

# APPUNTI DI STATISTICA MATEMATICA

Domenico Marinucci

Dipartimento di Matematica, Università di Roma Tor Vergata

December 12, 2024

## Abstract

Note di un corso di Statistica Matematica, con particolare enfasi sulla Teoria Asintotica

## 1 Disuguaglianze Fondamentali

In questo capitolo richiamiamo le disuguaglianze fondamentali che costituiscono gli strumenti principali per lo sviluppo di tutta la teoria asintotica.

**Proposition 1** (*Disuguaglianza di Markov*) Sia  $X : \Omega \rightarrow \mathbb{R}$  una variabile aleatoria a valori non-negativi (cioè  $X(\omega) \geq 0$  con probabilità 1) e con valor medio finito (cioè  $\mathbb{E}[|X|] = \mathbb{E}[X] < \infty$ ). Allora  $\forall t > 0$  abbiamo

$$\mathbb{P}(X \geq t) \leq \frac{\mathbb{E}[X]}{t}.$$

**Proof.** E' sufficiente osservare che

$$\begin{aligned} \mathbb{P}(X \geq t) &= \mathbb{E}[\mathbb{I}_{[t, \infty)}(X)] = \int_t^\infty dF_X(x) \\ &\leq \int_t^\infty \frac{x}{t} dF_X(x) \leq \frac{1}{t} \int_0^\infty x dF_X(x) \\ &= \frac{\mathbb{E}[X]}{t}. \end{aligned}$$

■

**Proposition 2** (*Disuguaglianza di Markov generalizzata*) Sia  $g : \mathbb{R} \rightarrow \mathbb{R}^+$  crescente e  $X : \Omega \rightarrow \mathbb{R}$  una variabile aleatoria tale che  $\mathbb{E}[g(X)] < \infty$ . Allora

$$\Pr(X \geq t) \leq \frac{\mathbb{E}[g(X)]}{g(t)}.$$

**Proof.** La dimostrazione è praticamente identica a quella già vista della disuguaglianza di Markov. In particolare, chiamando  $Y = g(X)$  otteniamo

$$\begin{aligned}
\mathbb{P}(X \geq t) &= \mathbb{P}(g(X) \geq g(t)) = \mathbb{E}[\mathbb{I}_{[g(t), \infty)}(Y)] \\
&= \int_{g(t)}^{\infty} dF_Y(y) \leq \int_{g(t)}^{\infty} \frac{Y}{g(t)} dF_Y(y) \\
&= \int_t^{\infty} \frac{g(x)}{g(t)} dF_X(x) \leq \frac{1}{g(t)} \int_0^{\infty} g(x) dF_X(x) \\
&= \frac{\mathbb{E}[g(X)]}{g(t)} .
\end{aligned}$$

■

**Remark 3** La disuguaglianza di Markov e la sua versione generalizzata sono il primo esempio di disuguaglianze di concentrazione; in pratica, implicano che tanto più una variabile aleatoria ammette momenti finite, tanto più le code della sua distribuzione vanno a zero velocemente. Ad esempio, consideriamo una variabile aleatoria che ammetta funzione generatrice dei momenti finita in un intorno dell'origine, cioè tale per cui

$$m_X(u) := \mathbb{E}[\exp(uX)] < \infty \text{ per } u \in [0, T], \quad T > 1 .$$

Segue immediatamente che

$$\Pr(X \geq t) \leq \frac{\mathbb{E}[\exp(X)]}{\exp(t)} = \frac{m_X(1)}{\exp(t)} ,$$

cioè le code della distribuzione di  $X$  devono decadere almeno esponenzialmente.

**Proposition 4** (Disuguaglianza di Chebyshev) Sia  $X : \Omega \rightarrow \mathbb{R}$  una variabile aleatoria con momento secondo finito (cioè  $\mathbb{E}[X^2] < \infty$ ). Allora  $\forall t > 0$  abbiamo

$$\mathbb{P}(|X - \mathbb{E}[X]| \geq t) \leq \frac{\text{Var}[X]}{t^2} .$$

**Proof.** E' sufficiente osservare che, utilizzando la disuguaglianza di Markov per  $Y = |X - \mathbb{E}[X]|^2$

$$\begin{aligned}
\mathbb{P}(|X - \mathbb{E}[X]| \geq t) &= \mathbb{P}(|X - \mathbb{E}[X]|^2 \geq t^2) \\
&= \frac{\mathbb{E}[|X - \mathbb{E}[X]|^2]}{t^2} .
\end{aligned}$$

■

**Remark 5** Come discusso nei cenni storici più avanti, la disuguaglianza di Chebyshev è precedente a quella di Markov; in effetti Markov è stato un alunno di Chebyshev a San Pietroburgo.

