DISUGUAGLIANZA INTEGRALE DI JENSEN IN DIMENSIONE FINITA

LIBOR VESELY

1. Prima disuguaglianza integrale di Jensen

1.1. **Motivazione.** Siano C un insieme convesso in uno spazio vettoriale, $f\colon C \to (-\infty, +\infty]$ una funzione convessa. Secondo la disuguaglianza ("finita") di Jensen, se $x_1, \ldots, x_n \in C$, $\lambda_1, \ldots, \lambda_n \geq 0$ e $\sum_1^n \lambda_i = 1$, allora

$$f\left(\sum_1^n \lambda_i x_i\right) \leq \sum_1^n \lambda_i f(x_i)\,.$$

Consideriamo la misura $\mu = \sum_1^n \lambda_i \delta_{x_i}$, dove δ_z denota la misura di Dirac del punto z. Allora μ è una misura di probabilità (ad es., su C), e la disuguaglianza di Jensen diventa

$$f\left(\int_C x\, d\mu(x)\right) \leq \int_C f(x)\, d\mu(x)\,.$$

In questo testo, presenteremo una generalizzazione dell'ultima disuguaglianza a misure di probabilità su C più generali. Per semplicità ci limitiamo agli spazi finito-dimensionali.

integrazione

1.2. **Integrazione di funzioni vettoriali.** Siano (C, Σ, μ) uno spazio di probabilità con $C \subset \mathbb{R}^d$, $g\colon C \to \mathbb{R}^m$ una funzione (vettoriale) Σ-misurabile (nel senso che le controimmagini degli aperti appartengono a Σ), e $e_1^*, \ldots, e_m^* \in (\mathbb{R}^m)^*$ i funzionali delle coordinate, diciamo che $g \in L_1(\mu)$ se $e_i^* \circ g \in L_1(\mu)$ per ogni i. In tal caso definiamo l'integrale di g semplicemente per coordinate:

$$\int_C g\, d\mu := \left(\int_C (e_1^* \circ g)\, d\mu\,, \ldots,\, \int_C (e_m^* \circ g)\, d\mu\right).$$

Si noti che questa definizione dice esattamente che

$$e_i^*\left(\int_C g\, d\mu\right) = \int_C (e_i^* \circ g)\, d\mu \quad \text{per ogni } i = 1, \ldots, m.$$

Siccome ogni $\ell \in (\mathbb{R}^m)^*$ è una combinazione lineare dei funzionali delle coordinate, otteniamo:

$$\ell\left(\int_C g\, d\mu\right) = \int_C (\ell \circ g)\, d\mu \quad \text{per ogni } \ell \in (\mathbb{R}^m)^*.$$

Lasciamo come semplice **esercizio per il lettore** la dimostrazione del fatto che *le seguenti affermazioni sono equivalenti:*

(i) $g \in L_1(\mu)$*;*
(ii) $e_i^* \circ g \in L_1(\mu)$ *per ogni* $i = 1, \dots, d$*;*
(iii) $\ell \circ g \in L_1(\mu)$ *per ogni* $\ell \in (\mathbb{R}^d)^*$*;*
(iv) $\|g\| \in L_1(\mu)$*.*

Inoltre, in tal caso, $\left\|\int_C g\,d\mu\right\| \le \int_C \|g\|\,d\mu$ *per ogni norma* $\|\cdot\|$ *su* $\mathbb{R}^m$*.*

ipotesi

1.3. **Le ipotesi base.** Sia $C \subset \mathbb{R}^d$ un *insieme convesso.* Per poter integrare la funzione $x \mapsto x$, abbiamo bisogno che la misura sia definita almeno sui sottoinsiemi boreliani di C. Anche se C può non essere boreliano (a meno che non si sia nel caso $C \subset \mathbb{R}$), possiamo considerare la σ-algebra $\mathcal{B}(C)$ dei suoi *boreliani relativi*, cioè, i boreliani dello spazio metrico C. Ovviamente,

$$\mathcal{B}(C) = \{B \cap C : B \in \mathit{Borel}(\mathbb{R}^d)\}.$$

Supponiamo quindi che μ *sia una misura di probabilità definita su una* σ*-algebra* Σ *di sottoinsiemi di* C*, tale che*

$$\mathcal{B}(C) \subset \Sigma.$$

1.4. **Il baricentro di una misura.** Supponiamo che $C \subset \mathbb{R}^d$ e la misura di probabilità μ soddisfino le ipotesi base 1.3. Il *baricentro* di μ è (se esiste) il punto

$$x_\mu := \int_C x\,d\mu(x) \ \in \mathbb{R}^d.$$

Per quanto detto in 1.2,

$$x_\mu \text{ esiste se e solo se } \|\cdot\| \in L_1(\mu).$$

In particolare, x_μ esiste se C è limitato.

esercizio

Remark 1.1 (Esercizio). Supponiamo che x_μ esista. Allora

$$a(x_\mu) = \int_C a\,d\mu \quad \text{per ogni funzione affine } a\colon \mathbb{R}^d \to \mathbb{R}.$$

(Suggerimento: scrivere $a = \ell + \beta$ dove ℓ è lineare e β costante.)

