semplicità, adottiamo il seguente modello:

Se l'intervallo di vincolo è lungo $\lambda > 0$ (rispetto all'unità che abbiamo scelto) allora il corrispondente tasso di interesse è $r = \lambda r_0$.

Calcolando tutto come prima ma tenendo conto del parametro λ , vediamo che al solito $c(0) = c_0$, mentre

$$c(n\lambda) = (1 + r_0\lambda)c((n-1)\lambda) = (1 + r_0\lambda)^n c_0 .$$

Adesso vogliamo capire, per esempio, cosa succede quando $\lambda \rightarrow +\infty$ oppure $\lambda \rightarrow 0^+$ e $n \rightarrow +\infty$. Fissiamo per esempio una successione crescente divergente della forma $\lambda_n = \lambda_0 n$. Vediamo allora che la successione dei capitali al tempo λ_0 :

$$c_n(\lambda_0) = (1 + r_0\lambda_0 n)c_0$$

è crescente e diverge a $+\infty$. Questo non è sorprendente. Un comportamento più interessante si ha prendendo $\lambda_n = \lambda_0/n$ che converge a 0. Allora si ha:

$$c_n(\lambda_0) = \left(1 + \frac{r_0\lambda_0}{n}\right)^n c_0 .$$

Usando fatti noti, vediamo che $c_n(\lambda_0)$ è ancora una successione crescente che però adesso converge al limite finito $e^{r_0\lambda_0}$. In particolare se $\lambda_0 = 1 = r_0$, il limite è esattamente la costante di Nepero.

5. MODELLI DI TIPO LOGARITMICO

Un modello di tipo logaritmico è della forma

$$y = A \log x + a$$

definito per $x > 0$, dove A e a sono due costanti reali. Un importante esempio è la definizione dell'entropia data da Boltzman. Immaginiamo, per fissare le idee, di avere a che fare con un campione di gas, composto da N molecole, in un contenitore nello spazio 3-dimensionale. Uno stato microscopico del sistema è dato dall'insieme delle posizioni e delle velocità (istantanee) di tutte le molecole (6 gradi di libertà per ogni molecola). Dunque l'insieme \mathcal{S} di tutti gli stati microscopici (detto anche lo spazio delle fasi del sistema) può essere interpretato come una regione di \mathbb{R}^{6N} . La determinazione effettiva degli stati microscopici può essere impraticabile, per cui si può cercare una riduzione di grana grossa dello spazio delle fasi: decomponiamo \mathcal{S} in celle con la proprietà che due stati che stanno nella stessa cella sono macroscopicamente indistinguibili, mentre stati in celle diverse lo sono. Allora l'entropia S di uno stato corrispondente ad un punto $x \in \mathcal{S}$ è definita da:

$$S = k \log(V)$$

dove V è il volume della cella che contiene x , e k è la costante di Boltzman ($k = 1.38 \times 10^{-23}$ per una scelta opportuna delle unità di misura in gioco). L'uso del logaritmo ha un'importanza pratica perchè permette di maneggiare meglio i numeri molto grandi che di fatto intervengono maneggiando l'entropia (vedi la discussione qui sotto sulla "scala logaritmica") e, soprattutto ha un'importanza strutturale; infatti risulta che se un sistema $\mathcal{S} = \mathcal{S}_1 \cup \mathcal{S}_2$ è l'unione di due sistemi indipendenti, allora l'entropia associata al sistema totale è la somma delle entropie associate ai sistemi costituenti. Il secondo principio della termodinamica si può esprimere dicendo che ogni stato libero di evolvere nello spazio delle fasi, tende a muoversi secondo traiettorie che incontrano celle di volume strettamente crescente. Spesso il sistema ha una cella di volume molto più grande di tutte le altre, corrispondente agli stati in equilibrio termico.

6. ALTRI MODELLI REALISTICI

Modelli del tipo

$$y = Ae^{bx} \cos(\omega x + \phi)$$

ottenuti come prodotto di una funzione sinusoidale e di una esponenziale, rappresentano bene sistemi in cui sono presenti oscillazioni smorzate, per esempio quando $x \rightarrow +\infty$ se il parametro $b < 0$.

Il battimento è un fenomeno che avviene per somma di due vibrazioni di pari ampiezza, ma che differiscono l'una dall'altra per una differenza di frequenza, in modo che per interferenza si sommano e si cancellano periodicamente, dando luogo ad una vibrazione risultante con un andamento che può essere racchiuso tra due onde sinusoidali identiche ma sfasate tra loro di π . Fenomeni di battimento intervengono per esempio quando si accordano gli strumenti musicali. Modelli di battimento sono dati da funzioni elementari del tipo:

$$y = \cos(\pi a x) \sin(2\pi b x)$$

per una scelta opportuna delle costanti a e b .