**Example 6** (*Legge debole dei grandi numeri*) Anticipando un poco la discussione nei prossimi capitoli, possiamo illustrare immediatamente l'applicazione più importante della disuguaglianza di Chebyshev. Siano infatti  $X_i : \Omega \rightarrow \mathbb{R}$  variabili aleatorie indipendenti con momento secondo finito, e definiamo come al solito valor medio e varianza  $\mu := \mathbb{E}[X]$ ,  $\sigma^2 := \mathbb{E}[(X - \mu)^2]$ . Introduciamo altresì la media aritmetica  $\bar{X}_n := \frac{1}{n} \sum_{i=1}^n X_i$ ; dalla disuguaglianza di Chebyshev abbiamo immediatamente che, per ogni  $\varepsilon > 0$

$$\Pr \{ |\bar{X}_n - \mu| > \varepsilon \} \leq \frac{\text{Var} \{ \bar{X}_n \}}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2} \rightarrow 0 \text{ per } n \rightarrow \infty .$$

E' immediato verificare che l'ipotesi di indipendenza può essere generalizzata a quella di incorrelazione; anche le ipotesi di identica distribuzione può essere facilmente abbandonata, come vedremo nei prossimi capitoli.

Nelle precedenti disuguaglianze abbiamo mostrato che per variabili aleatorie che ammettono un numero di momenti maggiore le code della funzione di distribuzione decadono più velocemente. E' pertanto naturale chiedersi se si possano ottenere risultati più stringenti focalizzandosi su classi di variabili aleatorie che abbiano supporto compatto, risultando quindi uniformemente limitate. La risposta è affermativa, come mostrato nel prossimo risultato.

**Proposition 7** (*Disuguaglianza di Hoeffding*) Sia  $X_1, \dots, X_n$  una successione di variabili aleatorie indipendenti e tali che esistano due sequenze di numeri reali  $a_i, b_i$ ,  $i = 1, 2, \dots, n$  per cui valga  $a_i \leq X_i \leq b_i$  per ogni  $i$ ; assumiamo inoltre che  $\mathbb{E}[Y_i] = 0$  per  $i = 1, \dots, n$ . Per ogni  $t > 0$  vale la disuguaglianza

$$\Pr \left\{ \sum_{i=1}^n X_i \geq t \right\} \leq \inf_{u \geq 0} \left\{ e^{-ut} \prod_{i=1}^n e^{u^2(b_i - a_i)^2/8} \right\} .$$

**Proof.** Per la disuguaglianza di Markov abbiamo, quale che sia  $u > 0$

$$\begin{aligned} \Pr \left\{ \sum_{i=1}^n X_i \geq t \right\} &= \Pr \left\{ \sum_{i=1}^n u X_i \geq ut \right\} \\ &= \Pr \left\{ e^{u \sum_{i=1}^n X_i} \geq e^{ut} \right\} \\ &\leq e^{-ut} \mathbb{E} \left[ e^{u \sum_{i=1}^n X_i} \right] \\ &= e^{-ut} \prod_{i=1}^n \mathbb{E} \left[ e^{u X_i} \right] . \end{aligned}$$

Scriviamo ora

$$Y_i = \alpha b_i + (1 - \alpha) a_i, \quad \alpha = \frac{Y_i - a_i}{b_i - a_i} ;$$

per la convessità della funzione esponenziale, si ha che

$$\begin{aligned} e^{u X_i} &\leq \frac{Y_i - a_i}{b_i - a_i} e^{u b_i} + \left( 1 - \frac{Y_i - a_i}{b_i - a_i} \right) e^{u a_i} \\ &= \frac{Y_i - a_i}{b_i - a_i} e^{u b_i} + \frac{b_i - Y_i}{b_i - a_i} e^{u a_i} , \end{aligned}$$

da cui, ricordando che  $\mathbb{E}[Y_i] = 0$

$$\mathbb{E} [e^{uX_i}] \leq -\frac{a_i}{b_i - a_i} e^{ub_i} + \frac{b_i}{b_i - a_i} e^{ua_i} .$$

