2. Funzioni convesse

2.1 Definizioni e proprietà

La nozione di funzione convessa è piuttosto recente e risale alla fine dell'ottocento. Gibbs e Maxwell la utilizzarono per modellare le relazioni fra le variabili termodinamiche. È rilevante in molti contesti, perché si potrebbe dire che, tra le funzioni non lineari, quelle convesse sono le più vicine alle lineari.

L'ambiente più naturale in cui parlare di funzioni convesse è quello multidimensionale. Qui discutiamo le funzioni convesse di una variabile reale.

2.1 Definizione. *Un insieme E del piano si dice* convesso *se per ogni coppia di punti P_1, $P_2 \in E$ il segmento P_1P_2 di estremi P_1 e P_2 è contenuto in E.*

2.2 Definizione. *Una funzione $f : [a, b] \to \mathbb{R}$ si dice* convessa *in $[a, b]$ se $\forall\ x_1, x_2 \in [a, b]$ e $\lambda \in [0, 1]$ si ha*

$$f((1-\lambda)x_1 + \lambda x_2) \leq (1-\lambda)f(x_1) + \lambda f(x_2). \tag{2.1}$$

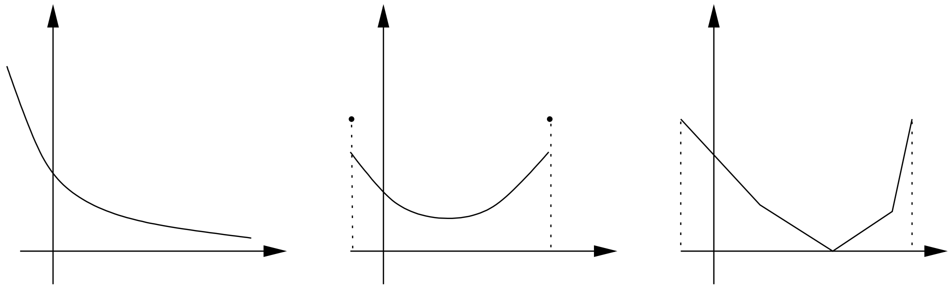

Figura 2.1. Esempi di funzioni convesse.

Posto $x = (1-\lambda)x_1 + \lambda x_2$ la (2.1) si può riscrivere come

$$f(x) \leq f(x_1) + \frac{f(x_2) - f(x_1)}{x_2 - x_1}(x - x_1), \qquad x \in [x_1, x_2].$$

Dunque *f è convessa se e solo se per ogni x_1, x_2 con $x_1 < x_2$, il grafico di f in $[x_1, x_2] \times \mathbb{R}$ è al di sotto della corda per $(x_1, f(x_1))$ e $(x_2, f(x_2))$.*

2.3 Proposizione. *Sia $f : [a, b] \to \mathbb{R}$. Sono fatti equivalenti*

(i) *f è convessa in $[a, b]$.*

(ii) *Per ogni $x_1, x_2 \in [a, b]$ con $x_1 < x_2$ il grafico di f in $[a, b] \setminus [x_1, x_2]$ è sopra la corda per $(x_1, f(x_1))$ e $(x_2, f(x_2))$*

$$f(x) \geq f(x_1) + \frac{f(x_2) - f(x_1)}{x_2 - x_1}(x - x_1), \qquad \forall x \notin [x_1, x_2].$$

(iii) *L'*epigrafico *di f definito da*

$$E_f := \{(x, y) \in [a, b] \times \mathbb{R} \ \mid \ y \geq f(x)\}$$

è un insieme convesso.

(iv) $\forall\, x, y, z \in [a, b]$ *con* $x < y < z$ *si ha*

$$\frac{f(y)-f(x)}{y-x} \leq \frac{f(z)-f(x)}{z-x} \leq \frac{f(z)-f(y)}{z-y}. \tag{2.2}$$

(v) *Per ogni* $x_0 \in [a, b]$ *il coefficiente angolare delle corde*

$$F_{x_0}(x) := \frac{f(x)-f(x_0)}{x-x_0}$$

è una funzione crescente in $[a, b]$.

Dimostrazione. Indichiamo per brevità con $r_{uv}(x)$ la retta per $(u, f(u))$ e $(v, f(v))$.
(i) $\Leftrightarrow$ (ii) Se (ii) non fosse vera, esisterebbe $x_3 \notin [x_1, x_2]$ per cui

$$f(x_3) < r_{x_1x_2}(x_3)$$

Supponendo che $x_3 < x_1 < x_2$ si avrebbe allora che

$$f(x_1) > r_{x_3x_2}(x_1)$$

che contraddice la (i). In modo del tutto simile si dimostra l'implicazione opposta.
(i) $\Leftrightarrow$ (iii) Si lascia al lettore.
(i) $\Leftrightarrow$ (iv) $\Leftrightarrow$ (v). Dalla diseguaglianza di convessità

$$f(y) - f(x) \leq \frac{f(z)-f(x)}{z-x}(y-x)$$

si ricava la prima diseguaglianza di (iv) dividendo per $y - x$. D'altra parte dalla (ii)

$$f(x) - f(z) \geq \frac{f(y)-f(z)}{y-z}(x-z)$$

da cui si ricava immediatamente la seconda diseguaglianza in (iv). (v) è una riscrittura della (iv). Infine se $x_1 < x < x_2$, dalla prima diseguaglianza in (iv) si ricava