Consideriamo le funzioni elementari del tipo:

$$y = f(x) = a(1 - e^{-k(x-x_0)}) + b$$

dove le costanti $a, k, x_0, b \in \mathbb{R}$, $a, k > 0$. La derivata $f'(x) = kae^{-k(x-x_0)}$, dunque è sempre strettamente positiva, dunque f è crescente. Inoltre $\lim_{x \rightarrow +\infty} f(x) = a + b$, così che f ha un asintoto orizzontale $y = a + b$. Una variante sono le funzioni elementari del tipo:

$$y = f(x) = \frac{a}{1 + e^{-k(x-x_0)}} + b$$

si verifica che anche in questo caso la funzione è crescente, ha l'asintoto orizzontale $y = b$ per $x \rightarrow -\infty$, l'asintoto orizzontale $y = a + b$ per $x \rightarrow +\infty$. Il primo tipo di funzione si presta ad essere modello di sistemi che presentano un fenomeno di saturazione per $x \rightarrow +\infty$, mentre il secondo tipo si presta quando è presente anche un fenomeno di saturazione inversa.

Esempio 6.1. Una modello del tipo

$$y = f(x) = \frac{a}{1 + e^{-k(x-x_0)}}$$

può essere realistico quando per esempio x rappresenta la concentrazione di un certo antibiotico, mentre y rappresenta la mortalità di un certo batterio. L'interpretazione qualitativa del modello dice che a "basse" concentrazioni la mortalità è praticamente nulla (saturazione inversa); ad "alte" concentrazioni la mortalità è praticamente indipendente dal valore della concentrazione (saturazione); in un intervallo di valori centrali della concentrazione l'efficacia dell'antibiotico è massima, infatti la derivata è positiva, è piuttosto grande (cioè il grafico è piuttosto ripido), così che anche piccole variazioni della concentrazione provocano un significativo aumento della mortalità.

Un altro esempio notevole di funzione elementare che incorpora un fenomeno di saturazione è il seguente

$$I = f(\nu) = 2h\nu^3(e^{\frac{h\nu}{kT}} - 1)$$

dove si considera l'intensità (l'ampiezza) I in funzione della frequenza (colore) ν della radiazione di un corpo radiante (per esempio un forno) contenuto in una cavità "nera" che assicura che la radiazione è in equilibrio termico a temperatura T con il materiale circostante; k è la costante di Boltzman già incontrata nella definizione dell'entropia, mentre h è la *costante di Planck* che in un opportuno sistema di unità di misura condiviso con k , prende un valore molto piccolo (circa $h = 6.62 \times 10^{-34}$). Studiando il grafico si vede che $f(0) = 0$, la funzione è positiva, cresce fino a raggiungere un valore di massimo assoluto, poi decresce ed ha l'asintoto orizzontale $I = 0$ per $\nu \rightarrow +\infty$. Questo modello proposto da Planck nel 1900, corregge quello quadratico proposto in precedenza da Rayleigh-Jeans, sulla base della teoria ondulatoria classica:

$$I = 2kT\nu^2 .$$

Il problema con questo precedente modello è che pur essendo in accordo con i risultati sperimentali per frequenze basse (nell'infrarosso) non lo è in modo radicale quando ν è molto grande (la cosiddetta "catastrofe ultravioletta"). Il modello di Planck oltre ad essere del tutto corroborato dai risultati sperimentali per ogni valore di ν , è di fondamentale importanza, anche storica, per il modo in cui fu derivato; la sofisticata analisi statistica di Planck che porta alla formula è basata sul postulato cruciale che h è una nuova costante fondamentale della natura e che le oscillazioni elettromagnetiche possono essere emesse o assorbite solo in modo "quantizzato", cioè in pacchetti di una specifica energia E , legata alla frequenza secondo la relazione lineare $E = h\nu$.

Il seguente esempio geometrico è piuttosto rivolto ad un lettore particolarmente interessato.