Scriviamo ora

$$-\frac{a_i}{b_i - a_i} e^{ub_i} + \frac{b_i}{b_i - a_i} e^{ua_i} = e^{g_i(v_i)}$$

dove

$$g_i(v_i) := -\gamma_i v_i + \log(1 - \gamma_i + \gamma_i e^{v_i}) , \quad \gamma_i = -\frac{a_i}{b_i - a_i} , \quad v_i := u(b_i - a_i) ;$$

notiamo infatti che

$$\begin{aligned} e^{g_i(v_i)} &= e^{-\gamma_i v_i} \left(1 + \frac{a_i}{b_i - a_i} - \frac{a_i}{b_i - a_i} e^{u(b_i - a_i)}\right) \\ &= e^{ua_i} \left(1 + \frac{a_i}{b_i - a_i} - \frac{a_i}{b_i - a_i} e^{u(b_i - a_i)}\right) \\ &= \left(\frac{b_i - a_i}{b_i - a_i} e^{ua_i} + \frac{a_i}{b_i - a_i} e^{ua_i} - \frac{a_i}{b_i - a_i} e^{ub_i}\right) \\ &= \frac{b_i}{b_i - a_i} e^{ua_i} - \frac{a_i}{b_i - a_i} e^{ub_i} . \end{aligned}$$

Notiamo che  $g_i(0) = 0$  ed inoltre

$$g'_i(0) = -\gamma_i + \frac{\gamma_i e^{v_i}}{(1 - \gamma_i + \gamma_i e^{v_i})} \Big|_{v_i=0} = 0 ,$$

$$g''_i(v) = \frac{\gamma_i e^v (1 - \gamma_i + \gamma_i e^{v_i}) - \gamma_i^2 e^{2v}}{(1 - \gamma_i + \gamma_i e^{v_i})^2} = \frac{\gamma_i e^v (1 - \gamma_i)}{(1 - \gamma_i + \gamma_i e^v)^2} \leq \frac{1}{4} \text{ per ogni } \gamma_i, v .$$

Infatti, poiché  $\gamma_i, v > 0$

$$\begin{aligned} \frac{\gamma_i e^v (1 - \gamma_i)}{(1 - \gamma_i + \gamma_i e^v)^2} &= \frac{\gamma_i (1 - \gamma_i)}{e^{v/2}} \frac{1}{(1 - \gamma_i + \gamma_i e^{v/2})^2} \\ &\leq \frac{1}{4} \frac{1}{(1 + \gamma_i (e^{v/2} - 1))^2} \leq \frac{1}{4} . \end{aligned}$$

Per il teorema del valor medio di Lagrange, esiste  $\xi \in (0, v_i)$  tale per cui

$$\begin{aligned} g_i(v_i) &= g_i(0) + g'_i(0)v_i + g''_i(\xi) \frac{v_i^2}{2} \\ &= g''_i(\xi) \frac{v_i^2}{2} \leq \frac{u^2 (b_i - a_i)^2}{8} . \end{aligned}$$

Ne segue che

$$\mathbb{E} [e^{tX_i}] \leq e^{g_i(v_i)} \leq e^{u^2 (b_i - a_i)^2 / 8} ,$$

ed la Proposizione è dimostrata. ■

**Corollary 8** (*Media aritmetica di variabili aleatorie di Bernoulli*) Siano  $Y_1, \dots, Y_n$  variabili aleatorie indipendenti identicamente distribuite con legge di Bernoulli  $Ber(p)$ . Per ogni  $\varepsilon > 0$ , abbiamo che

$$\Pr \{ |\bar{Y}_n - p| > \varepsilon \} \leq 2e^{-2n\varepsilon^2}.$$

**Proof.** Si prenda  $X_i := \frac{1}{n}(Y_i - p)$ . E' immediato verificato che  $\mathbb{E}[X_i] = 0$  e  $a_i \leq X_i \leq b_i$  per ogni  $i = 1, \dots, n$ , con  $a_i = -\frac{p}{n}$  e  $b_i = \frac{1-p}{n}$ , da cui  $(b_i - a_i)^2 = \frac{1}{n^2}$ . Applicando la disuguaglianza di Hoeffding si ha

$$\Pr \{ \bar{Y}_n - p > \varepsilon \} \leq \inf_u \left\{ e^{-u\varepsilon} \prod_{i=1}^n e^{u^2/8n^2} \right\} = \inf_u \left\{ e^{-u\varepsilon} e^{u^2/8n} \right\}$$

e prendendo  $u = 4n\varepsilon$  otteniamo

$$\inf_u \left\{ e^{-u\varepsilon} e^{u^2/8n} \right\} \leq e^{-4n\varepsilon \times \varepsilon} e^{16n^2\varepsilon^2/8n} = e^{-2n\varepsilon^2}.$$