$$\frac{f(x)-f(x_1)}{x-x_1} \leq \frac{f(x_2)-f(x_1)}{x_2-x_1}$$

vale a dire la diseguaglianza di convessità. $\square$

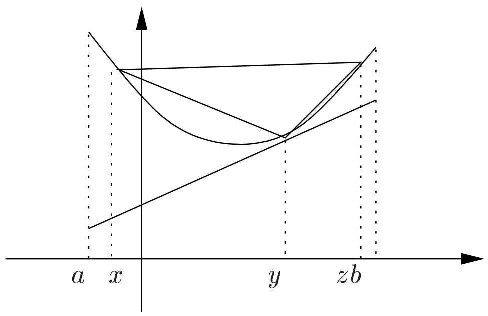

Figura 2.2. Pendenza delle corde e retta tangente a una funzione convessa.

Se la funzione f è derivabile almeno una volta, la convessità si puo' caratterizzare in altri modi. Si ha

2.4 Proposizione. *Sia* $f : [a, b] \to \mathbb{R}$ *una funzione derivabile. Sono fatti equivalenti*

(i) f *è convessa,*
(ii) $f'(x)$, $x \in [a, b]$, *è crescente,*

(iii) *Il grafico di f è sopra tutte le sue tangenti, cioè per ogni x, $x_0 \in [a,b]$*

$$f(x) \geq f(x_0) + f'(x_0)(x - x_0).$$

Se poi f è derivabile due volte, f è convessa in $[a,b]$ se e solo se $f''(x) \geq 0$ $\forall x \in (a,b)$.

Dimostrazione. (i) $\Rightarrow$ (ii) Siano $x < y < z \in [a,b]$. Passando al limite nella (2.2) per $y \to x^+$ e $y \to z^-$ si trova

$$f'(x) = f'_+(x) \leq \frac{f(z) - f(x)}{z - x} \leq f'_-(z) = f'(z) \tag{2.3}$$

il che prova che f' è crescente.

(ii) $\Rightarrow$ (iii) . Siano $x, x_0 \in [a,b]$. Se $x < x_0$ la (2.3) con $z = x_0$ dà

$$\frac{f(x_0) - f(x)}{x_0 - x} \leq f'(x_0)$$

i.e.,

$$f(x) \geq f(x_0) + f'(x_0)(x - x_0)$$

se invece $x > x_0$ sempre la (2.3), questa volta con $z = x$ e $x = x_0$, dà

$$f'(x_0) \leq \frac{f(x) - f(x_0)}{x - x_0}$$

(iii) $\Rightarrow$ (i). Se $x_1 < x < x_2 \in [a,b]$ allora dalla (iii)

$$f(x_1) \geq f'(x)(x_1 - x) + f(x) \quad \text{e} \quad f(x_2) \geq f'(x)(x_2 - x) + f(x).$$

Moltiplicando la prima diseguaglianza per $\lambda := (x_2 - x)/(x_2 - x_1)$ e la seconda per $1 - \lambda = (x - x_1)/(x_2 - x_1)$, si trova

$$f(x) \leq \lambda f(x_1) + (1 - \lambda) f(x_2), \text{ con } x = \lambda x_1 + (1 - \lambda) x_2,$$

i.e., f è convessa. □

2.5 ¶ Si può in realtà dire di più. Ripercorrendo la dimostrazione della proposizione Proposizione 2.4, si dimostra

Theorem. *Sia $f : [a,b] \to \mathbb{R}$ una funzione convessa. Allora*
(i) *f ha derivata destra e sinistra finite in ogni punto $x \in]a,b[$ interno.*
(ii) *Le funzione $f'_+(x)$ e $f'_-(x)$ sono crescenti in $]a,b[$,*
(iii) *Per ogni c, d con $a < c < d < b$ f ha rapporto incrementale limitato in $[c,d]$*
(iv) *f è continua in ogni punto $x \in]a,b[$ interno.*

Dimostrazione. (i) Per ogni $x_0 \in [a,b]$ il rapporto $F_{x_0}(x)$ è crescente in x e dunque esistono, per il teorema sulla esistenza dei limiti di funzioni monotone,

$$f'_+(x_0) := \lim_{x \to x_0^+} \frac{f(x) - f(x_0)}{x - x_0}$$

$$f'_-(x_0) := \lim_{x \to x_0^-} \frac{f(x) - f(x_0)}{x - x_0}.$$

Inoltre $f'_-(x_0)$ e $f'_+(x_0)$ sono finiti se $x_0 \in]a,b[$.

Poi per $x, y \in [a,b]$

$$f'_-(x) \leq f'_+(x) \leq \frac{f(y) - f(x)}{y - x} = \frac{f(x) - f(y)}{x - y} \leq f'_-(y) \leq f'_+(y), \tag{2.4}$$

e (ii) è provata.

(iii) e (iv). Sempre dalla (2.4) per ogni $x, y \in [c,d]$, $x < y$

$$-\infty < f'_-(c) \leq f'_-(x) \leq \frac{f(y) - f(x)}{y - x} \leq f'_+(y) \leq f'_+(d) < \infty,$$

il che prova (iii). La continuità di f all'interno di $[a,b]$ è ora una semplice conseguenza. □

2.6 ¶ Sia $f : [a,b] \to \mathbb{R}$ convessa e continua. Se $f(a)f(b) < 0$ allora f si annulla in un unico punto.