Vediamo ora come opportune funzioni elementari concorrano nel costruire un modello di *geometria non euclidea*, detta *geometria iperbolica*. Dati due punti $P = (x, y)$ e $Q = (s, t)$ del piano \mathbb{R}^2 ; indichiamo con $PQ = \sqrt{(x-s)^2 + (y-t)^2}$ la distanza euclidea tra i due punti che è anche uguale alla lunghezza (euclidea) del segmento rettilineo (non orientato) $[PQ]$ di estremi P e Q . Si consideri ora nel piano il disco unitario *aperto* $\mathcal{D} = \{x^2 + y^2 < 1\}$ delimitato dalla circonferenza di bordo $\mathcal{S} = \{x^2 + y^2 = 1\}$. Dati due punti distinti $A, B \in \mathcal{D}$, il segmento $[AB]$ è tutto contenuto in \mathcal{D} (si dice che \mathcal{D} è *convesso*). Definiamo ora una nuova lunghezza, detta *iperbolica*, di $[AB]$ nel modo seguente. Si consideri la retta r che contiene $[AB]$ e siano P e Q i due punti dell'intersezione $r \cap \mathcal{S}$ distribuiti rispetto ad una orientazione ausiliaria di r in modo che $P < A < B < Q$. Poniamo allora

$$d_h([AB]) = \frac{1}{2} \log\left(\frac{PB \cdot AQ}{PA \cdot BQ}\right) .$$

Se per semplicità supponiamo che $A = O$ sia il centro di \mathcal{D} e $0 < r = OB < 1$ allora il raggio iperbolico $r_h = d_h([OB])$ si esprime in funzione del raggio euclideo r nel modo seguente.

$$r_h = \frac{1}{2} \log\left(\frac{1+r}{1-r}\right) .$$

Si vede allora che $\lim_{r \rightarrow 0^+} r_h = 0$ mentre $\lim_{r \rightarrow 1^-} r_h = +\infty$. Possiamo quindi dire che \mathcal{S} è il *bordo all'infinito* di \mathcal{D} . Nella nuova geometria che stiamo descrivendo, avendo come termine di confronto quella euclidea,