La dimostrazione è completata ripetendo lo stesso ragionamento per  $\Pr \{ \bar{Y}_n - p < -\varepsilon \}$ .

■

**Remark 9** (*Un confronto tra la disuguaglianza di Chebyshev e quella di Hoeffding*). Nel caso di variabili aleatorie limitate, la disuguaglianza di Hoeffding è enormemente più efficiente della disuguaglianza di Chebyshev. Ad esempio, nel caso di variabili Bernoulliane di parametro  $p = \frac{1}{2}$  con  $n = 100$  ed  $\varepsilon = 0.2$  la disuguaglianza di Chebyshev ci dà il seguente limite superiore:

$$\Pr \{ |\bar{Y}_n - p| > 0.2 \} \leq \frac{\text{Var} \{ \bar{Y}_n \}}{\varepsilon^2} = \frac{1}{4n\varepsilon^2} \simeq 0.0625$$

mentre con la disuguaglianza di Hoeffding si ottiene

$$\Pr \{ |\bar{Y}_n - p| > 0.2 \} \leq 2e^{-2 \times 100 \times (0.2)^2} \simeq 0.00067.$$

**Remark 10** (*Concentrazione del volume di un ipercubo*). La disuguaglianza di Hoeffding ammette una interessante interpretazione geometrica quando viene applicata al comportamento del volume di un ipercubo  $[0, 1]^n$ , nel limite in cui  $n \rightarrow \infty$ . Consideriamo infatti l'iperpiano  $\Gamma_n := \{x_i : x_1 + x_2 + \dots + x_n = n/2\}$ ; definiamo un "Tubo" di raggio  $\varepsilon$  intorno a  $\Gamma_n$  come

$$\text{Tub}_\varepsilon(\Gamma_n) := \{ \mathbf{x} \in \mathbb{R}^n : \text{dist} \{ \mathbf{x}, \Gamma_n \} \leq \varepsilon \}.$$

Per ogni  $\varepsilon > 0$ , il volume dell'intersezione tra  $\text{Tub}_\varepsilon(\Gamma_n)$  e  $[0, 1]^n$  converge esponenzialmente a 1 quando  $n$  diverge all'infinito; in altre parole, la massa del cubo si concentra esponenzialmente in un intorno arbitrariamente piccolo della diagonale principale.

Le precedenti disuguaglianze ci permettono di avere dei limiti superiori per il comportamento delle code di variabili aleatorie generiche. E' interessante confrontare il loro comportamento con quello delle code di una variabile Gaussiana standard; a questo scopo introduciamo la prossima proposizione.

**Proposition 11** (*Disuguaglianza di Mill-Gordon*) Sia  $Z \sim N(0, 1)$  una variabile Gaussiana standard. Si ha che

$$\sqrt{\frac{1}{2\pi}} \frac{t}{1+t^2} \exp\left(-\frac{t^2}{2}\right) \leq \Pr\{Z > t\} \leq \sqrt{\frac{1}{2\pi}} \frac{1}{t} \exp\left(-\frac{t^2}{2}\right).$$

**Proof.** La disuguaglianza di destra può essere stabilita con un semplice cambio di variabile; abbiamo:

$$\begin{aligned} \Pr\{Z > t\} &= \int_t^\infty \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx \\ &\leq \int_t^\infty \frac{x}{t} \exp\left(-\frac{x^2}{2}\right) dx \\ &= \frac{1}{t} \frac{1}{\sqrt{2\pi}} \int_{t^2/2}^\infty \exp(-y) dy \\ &= \frac{1}{t} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right). \end{aligned}$$

dopo il cambio di variabile  $y = \frac{x^2}{2}$ . Per il limite inferiore, possiamo ragionare come segue:

$$\begin{aligned} \Pr\{Z > t\} &= \int_t^\infty \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx \\ &\geq \int_t^\infty \frac{x^4 + 2x^2 - 1}{x^4 + 2x^2 + 1} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx \\ &= \int_t^\infty \frac{x^2(x^2 + 1) + x^2 - 1}{(x^2 + 1)^2} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx \\ &= \int_t^\infty \left\{ \frac{x^2}{(x^2 + 1)^2} + \frac{x^2 - 1}{(x^2 + 1)^2} \right\} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx. \end{aligned}$$