2.7 ¶ Se $f : [a,b] \to \mathbb{R}$ è lineare, allora f ha massimo e minimo agli estremi. Se $f : [a,b] \to \mathbb{R}$ è convessa, allora f ha massimo agli estremi. In formula $\max_{x\in[a,b]} f(x) = \max(f(a), f(b))$.

2.8 ¶ Mostrare che f è convessa se e solo se per ogni $x_0 \in (a,b)$ esiste $\lambda(x_0)$ tale che

$$f(x) \geq f(x_0) + \lambda(x_0)(x - x_0).$$

La retta $y = f(x_0) + \lambda(x_0)(x - x_0)$ si chiama *retta di appoggio* a f in $(x_0, f(x_0))$. Osservare che una funzione convessa può avere in un punto più rette di appoggio.

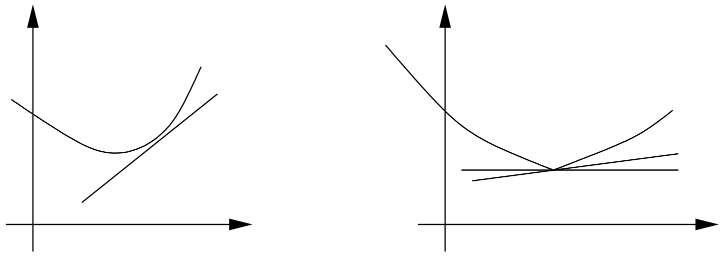

Figura 2.3. Rette di appoggio.

2.2 Qualche diseguaglianza elementare

In questa sezione abbiamo raccolto varie uguaglianze e diseguaglianze di uso comune. Anche se si possono fare dimostrazioni differenti, i metodi del calcolo e l'uso della convessità forniscono spesso dimostrazioni piuttosto semplici.

2.9 (Seno) Sappiamo già che $\sin x \leq x\ \forall x \geq 0$. Si ha anche

$$\sin x \geq x - \frac{x^3}{6}. \tag{2.5}$$

Infatti posto $f(x) := \sin x - x + x^3/6$, $x \geq 0$, si ha $f'(x) = \cos x - 1 + x^2/2$, $f''(x) = -\sin x + x \geq 0$ per ogni $x \geq 0$. Dunque f è convessa su $\{x \geq 0\}$, in particolare

$$f(x) \geq f(0) + f'(0)(x - 0) = 0$$

i.e., la (2.5).

2.10 (Coseno) Analogamente si prova che

$$1 - \frac{x^2}{2} \leq \cos x \leq 1 - \frac{x^2}{2} + \frac{x^4}{24} \qquad \forall x \in \mathbb{R} \tag{2.6}$$

La diseguaglianza a sinistra nella (2.6) si prova facilmente perché

$$1 - \cos x = 2\sin^2\left(\frac{x}{2}\right) \leq 2\frac{x^2}{4} = \frac{x^2}{2}.$$

Per l'altra, si pone $f(x) := 1 - \cos x - x^2/2 + x^4/24$, $x \in \mathbb{R}$. e si calcola $f'(x) = \sin x - x + x^3/6$, $f''(x) = \cos x - 1 + x^2/2$ Percio' $f''(x) \geq 0$ su $\mathbb{R}$, i.e., f è convessa e dunque

$$f(x) \geq f(0) + f'(0)(x - 0) = 0$$

i.e., la diseguaglianza a destra nella (2.6).

2.11 (Logaritmo e esponenziale) La convessità dell'esponenziale e la concavità del logaritmo danno immediatamente ad esempio

$$e^x \geq e^0 + e^0(x-0) = 1 + x, \qquad \forall x \in \mathbb{R},$$
$$\log x \leq \frac{1}{1}(x-1) + \log 1 = x - 1, \qquad \forall x > 0.$$

Volendo una stima dal basso si può utilizzare una secante al grafico. Ad esempio se $a \in \mathbb{R}$, $a \neq 0$

$$e^x \leq \frac{e^a - 1}{a} x + 1, \qquad \forall x \in [0, a]$$

e se $a > 1$

$$\log x \geq \frac{\log a}{a-1}(x-1), \qquad \forall x \in [1, a].$$

a. Diseguaglianza di Bernoulli

2.12 Proposizione. *Se $\alpha \geq 1$ allora*

$$(1+x)^\alpha \geq 1 + \alpha x \qquad \forall x > -1$$

Dimostrazione. Sia $f(x) := (1+x)^\alpha$, $x > -1$. Derivando si ottiene

$$f'(x) = \alpha(1+x)^{\alpha-1}, \qquad f''(x) = \alpha(\alpha-1)(1+x)^{\alpha-2}.$$

$f''(x)$ è dunque positiva, quindi f è convessa. In particolare

$$f(x) \geq f(0) + f'(0)x = 1 + \alpha x.$$

□

In particolare se $a > 1$ si ha per ogni $n \in \mathbb{N}$

$$a^n \geq 1 + n(a-1), \qquad a^{-n} \leq \frac{1}{1 + n(a-1)}.$$

2.13 ¶ Fare una dimostrazione per induzione su n della diseguaglianza precedente.