\mathcal{D} gioca il ruolo di \mathbb{R}^2 . I triangoli in \mathcal{D} sono i (soliti) triangoli tutti contenuti in \mathcal{D} ; per la convessità di \mathcal{D} questo è equivalente a richiedere che i vertici siano punti di \mathcal{D} . Nella nuova geometria però la lunghezza dei lati di un triangolo non è quella euclidea che viene rimpiazzata da d_h . Il primo fatto che ci dice che stiamo trattando una onesta geometria è che anche le nuove lunghezze verificano le *disuguaglianze triangolari*: per ogni triangolo, comunque si scelga un lato la sua lunghezza è minore o uguale alla somma delle lunghezze degli altri due. Omettiamo la verifica di questa proprietà; ci limitiamo a segnalare che la proprietà funzionale di \log gioca un ruolo importante. Usando questi fatti, generalizzando quanto fatto prima in un caso particolare, si verifica che per ogni punto A di \mathcal{D} e ogni punto $S \in \mathcal{S}$ del bordo di \mathcal{D} , se $B \in \mathcal{D}$ appartiene al segmento di estremi A e S e $B \rightarrow A$ allora $d_h([AB]) \rightarrow 0$, mentre se $B \rightarrow S$, allora $d([AB]) \rightarrow +\infty$. Dunque, ponendo $d_h([AA]) = 0$ per ogni $A \in \mathcal{D}$, analogamente al caso euclideo $d_h([AB]) = d_h(A, B)$ definisce una *distanza* su \mathcal{D} tale che per ogni segmento $[AB]$ contenuto in \mathcal{D} la lunghezza del segmento è uguale alla distanza tra i suoi estremi. Per capire un po' il carattere di questa nuova geometria, si consideri il seguente problema: sia data una corda $[ST]$ di \mathcal{D} con estremi $S, T \in \mathcal{S}$. Siano $A, B \in [ST]$ due punti di \mathcal{D} giacenti su questa corda. Sia D il diametro di \mathcal{D} parallelo alla corda $[ST]$ di estremi $S', T' \in \mathcal{S}$ tali che $[SS']$ e $[TT']$ sono lati del trapezio di vertici S, T, S', T' . Vogliamo determinare A', B' su D in modo tale che $d_h([AB]) = d_h([A'B'])$. Questi si determinano geometricamente nel modo seguente: si considerano le semirette di origine S' e T' rispettivamente, passanti per i punti S e T rispettivamente. Queste due semirette si incontrano in un punto P esterno a \mathcal{D} . Allora A' è dato dall'intersezione con il diametro D della semiretta di origine P passante per A (analogamente per B'). Il fatto che le lunghezze d_h siano le stesse segue dalla formula che definisce d_h e dal teorema di Talete. Dunque muovendo la corda parallelamente in modo che gli estremi di avvicinino rispetto alla distanza euclidea, troviamo segmenti di lunghezza iperbolica costante che “appaiono” sempre più corti da un punto di vista euclideo. Analogamente alla geometria euclidea definiamo le “rette” iperboliche come i segmenti aperti non estendibili; dunque queste coincidono con i segmenti (aperti) in \mathcal{D} con estremi all'infinito cioè su \mathcal{S} . Come nella geometria euclidea, due rette iperboliche o non si intersecano (e diciamo allora che sono *parallele*) oppure si intersecano in un solo punto (sono *incidenti*); inoltre, per ogni punto passano infinite rette, per due punti distinti passa una sola retta. Ci sono però due modi in cui due rette iperboliche possono essere parallele: possono avere coppie di estremi su \mathcal{S} distinte (diciamo allora che sono *ultraparallele*), oppure avere un estremo in comune sul bordo di \mathcal{D} (parallele *incidenti all'infinito*). Preso A in \mathcal{D} e una retta iperbolica r che non passa per A , ci sono due rette passanti per A e incidenti all'infinito con r , mentre ci sono infinite rette ultraparallele a r passanti per A . Quindi in questa nuova geometria non vale il postulato delle parallele che vale per la geometria euclidea. Si ricordi che per secoli è stato discusso se l'assioma delle parallele fosse o no indipendente dagli altri assiomi della geometria euclidea. Il modello di geometria che stiamo descrivendo mostra che in effetti quell'assioma è indipendente dagli altri. Il trattamento della misura degli angoli nella geometria iperbolica è più involuto. Si potrebbe congetturare che la misura dell'angolo formato da due rette iperboliche ordinate coincida con la misura euclidea. Ma è questo il caso solo se le due rette sono diametri. In generale descriviamo la ricetta senza giustificarla. Prendiamo una retta r in \mathcal{D} di estremi S e T in \mathcal{S} che non sia un diametro e sia A un punto di r . Esiste un'unica circonferenza C di centro esterno a \mathcal{D} tale che C e \mathcal{S} si intersecano ortogonalmente in S e T . Consideriamo l'arco (aperto) di circonferenza $\Gamma = C \cap \mathcal{D}$. Sia A' il punto di Γ ottenuto intersecando Γ con la semiretta di origine O , passante per A . Se la retta r è invece un diametro, poniamo $\Gamma = r$ e $A' = A$. Se ora abbiamo due rette ordinate r e r' che si intersecano in A possiamo realizzare A' come intersezione dei due rispettivi archi ordinati Γ e Γ' . Considerando le rispettive rette tangenti in A' , questi archi determinano un angolo di una data misura *euclidea* α . Allora α è anche la misura *iperbolica* dell'angolo formato dalle due rette ordinate r e r' . Ne segue che la somma delle misure degli angoli interni di un triangolo iperbolico è $< \pi$ (contro il fatto che nella geometria euclidea tale somma è uguale a π). Si può sviluppare una trigonometria iperbolica in cui le funzioni \cosh e \sinh (introdotte alla fine di [EC]) svolgono un ruolo analogo alle funzioni \cos e \sin nella trigonometria classica. Inoltre la geometria iperbolica è essenziale per costruire modelli della cinematica della relatività ristretta.

Cogliamo qui l'occasione per aggiungere qualche informazione sulle funzioni \cosh e \sinh . Ricordiamo le definizioni (su tutto \mathbb{R})

$$\sinh(x) := \frac{e^x - e^{-x}}{2}, \quad \cosh(x) := \frac{e^x + e^{-x}}{2}, \quad \tanh(x) = \frac{\sinh(x)}{\cosh(x)}.$$

Notiamo intanto che \cosh è pari mentre \sinh è dispari. Per ogni $x \in \mathbb{R}$, $\cosh^2(x) - \sinh^2(x) = 1$. Inoltre si verifica direttamente che $\cosh' = \sinh$, $\sinh' = \cosh$. Su l'intervallo $[0, +\infty)$, entrambe \cosh e \sinh sono > 0 . Quando $x \rightarrow +\infty$, $\cosh(x) > e^x$ dunque tende a $+\infty$. D'altra parte, $\cosh(x) - \sinh(x) = e^{-x}$ che tende a 0. Ne deduciamo che \sinh è strettamente crescente su \mathbb{R} , si annulla solo per $x = 0$, ed è bigettivo, per cui è definita su tutto \mathbb{R} la funzione inversa $\operatorname{arcsinh}$. La funzione \cosh è strettamente crescente per $x > 0$, decrescente per $x < 0$, ed ha un punto di minimo in $x = 0$, $\cosh(0) = 1$. La restrizione di \cosh su $[0, +\infty)$ è bigettiva sopra $[1, +\infty)$ e possiamo definire la funzione inversa $\operatorname{arcosh} : [1, +\infty) \rightarrow [0, +\infty)$. Lasciamo al lettore di aggiungere informazioni a proposito della concavità di queste funzioni. Per quanto riguarda la tangente iperbolica, si verifica la relazione $\tanh^2(x) = 1 - \frac{1}{\cosh^2(x)}$.