Chiamiamo ora

$$\psi(x) = \frac{x}{1+x^2} \exp\left(-\frac{x^2}{2}\right)$$

e notiamo che

$$\begin{aligned} \psi'(x) &= \frac{d\psi(x)}{dx} = \frac{1+x^2-2x^2}{(1+x^2)^2} \exp\left(-\frac{x^2}{2}\right) - \frac{x^2}{1+x^2} \exp\left(-\frac{x^2}{2}\right) \\ &= -\left\{ \frac{x^2}{(x^2+1)^2} + \frac{x^2-1}{(x^2+1)^2} \right\} \exp\left(-\frac{x^2}{2}\right). \end{aligned}$$

La dimostrazione è dunque completata dal Teorema Fondamentale del Calcolo:

$$-\frac{1}{\sqrt{2\pi}} \int_t^\infty \psi'(x) dx = -\frac{1}{\sqrt{2\pi}} [\psi(x)]_t^\infty = \sqrt{\frac{1}{2\pi}} \frac{t}{1+t^2} \exp\left(-\frac{t^2}{2}\right).$$

■

**Remark 12** Queste due disuguaglianze sono piuttosto efficienti; ad esempio, per  $t = 2$  abbiamo

$$\Pr\{Z > 2\} = \int_2^\infty \frac{1}{\sqrt{2\pi}} \exp(-\frac{x^2}{2}) dx = 0.02275$$

mentre le stime di Mill danno

$$\frac{1}{\sqrt{2\pi}} \frac{t}{1+t^2} \exp(-\frac{t^2}{2}) \Big|_{t=2} = 0.021596 \leq \Pr\{Z > 2\} \leq \frac{1}{\sqrt{2\pi}} \frac{1}{t} \exp(-\frac{t^2}{2}) \Big|_{t=2} = 0.027 .$$

Per  $t = 3$  il risulta diventa ancora più preciso; otteniamo

$$\Pr\{Z > 3\} = \int_3^\infty \frac{1}{\sqrt{2\pi}} \exp(-\frac{x^2}{2}) dx = 0.001349$$

mentre le stime di Mill danno

$$\begin{aligned} \frac{1}{\sqrt{2\pi}} \frac{t}{1+t^2} \exp(-\frac{t^2}{2}) \Big|_{t=3} &= 0.001329 \\ &\leq \Pr\{Z > 3\} \\ &\leq \frac{1}{\sqrt{2\pi}} \frac{1}{t} \exp(-\frac{t^2}{2}) \Big|_{t=3} = 0.0014773 . \end{aligned}$$

**Remark 13** Vediamo quale risultato otterremo per la somma di variabili Bernoulliane considerando l'approssimazione asintotica che segue dal teorema del Limite Centrale; prendendo per semplicità  $p = \frac{1}{2}$  abbiamo:

$$\bar{Y}_n - \mathbb{E}[\bar{Y}_n] \simeq \mathcal{N}(0, \frac{p(1-p)}{n}) = \mathcal{N}(0, \frac{1}{4n})$$

ed utilizzando la disuguaglianza di Mill abbiamo dunque

$$\begin{aligned} \Pr\left\{|\bar{Y}_n - \frac{1}{2}| > 0.2\right\} &= \Pr\left\{|\mathcal{N}(0, \frac{1}{4n})| > 0.2\right\} + o_{n \rightarrow \infty}(1) \\ &= \Pr\{|\mathcal{N}(0, 1)| > 0.2 \times 2\sqrt{n}\} \\ &\leq \frac{2}{\sqrt{2\pi}} \frac{1}{0.2 \times 2\sqrt{n}} e^{-4 \times n \times (0.2)^2 / 2} . \end{aligned}$$