b. Binomio di Newton

2.14 Proposizione. *Per ogni a, $b \in \mathbb{R}$ e $n \in \mathbb{N}$ si ha*

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

Dimostrazione. Se entrambi a e b sono nulli la tesi è ovvia. Se almeno uno dei due, diciamo a è non nullo, dividendo per a^n, ci si riduce a provare che

$$(1+x)^n = \sum_{k=0}^{n} \binom{n}{k} x^k \qquad \forall x \tag{2.7}$$

È facile verificare che $\sum_{k=0}^{n} \binom{n}{k} x^k$ è lo sviluppo di Taylor centrato in 0 di ordine n di $(1+x)^n$. D'altra parte $(1+x)^n$ è un polinomio di ordine n e dunque coincide con $\sum_{k=0}^{n} \binom{n}{k} x^k$. □

Altra dimostrazione della (2.7). Sia $g(x) := \sum_{k=0}^n \binom{n}{k} x^k$, $x \in \mathbb{R}$. Derivando in x si ottiene

$$\begin{aligned}(1+x)g'(x) &= \sum_{k=1}^n \binom{n}{k} k x^{k-1} + \sum_{k=1}^n \binom{n}{k} k x^k \\ &= \binom{n}{1} x^0 + \sum_{k=1}^{n-1} \left((k+1)\binom{n}{k+1} + k\binom{n}{k} \right) x^k + n\binom{n}{n} x^n \\ &= n + n\sum_{k=1}^{n-1} \binom{n}{k} x^k + n x^n = n \sum_{k=0}^n \binom{n}{k} x^k = n g(x)\end{aligned}$$

perché

$$(k+1)\binom{n}{k+1} + k\binom{n}{k} = n\binom{n}{k}.$$

Se ora $f(x) := (1+x)^n$, si ha $f(0) = 1 = g(0)$ e

$$D\frac{g(x)}{f(x)} = \frac{g'(x)}{(1+x)^n} - \frac{ng(x)}{(1+x)^{n+1}} = 0.$$

La funzione $g(x)/(1+x)^n$ è dunque identicamente 1 e la tesi è provata. □

Un'altra dimostrazioe ancora per induzione diretta su n si trova al Cap. II, Esempio 1.6.

2.3 Diseguaglianza di Jensen

Si puo' riscrivere la condizione di convessità per una funzione $f : [a, b] \to \mathbb{R}$ come

$$f(\alpha_1 x_1 + \alpha_2 x_2) \le \alpha_1 f(x_1) + \alpha_2 f(x_2)$$

per ogni $x_1, x_2 \in [a, b]$ e ogni α_1, α_2 con $\alpha_1, \alpha_2 \ge 0$ e $\alpha_1 + \alpha_2 = 1$. Si noti che i punti $\alpha_1 x_1 + \alpha_2 x_2$ descrivono tutto il segmento $[x_1, x_2]$ al variare di α_1, α_2 con i vincoli $\alpha_1, \alpha_2 \ge 0$ e $\alpha_1 + \alpha_2 = 1$. Piu' in generale

2.15 Proposizione (Diseguaglianza di Jensen discreta). *Sia $f : [a, b] \to \mathbb{R}$ una funzione convessa e siano $n \ge 2$, $x_1, x_2, \dots, x_n$ n punti di $[a, b]$ e $\alpha_1, \alpha_2, \dots, \alpha_n$ n numeri reali con $\alpha_i \ge 0$ e $\sum_{i=1}^n \alpha_i = 1$. Allora $\sum_{i=1}^n \alpha_i x_i \in [a, b]$ e*

$$f\Big(\sum_{i=1}^n \alpha_i x_i\Big) \le \sum_{i=1}^n \alpha_i f(x_i)$$

Se poi f è strettamente convessa e

$$f\Big(\sum_{i=1}^n \alpha_i x_i\Big) = \sum_{i=1}^n \alpha_i f(x_i)$$

per una scelta dei coefficienti $\alpha_1, \alpha_2, \dots, \alpha_n$ tutti non nulli e con $\sum_{i=1}^n \alpha_i = 1$, allora $x_1 = x_2 = \dots = x_n$.

Dimostrazione. Si procede per induzione su n. Per $n = 2$ la tesi è la definizione di funzione convessa. Supposto ora di aver provato la proprietà per $n-1$ punti, proviamola per n punti. Sia $\alpha := \alpha_1 + \dots + \alpha_{n-1}$, e quindi $\alpha_n := 1 - \alpha$. Se $\alpha = 0$ o $\alpha = 1$ per l'ipotesi induttiva la tesi è provata, altrimenti $0 < \alpha < 1$ e posto $\beta_i := \alpha_i/\alpha$, $i = 1, \dots, n-1$,

$$x := \sum_{i=1}^{n-1} \beta_i x_i$$

si ha

$$0 \le \beta_i \le 1, \quad \sum_{i=1}^{n-1} \beta_i = 1, \quad \sum_{i=1}^{n} \alpha_i x_i = \alpha x + (1-\alpha) x_n$$

e quindi, essendo $x \in [a,b]$ per l'ipotesi di induzione, si ha $\sum_{i=1}^{n} \alpha_i x_i \in [a,b]$ e

$$f(\sum_{i=1}^{n} \alpha_i x_i) = f(\alpha x + (1-\alpha)x_n) \le \alpha f(x) + (1-\alpha) f(x_n) \le \sum_{i=1}^{n} \alpha_i f(x_i).$$