Prima di dimostrare che $x_\mu \in C$, ricordiamo il seguente teorema di separazione.

separazione

Theorem 1.2. *Siano* $A, B \subset \mathbb{R}^d$ *due insiemi convessi tali che*

$$(\mathrm{ri}\,A) \cap (\mathrm{ri}\,B) = \emptyset$$

(dove $\mathrm{ri}\,E$ *denota l'inerno relativo di* E*). Allora esiste* $\ell \in (\mathbb{R}^d)^*$ *tale che*

$$\ell(a) < \ell(b) \quad \text{per ogni } a \in \mathrm{ri}\,A,\ b \in \mathrm{ri}\,B.$$

In particolare, $\sup \ell(A) \le \inf \ell(B)$*.*

baricentro **Theorem 1.3.** *Siano $C \subset \mathbb{R}^d$ un insieme convesso, μ una misura di probabilità su C definita su na σ-algebra contenente $\mathcal{B}(C)$. Supponiamo che $\|\cdot\| \in L_1(\mu)$. Allora il baricentro x_μ appartiene a C.*

Proof. Procediamo per induzione rispetto alla dimensione di C. Se $\dim C = 0$, abbiamo $C = \{c\}$ e quindi $\mu = \delta_c$, $x_\mu = c \in C$.

Passo induttivo. Sia ora $n \geq 0$ un intero tale che il teorema valga per insiemi convessi di dimensione $\leq n$. Sia $\dim C = n+1$.

Procedendo per assurdo, supponiamo che $x_\mu \notin C$. Applicando il Teorema 1.2 agli insiemi convessi C e $\{x_\mu\}$, otteniamo $\ell \in (\mathbb{R}^d)^*$ tale che

$$\ell(c) < \ell(x_\mu) \quad \text{per ogni } c \in \operatorname{ri} C.$$

Abbiamo $\ell - \ell(x_\mu) \geq 0$ su C, ma anche

$$\int_C [\ell - \ell(x_\mu)]\, d\mu = x_\mu - x_\mu = 0.$$

Per il teorema di annullamento, dobbiamo avere $\ell(x) = \ell(x_\mu)$ per μ-quasi ogni punto $x \in C$. In altre parole, μ è concentrata sull'insieme convesso $C_1 := C \cap [\ell = \ell(x_\mu)]$. Visto che C_1 non interseca $\operatorname{ri} C$, si ha $\dim C_1 \leq n$. Per la nostra ipotesi induttiva, $x_\mu \in C_1 \subset C$; e questo assurdo completa la dimostrazione. $\square$

1.5. **Disuguaglianza integrale di Jensen.** Ci servirà il seguente lemma.

minorazione **Lemma 1.4.** *Siano $C \subset \mathbb{R}^d$ un insieme convesso, $f\colon C \to \mathbb{R}$ una funzione convessa, $x_0 \in \operatorname{ri} C$. Supponiamo che $-\infty < t_0 < f(x_0)$. Allora esiste una funzione affine $a\colon \mathbb{R}^d \to \mathbb{R}$ tale che $a \leq f$ in C, e $t_0 < a(x_0)$.*

Proof. Fissiamo un qualsiasi numero reale $t_1 > f(x_0)$. È facile vedere che il punto (x_0, t_1) appartiene all'iterno relativo dell'epigrafo

$$\operatorname{epi} f := \{(x,t) \in C \times \mathbb{R} : t \geq f(x)\} \subset \mathbb{R}^d \times \mathbb{R}.$$

Sia $\varepsilon > 0$ tale che $t_0 + \varepsilon < f(x_0)$. Applicando il Teorema 1.2 agli insiemi convessi $\operatorname{epi} f$ e $\{(x_0, t_0+\varepsilon)\}$, otteniamo l'esistenza di $\Lambda \in (\mathbb{R}^d \times \mathbb{R})^*$ tale che $\Lambda(x_0, t_0+\varepsilon) \leq \alpha := \inf \Lambda(\operatorname{epi} f)$. Siccome Λ separa strettamente i punti (x_0, t_0) e (x_0, t_1), sappiamo che l'iperpiano $[\Lambda = \alpha] \subset \mathbb{R}^d \times \mathbb{R}$ coincide con il grafico di una funzione affine $a\colon \mathbb{R}^d \to \mathbb{R}$. Quindi,

$$t_0 < t_0 + \varepsilon \leq a(x_0) \quad \text{e} \quad a|_C \leq f.$$

$\square$

Jensen1 **Theorem 1.5** (disuguaglianza di Jensen). *Siano $C \subset \mathbb{R}^d$ un insieme convesso, μ una misura di probabilità su C definita su una σ-algebra Σ, $f\colon C \to (-\infty, +\infty]$ una funzione convessa Σ-misurabile. Supponiamo che $\mathcal{B}(C) \subset \Sigma$ e $\|\cdot\| \in L_1(\mu)$. Allora:*