Per  $n = 100$  otteniamo

$$\frac{2}{\sqrt{2\pi}} \frac{1}{0.2 \times 20} e^{-2 \times 100 \times (0.2)^2} = 6.6915 \times 10^{-5} ,$$

un valore circa 10 volte inferiore a quello ottenuto dalla disuguaglianza di Hoeffding. Il risultato però non deve ingannare: la disuguaglianza di Hoeffding vale in senso stretto, mentre qui stiamo trovando un limite superiore alla probabilità Gaussiana, tralasciando il fatto che il termine di approssimazione  $o_{n \rightarrow \infty}(1)$  è molto più grande della maggiorazione ottenuta: si può dimostrare solo l'ordine  $O(n^{-1/2})$ . In altre parole, il Teorema del Limite Centrale NON può essere sfruttato per ottenere un limite così efficiente per la probabilità sulle code della variabile  $\bar{Y}_n$ .

La prossima disuguaglianza dovrebbe essere ben nota dai corsi di Geometria ed Analisi.

**Proposition 14** (*Disuguaglianza di Cauchy-Schwartz*) Siano  $X, Y : \Omega \rightarrow \mathbb{R}$  variabili aleatorie con momento secondo finito. Abbiamo che

$$|\mathbb{E}[XY]|^2 \leq \mathbb{E}[X^2]\mathbb{E}[Y^2] .$$

La disuguaglianza è stretta, a meno che le variabili aleatorie siano tra loro in relazione lineare con probabilità 1, cioè esista un numero reale  $t^*$  tale per cui

$$\mathbb{E}[(Y - t^*X)^2] = 0 .$$

**Proof.** Consideriamo la funzione quadratica (in  $t$ )

$$g(t) : \mathbb{E}[(Y - tX)^2] = \mathbb{E}[Y^2] - 2t\mathbb{E}[XY] + t^2\mathbb{E}[X^2] \geq 0 .$$

L'equazione  $g(t) = 0$  ovviamente ammette al più una radice reale di molteplicità due; quindi il discriminante deve avere valore non-positivo, in particolare

$$4(\mathbb{E}[XY])^2 - 4\mathbb{E}[X^2]\mathbb{E}[Y^2] \leq 0$$

da cui segue la disuguaglianza, che diviene una uguaglianza se e solo se esiste  $t^*$  tale per cui  $g(t^*) = 0$ . ■

**Remark 15** *Il richiamo alla geometria ci anticipa una idea che avrà grande importanza nei corsi di probabilità più avanzati: lo spazio delle variabili aleatorie di momento secondo finito può essere visto come uno spazio vettoriale dotato di un prodotto interno (prodotto scalare) definito da*

$$\langle X, Y \rangle := \mathbb{E}[XY] .$$

*Nel caso particolare di variabili con valor medio nullo questo prodotto interno è la covarianza.*

Per la prossima disuguaglianza ricordiamo innanzitutto che una funzione  $g(\cdot) : \mathbb{R} \rightarrow \mathbb{R}$  si dice convessa se per ogni  $\alpha \in (0, 1)$  e per ogni  $x_1$  e  $x_2$  nel suo dominio (connesso), si ha che

$$g(\alpha x_1 + (1 - \alpha)x_2) \leq \alpha g(x_1) + (1 - \alpha)g(x_2) .$$

Data una funzione convessa per ogni punto  $x_0$  esiste una costante  $L = L_{x_0}$  tale per cui

$$g(x) \geq g(x_0) + L(x - x_0) ;$$

la costante non è necessariamente unica (se la funzione è derivabile, coincide con il valore della derivata in quel punto).

**Proposition 16** (*Disuguaglianza di Jensen*) Sia  $X$  una variabile aleatoria con valor medio finito e  $g(\cdot)$  una funzione convessa. Abbiamo che

$$\mathbb{E}[g(X)] \geq g(\mathbb{E}[(X)]) ,$$

dove la grandezza a sinistra della disuguaglianza può essere infinita.

**Proof.** E' sufficiente osservare che, per la monotonia del valor medio e prendendo  $x_0 = \mathbb{E}[(X)]$

$$\begin{aligned} \mathbb{E}[g(X)] &\geq \mathbb{E}[g(x_0) + L(X - x_0)] = \mathbb{E}[g(x_0)] + \mathbb{E}[L(X - x_0)] \\ &= g(\mathbb{E}[(X)]) + L\mathbb{E}[(X - \mathbb{E}[(X)])] = g(\mathbb{E}[(X)]) . \end{aligned}$$

■