Proviamo ora la seconda parte. Supponiamo ora che la funzione f sia strettamente convessa. Procediamo anche questa volta per induzione su n. Per $n = 2$, se per assurdo fosse $x_1 < x_2$, e $0 < \lambda < 1$, si avrebbe che $(1-\alpha)x_1 + \alpha x_2$ è *interno* al segmento $[x_1, x_2]$ e dunque $f((1-\lambda)x_1 + \lambda x_2) < (1-\lambda) f(x_1) + \lambda f(x_2)$, il che contraddice l'ipotesi

$$f((1-\lambda)x_1 + \lambda x_2) = (1-\lambda) f(x_1) + \lambda f(x_2).$$

Supponiamo ora la tesi vera per $n-1$ punti e dimostrimola per n punti. Se αm x, β_i sono definiti come prima, $\sum_{k=1}^{n} \alpha_i x_i = \alpha x + (1-\alpha)x_n$, con $0 < \alpha < 1$ per ipotesi. Pertanto

$$f(\alpha x + (1-\alpha)x_n) = \sum_{i=1}^{n-1} \alpha_i f(x_i) + (1-\alpha) f(x_n) = \alpha \sum_{i=1}^{n-1} \beta_i f(x_i) + (1-\alpha) f(x_n) \ge \alpha f(x) + (1-\alpha) f(x_n)$$

e essendo f convessa,

$$f(\alpha x + (1-\alpha)x_n) = \alpha \sum_{i=1}^{n-1} \beta_i f(x_i) + (1-\alpha) f(x_n) = \alpha f(x) + (1-\alpha) f(x_n).$$

Da

$$\sum_{i=1}^{n-1} \beta_i f(x_i) = f(x) = f\Big(\sum_{i=1}^{n-1} \beta_i x_i\Big)$$

e dall'ipotesi induttiva segue che $x_1 = x_2 = \dots = x_{n-1} =: \overline{x}$ In particolare $x = \sum_{i=1}^{n-1} \beta_i x_i = \overline{x}$. D'altra parte da $f(\alpha x + (1-\alpha)x_n) = \alpha f(x) + (1-\alpha) f(x_n)$ e dal caso $n = 2$, segue che $x_n = x = \overline{x}$. □

Un caso tipico è la scelta di $(\alpha_1, \alpha_2, \dots, \alpha_n) = (1/n, \dots, 1/n)$. La diseguaglianza di Jensen ha allora l'aspetto di

$$f\Big(\frac{1}{n} \sum_{k=1}^{n} x_k\Big) \le \frac{1}{n} \sum_{k=1}^{n} f(x_k) \tag{2.8}$$

2.4 Diseguaglianze e convessità

Molte disegueglianze utili si possono trattare in modo unitario in termini di convessità di opportune funzioni. Sono anzi spesso una semplice applicazione della disguaglianza di Jensen discreta.

a. Media aritmetica e media quadratica

La (2.8) applicata alla funzione convessa t^2 da subito la *diseguaglianza tra media aritmetica e media quadratica*

$$\overline{x} := \frac{1}{n} \sum_{k=1}^{n} x_k \le \Big(\frac{1}{n} \sum_{k=1}^{n} x_k^2\Big)^{1/2} \tag{2.9}$$

Si noti che

$$\frac{1}{n} \sum_{k=1}^{n} x_k^2 = \overline{x}^2 + \frac{1}{n} \sum_{k=1}^{n} (x_k - \overline{x})^2, \qquad \overline{x} := \frac{1}{n} \sum_{k=1}^{n} x_k. \tag{2.10}$$

2.16 ¶ Una dimostrazione della (2.9) per induzione diretta su n si trova al Cap. I, Esempio 3.16.

b. Media aritmetica e media geometrica

La convessità di e^x dà

$$\exp\left(\sum_{i=1}^n \alpha_i x_i\right) \le \sum_{i=1}^n \alpha_i e^{x_i} \tag{2.11}$$

per tutti gli $x_1, \dots, x_n \in \mathbb{R}$, $\alpha_i \ge 0$ con $\sum_{i=1}^n \alpha_i = 1$. Essendo e^x strettamente convessa l'uguaglianza nella stima precedente con tutti gli α_i non nulli implica che $x_1 = x_2 = \dots = x_n$.

Ponendo $y_i = e^{x_i}$ nella (2.11) otteniamo la

2.17 Proposizione. *Se $y_i \ge 0$, $\alpha_i \ge 0$ e $\sum_{i=1}^n \alpha_i = 1$ allora*

$$y_1^{\alpha_1} \cdot y_2^{\alpha_2} \cdots y_n^{\alpha_n} \le \sum_{i=1}^n \alpha_i y_i. \tag{2.12}$$

In particolare se $\alpha_1 = \dots = \alpha_n = 1/n$

$$(y_1 \cdot \dots \cdot y_n)^{1/n} \le \frac{1}{n}\sum_{i=1}^n y_i.$$

c. Diseguaglianza di Young

Una scelta possibile nella (2.12) è anche $n = 2$, $\alpha_1 = 1/p$, $\alpha_2 = 1/q$, $p, q > 0$, $1/p + 1/q = 1$, $y_1 = A \ge 0$, $y_2 = B \ge 0$. Si ha

$$A^{1/p} B^{1/q} \le \frac{1}{p}A + \frac{1}{q}B,$$

e scegliendo poi $A = \epsilon\, a^p$ e $B = \epsilon^{-1}\, b^q$ si trova la *diseguaglianza di Young*