Un altro esempio importante di funzione elementare è formato dalle funzioni *Gaussiane* cioè della forma

$$y = Ae^{-\beta x^2}$$

dove A e β sono costanti > 0 . Una tale funzione è pari, positiva, con 0 come punto di massimo assoluto, decrescente su $x > 0$ e con asintoto orizzontale $y = 0$ quando $x \rightarrow +\infty$. Per molti fenomeni aleatori (anche di grande importanza applicativa), le funzioni Gaussiane intervengono per definire la corrispondente distribuzione di probabilità.

7. SCALA LOGARITMICA ED ALTRI ESEMPI DI INTERPOLAZIONE

Dato un sistema di coordinate cartesiane x, y su \mathbb{R}^2 (rispetto ad una fissata unità di misura dei segmenti), la scala logaritmica su uno degli assi (per esempio l'asse delle y) si ottiene facendo il seguente cambio di coordinata sul semiasse positivo:

$$t = \log(y).$$

Ci sono delle circostanze pratiche in cui conviene fare questo cambiamento di scala (su uno o entrambi gli assi). Supponiamo per esempio che dobbiamo riportare su una pagina di un testo certe coppie di dati numerici (x_j, y_j) , $j = 1, \dots, n$, e che per esempio le $y_j > 0$ crescano molto rapidamente. E' facile allora che quelle coppie escano dalla pagina. Potremmo rimpicciolire l'unità di misura, cioè rimpicciolire linearmente la figura, con il rischio però che adesso questa diventi troppo piccola e poco leggibile. Siccome $\lim_{y \rightarrow +\infty} \frac{\log(y)}{y} = 0$, c'è più spazio sulla stessa pagina per riportare in modo efficace le coppie di dati (x_j, t_j) .

L'uso della scala logaritmica permette di estendere i risultati sull'interpolazione lineare discussi prima ad altri tipi di modello. Per analogia con la discussione fatta prima per i modelli lineari, supponiamo di avere a che fare con un dato sistema reale e qualche motivo per congetturare che qualitativamente il modello sia di tipo esponenziale:

$$y = Ae^{ax}, \quad A > 0.$$

Vogliamo allora corroborare (o meno) sperimentalmente questa ipotesi qualitativa e (nel caso fosse confermata) determinare con buona approssimazione l'effettiva legge (cioè la coppia di "costanti di struttura" del sistema (A, a)). Passando alla scala logaritmica rispetto all'asse delle y , $t = \log(y)$, riconduciamo questo problema a quello già risolto nel caso lineare; infatti si ottiene la funzione

$$t = ax + \log(A)$$

cioè nelle nuove coordinate il modello diventa formalmente lineare. Se per esempio $t = m_0x + q_0$ è la retta di regressione per certi dati (x_j, t_j) , otteniamo l'interpolazione esponenziale dei dati (x_j, y_j)

nelle coordinate originali:

$$y = e^{q_0} e^{m_0 x} .$$

Se per esempio il modello congetturato è di tipo potenza, definito per $x > 0$:

$$y = Ax^a = Ae^{a \log(x)} , \quad A > 0$$

allora prendendo la scala logaritmica su entrambi gli assi ($z = \log(x)$, $t = \log(y)$) ci riconduciamo ancora una volta ad un modello formalmente lineare:

$$t = az + \log(A) .$$

Se il modello congetturato è del tipo con saturazione $y = f(x) = a(1 - e^{-k(x-x_0)}) + b$, si vede che $e^{-k(x-x_0)} = 1 - \frac{f(x) - b}{a}$, da cui, posto $z = \log(1 - \frac{f(x) - b}{a})$, si ricava $z = -kx + kx_0$. Ammettendo di conoscere per altra via i valori a e b “giusti”, possiamo ricavare k e x_0 per interpolazione lineare.