2.18 Proposizione. *Per ogni $a, b \ge 0$, per ogni $p, q > 1$ con $1/p + 1/q = 1$ e ogni $\epsilon > 0$ si ha*

$$ab \le \epsilon \frac{a^p}{p} + \frac{1}{\epsilon}\frac{b^q}{q}. \tag{2.13}$$

d. Entropia

Siano (X, P) uno spazio di probabilità, n un intero fissato, e $A_1, A_2, \dots, A_n$ n eventi esclusivi ($\{A_k\}$ è una partizione di X) con probabilità $P(A_1), \dots P(A_n)$. Si ha $\sum_{i=1}^n P(A_i) = 1$. La funzione

$$F(A_1, A_2, \dots, A_n) := -\sum_{i=1}^n P(A_i)\log(P(A_i))$$

da' una misura di quanto gli eventi $A_1, \dots, A_n$ si discostano dall'essere equiprobabili. Si dimostra infatti che la funzione f ha massimo in corrispondenza dei soli eventi equiprobabili.

Matematicamente, si puo' considerare la funzione di n variabili,

$$f(x) = f(x_1, \dots, x_n) = \frac{1}{n}\sum_{i=1}^n x_i \log x_i$$

sull'insieme $\Delta = \{x \in \mathbb{R}^n \,|\, x_i \ge 0, \sum_{i=1}^n x_i = 1\}$. Ovviamente

$$F(A_1, \dots, A_n) = -f(P(A_1), \dots, P(A_n))$$

e per questo $-f$ è detta *funzione entropia*. Si ha

2.19 Proposizione (Massimo dell'entropia). *$f : \Delta \to \mathbb{R}$ ha un unico minimo nel punto*

$$\overline{x} := (1/n, \dots, 1/n) \in \Delta.$$

Dimostrazione. Per ogni $x = (x_1, x_2, \dots, x_n) \in \Delta$, indichiamo con $\overline{x}$ la media che in questo caso è costantemente $\overline{x} := \frac{1}{n}\sum_{i=1}^n x_i = \frac{1}{n}$. Sia poi $\varphi(t)$ la funzione $\varphi(t) := t\log t$, $t \in (0,1)$. Si ha $f(x) = \frac{1}{n}\sum_{i=1}^n \varphi(x_i)$ e, essendo φ convessa in $(0,1)$, segue dalla (2.8)

$$f(\overline{x}) = \varphi\Big(\frac{1}{n}\Big) = \varphi\Big(\frac{1}{n}\sum_{i=1}^n x_i\Big) \le \sum_{i=1}^n \varphi(x_i) = f(x).$$

Per l'abitrarietà di $x \in \Delta$, $\overline{x}$ è un minimo per f su Δ. Mostriamo che non ci sono altri punti di minimo. Se ve ne fosse un'altro, diciamo $y = (y_1, y_2, \dots, y_n) \neq= \overline{x}$, dovrebbe essere $f(\overline{x}) = f(y)$, vale a dire

$$\varphi\Big(\frac{1}{n}\sum_{i=1}^n y_i\Big) = \varphi(\frac{1}{n}) = f(\overline{x}) = f(y) = \frac{1}{n}\sum_{i=1}^n \varphi(y_i)$$

e, essendo φ strettamente convessa, dovrebbe essere $y_1 = y_2 = \dots = y_n$, vale a dire $y = \overline{x}$. Un assurdo. □

e. Medie

E' opportuno fare qualche considerazione ulteriore sulle medie. Supponiamo di aver svolto un certo numero di misure sulla lunghezza L di un oggetto e di aver trovato n numeri $a_1, a_2, \dots, a_n$ distinti. Quale sarà il valore che converrà assumere per lunghezza di L? La risposta non è univoca e dipende essenzialmente dall'analisi della *distribuzione degli errori* contenuti nella misura. Comunque una media M dovrà soddisfare certi requisiti minimi

(i) la media dovrà essere certamente compresa tra il massimo e il minimo valore, $\min(a_1, a_2, \dots, a_n) \le M \le \max(a_1, a_2, \dots, a_n)$,
(ii) la media dovrà essere omogenea di grado 1 rispetto ai valori mediati. Se moltiplico tutti i valori misurati per $\lambda > 0$, ad esempio cambiando unità di misura, anche la media dovrà moltiplicarsi per λ.

Una possibilità è la *media aritmetica*

$$\overline{a} := \frac{1}{n}\sum_{k=1}^n a_k,$$

scelta che può essere giustificata dalla seguente considerazione euristica sulla funzione distribuzione degli errori.

2.20 (Media aritmetica e distribuzione degli errori) Sia $x_0 \in \mathbb{R}$ una quantità reale che viene misurata ripetutamente e accuratamente. Le misure fatte $a_1, a_2, \dots, a_n$ differiscono in generale dal valore esatto x_0. Con opportune ipotesi sulla distribuzione di probabilità degli errori, è possibile scegliere a partire dalle misure $a_1, a_2, \dots, a_n$ il valore più probabile x.

In molti casi la probabilità di avere una misura a diversa da x_0 decade con l'aumentare della distanza $|a - x_0|$ della misura a da x_0. In generale la densità di probabilità $p(a)$ di avere una misura a sarà una funzione a campana centrata in x_0,

$$p(a) = \phi((a - x_0)^2), \qquad \phi(t) \text{ decrescente in } t \text{ per } t > 0.$$

Se le misure fra loro indipendenti, si puo' dimostrare che la relativa densita' di probabilita' tende verso una distribuzione gaussiana. Assumiamo dunque che la densita' di probablita' delle misure sia gaussiana, cioe' *densità gaussiana*

$$p(a, x) = \frac{1}{\sqrt{2\pi}\,\sigma}\exp\Big(-\frac{(a-x)^2}{2\sigma^2}\Big)$$

dove $\sigma > 0$ è un parametro che controlla l'ampiezza della campana, x è il centro della campana e la costante $\frac{1}{\sqrt{2\pi}\sigma}$ è necessaria affinché

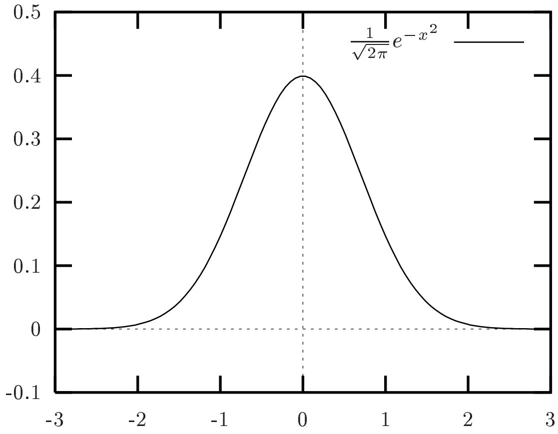

Figura 2.4. Grafico della distribuzione gaussiana $f(x) = \frac{1}{\sqrt{2\pi}e^{-x^2}}$.

$$\int_{-\infty}^{+\infty} p(a)\, da = 1.$$

Se ora si effettuano n misure indipendenti $a_1, a_2, \dots, a_n$, la densità di probabilità per questa serie di misure è

$$\begin{aligned} p(a_1, a_2, \dots, a_n, x, \sigma) &= p(a_1)p(a_2)\dots p(a_n) \\ &= \frac{1}{(2\pi)^{n/2}\,\sigma^n}\exp\Big(-\frac{1}{2\sigma^2}\sum_{k=1}^{n}(a_k - x)^2\Big). \end{aligned} \tag{2.14}$$

Se le misure sono state eseguite coscienziosamente, $p(a_1, a_2, \dots, a_n, x)$ è massima (*principio di massima verosimiglianza* per gli statistici): x e σ vanno dunque scelte in modo da massimizzare la (2.14). Se si tiene fisso σ, dovrà essere

$$\frac{dp(a_1, a_2, \dots, a_n, x, \sigma)}{dx} = 0$$

e, essendo la funzione $\exp(-s)$ decrescente, x minimizza il polinomio di secondo grado in x

$$\sum_{k=1}^{n}(x - a_k)^2$$

i.e.,

$$x = \overline{x} := \frac{1}{n}\sum_{k=1}^{n} a_k.$$

Si noti che $\overline{x}$ non dipende dall'ampiezza della campana. A questo punto dovrà anche essere

$$\frac{dp}{d\sigma}(a_1, a_2, \dots, a_n, \overline{x}, \sigma) = 0$$

e, eseguendo la derivata si trova che il miglior σ è lo *scarto quadratico medio dei dati*

$$\sigma_{\text{opt}} := \sqrt{\frac{1}{n}\sum_{k=1}^{n}(a_i - \overline{x})^2}.$$

Anche se la media aritmetica può sembrare, sulla base della discussione in 2.20, una definizione definitiva di media, tuttavia esperimenti differenti possono condurre a formule diverse per la media.

Supponiamo di misurare ripetutamente il lato di un quadrato di metallo con successive pesature $p_1, p_2, \dots, p_n$. In questo caso i lati misurati $a_1, a_2, \dots, a_k$ sono proporzionali alla radice quadrata dei pesi $p_1, p_2, \dots, p_k$. In presenza di una distribuzione gaussiana degli errori per i pesi, si prende per peso medio la media aritmetica

$$\overline{p} = \frac{1}{n}\sum_{k=1}^{n} p_k,$$

e dunque il lato medio è questa volta

$$||a||_2 := \sqrt{\frac{1}{n}\sum_{k=1}^{n} a_k^2}.$$

$||a||_2$ è la *media quadratica di* $a_1, a_2, \dots, a_n$.

Un altro caso è quello di voler misurare la lunghezza di una barra di bronzo misurandone la frequenza di vibrazione (è noto che le lunghezza di una barra vibrante è inversamente proporzionale alla frequenza di vibrazione). Se dunque le frequenze misurate ripetutamente sono $f_1, f_2, \dots, f_n$, e le rispettive lunghezze sono $a_1, a_2, \dots, a_n$, in presenza della solita distribuzione degli errori sulle frequenze si prende per frequenza media di vibrazione la media delle misure

$$\overline{f} := \frac{1}{n}\sum_{k=1}^{n} f_k$$

e dunque come media per le lunghezze la *media armonica*

$$||a||_{-1} := \Big(\frac{1}{n}\sum_{k=1}^{n}\frac{1}{a_k}\Big)^{-1}.$$

Un altro caso ancora è quello in cui le misure $a_1, a_2, \dots, a_n$ sono positive e i logaritmi delle misure $\log a_k$, $k = 1, n$ hanno una statistica gaussiana. In questo caso il valor medio dei logaritmi delle misure è la media aritmetica

$$\overline{y} = \frac{1}{n}\sum_{k=1}^{n}\log(a_k)$$

e il valor medio per le misure a_k è dunque la *media geometrica*

$$||a||_0 := \sqrt[n]{a_1 \cdot a_2 \cdot a_3 \dots a_n}.$$

Tutte queste medie sono in realtà molto legate fra loro. Siano infatti $a_1, a_2, \dots, a_n$ n numeri reali positivi. Per $p \neq 0$ si chiama *media p di* $a_1, a_2, \dots, a_n$ il numero

$$||a||_p := \Big(\frac{1}{n}(a_1^p + a_2^p + \dots + a_n^p)\Big)^{1/p} = \Big(\frac{1}{n}\sum_{k=1}^{n} a_k^p\Big)^{1/p}.$$

Casi particolari sono evidentemente la *media aritmetica*

$$||a||_1 := \frac{1}{n}\sum_{k=1}^{n} a_k,$$

la *media quadratica*

$$||a||_2 := \Big(\frac{1}{n}\sum_{k=1}^{n} a_k^2\Big)^{1/2},$$

e la *media armonica*

$$||a||_{-1} := \Big(\frac{1}{n}\sum_{k=1}^n \frac{1}{a_k}\Big)^{-1}.$$

Si ha

2.21 Proposizione. *Siano $a := (a_1, a_2, \dots, a_n)$ n numeri reali positivi. La funzione $p \to ||a||_p$ è crescente in p per $p \neq 0$. Inoltre*

(i) $||a||_p \to \max(a_1, a_2, \dots, a_n)$ *per* $p \to +\infty$,
(ii) $||a||_p \to \min(a_1, a_2, \dots, a_n)$ *per* $p \to -\infty$,
(iii) $||a||_p$ *tende alla media geometrica*

$$||a||_p \to \sqrt[n]{a_1 \cdot a_2 \cdot a_3 \dots a_n}$$

per $p \to 0$.
(iv) $||a||_p$ *è o strettamente crescente o costante. Questo ultimo caso accade se e solo se tutti i numeri* $a_1, a_2, \dots, a_n$ *sono uguali.*

Dimostrazione. Siano $0 < q < p$ applicando la diseguaglianza di Jensen alla funzione convessa $\phi(t) := t^{p/q}$, $t \geq 0$, e $\alpha_1 := a_1^q, \dots \alpha_n := a_n^q$, si trova esattamente

$$||a||_q \leq ||a||_p, \tag{2.15}$$

con l'uguaglianza se e solo se $a_1 = a_2 = \dots = a_n$. La crescenza di $||a||_p$ per $p < 0$ segue applicando la (2.15) ai numeri $(1/a_1, \dots, 1/a_n)$. $||a||_p$ è dunque crescente separatamente in $p < 0$ e $p > 0$ e strettamente crescente se e solo se i numeri $(a_1, a_2, \dots, a_n)$ non sono tutti uguali.

Proviamo ora (iii) dal che segue tra l'altro che $||a||_p$ è crescente su $\mathbb{R} \setminus \{0\}$. A tal fine, consideriamo la funzione definita per $p \neq 0$

$$f(p) := \frac{1}{n}\sum_{k=1}^n a_k^p - 1$$

Evidentemente per i teoremi sui limiti $f(p) \to 0$ per $p \to 0$ e

$$f'(p) = \frac{1}{n}\sum_{k=1}^n a_k^p \log(a_k) \to \frac{1}{n}\sum_{k=1}^n \log a_k = \sqrt[n]{a_1 \cdot a_2 \cdot a_3 \dots a_n}.$$

per $p \to 0$. Perciò il teorema di de l'Hôpital per la forma $0/0$ assicura che

$$\frac{f(p)}{p} \to \sqrt[n]{a_1 \cdot a_2 \cdot a_3 \dots a_n}.$$

e dunque anche

$$\frac{1}{p}\log(1+f(p)) = \frac{f(p)}{p}\,\frac{\log(1+f(p))}{f(p)} \to \sqrt[n]{a_1 \cdot a_2 \cdot a_3 \dots a_n},$$

i.e., la (iii).

(i) Sia $M := \max(a_1, a_2, \dots, a_n)$. Si ha evidentemente $||a||_p \leq M$ e tenendo conto che tutti gli a_k sono positivi,

$$||a||_p > \Big(\frac{M^p}{n}\Big)^{1/p} = \frac{M}{n^{1/p}} \to M$$

per $p \to +\infty$. Analogamente si procede per la (ii). □

In particolare si ritrova la diseguaglianza tra media aritmetica e media geometrica

2.22 Corollario. *Siano $a = (a_1, a_2, \dots, a_n)$ n numeri positivi. Allora la media armonica è minore della media geometrica e la media geometrica è a sua volta minore della media aritmetica,*

$$||a||_{-1} \leq \sqrt[n]{a_1 \cdot a_2 \cdot a_3 \dots a_n} \leq ||a||_1 \leq ||a||_2.$$