(a) il baricentro x_μ esiste ed appartiene a C;
(b) l'integrale $\int_C f\, d\mu$ esiste e appartiene a $(-\infty, +\infty]$;
(c) vale la disuguaglianza

$$f(x_\mu) \leq \int_C f\, d\mu\,.$$

Proof. (a) è stato già dimostrato. Dimostriamo ora (b). Il caso $f \equiv +\infty$ è ovvio; supponiamo quindi che $f \not\equiv +\infty$. L'insieme $D := [f < +\infty] \subset C$ è convesso e non vuoto, e quindi $\mathrm{ri}\, D \neq \emptyset$. Una facile applicazione del Lemma 1.4 alla funzione $f|_D$ mostra che esiste una funzione affine a su $\mathbb{R}^d$ tale che $a \leq f$. Le parti negative soddisfano $0 \leq f^- \leq a^-$; inoltre, $a \in L_1(\mu)$ in quanto $L_1(\mu)$ contiene le costanti e gli elementi di $(R^d)^*$. Ne segue che $f^- \in L_1(\mu)$ e vale (b). Rimane da dimostrare (c).

Se $t_0 := \int_C f\, d\mu = +\infty$, la disuguaglianza in (c) è ovvia. Sia quindi $t_0 \in \mathbb{R}$. Deve essere $\mu([f = +\infty]) = 0$, e quindi possiamo supporre che $f < +\infty$ su tutto C (sostituendo C con l'insieme convesso $[f < +\infty]$).

Procediamo per induzione rispetto alla dimensione di C, come nel Teorema 1.3. Il caso di $\dim C = 0$ è di nuovo ovvio. Per il passo induttivo, supponiamo che il teorema valga per gli insiemi convessi di dimensione $\leq n$, e supponiamo che $\dim C = n + 1$. Se $x_\mu \notin \mathrm{ri}\, C$, il Teorema 1.2 ci fornisce un $\ell \in (\mathbb{R}^d)^*$ tale che $\ell(x_\mu) > \ell(x)$ per ogni $x \in \mathrm{ri}\, C$. Come nel Teorema 1.3 si deduce μ è concentrata su $C_1 := C \cap [\ell = \ell(x_\mu)]$ che è un insieme convesso di dimensione $\leq n$. Applicando la nostra ipotesi induttiva, otteniamo subito la disuguaglianza in (c).

Sia quindi $x_\mu \in \mathrm{ri}\, C$. Vogliamo dimostrare che $f(x_\mu) \leq t_0$. Supponiamo per assurdo che $t_0 < f(x\mu)$. Per il Lemma 1.4, esiste una funzione affine a su $\mathbb{R}^d$ tale che $a \leq f$ su C e $t_0 < a(x_\mu)$. Possiamo scrivere $a = \ell + \beta$ dove $\ell \in (\mathbb{R}^d)^*$, $\beta \in \mathbb{R}$. Applicando l'esercizio 1.1, otteniamo

$$t_0 = \int_C f\, d\mu \geq \int_C a\, d\mu = a(x\mu) > t_0,$$

che è una contraddizione. $\square$

1.6. **Un'applicazione: disuguaglianza di Jensen per serie.** Consideriamo due successioni: $\{x_n\} \subset \mathbb{R}^d$ e $\{\lambda_n\} \subset [0, +\infty)$ con $\sum_1^{+\infty} \lambda_n = 1$. *Supponiamo che*

$$\sum_1^{+\infty} \lambda_n \|x_n\| < +\infty.$$

Allora:

(a) il punto $\bar{x} := \sum_1^{+\infty} \lambda_n x_n$ appartiene a $C := \mathrm{conv}\{x_n\}_{n\in\mathbb{N}}$;
(b) la serie $\sum_1^{+\infty} \lambda_n f(x_n)$ ammette somma (finita o $+\infty$);

(c) vale la disuguaglianza

$$f(\bar{x}) \leq \sum_{1}^{+\infty} \lambda_n f(x_n) \,.$$

Per dimostrarlo, è sufficiente considerare la misura di probabilità

$$\mu := \sum_{1}^{+\infty} \lambda_n \delta_{x_n}$$

su C, definita su $\Sigma := 2^C$. Più precisamente:

$$\mu(E) = \sum_{x_n \in E} \lambda_n \qquad (E \subset C).$$

Applicare la disuguaglianza integrale di Jensen.

2. Seconda disuguaglianza integrale di Jensen

2.1. **Il "push forward" di una misura.** Nella seguente proposizione definiamo la *misura immagine* di una misura μ, o il *push forward* di μ, una tecnica semplice ma importante nella teoria della misura.

pf **Proposition 2.1** (push forward)**.** *Siano $(\Omega, \mathcal{A}, \mu)$ uno spazio di misura (nonnegativa), (C, Σ) uno spazio misurabile, $g \colon \Omega \to C$ una funzione ($\mathcal{A}$-Σ)–misurabile (nel senso che $g^{-1}(E) \in \mathcal{A}$ per ogni $E \in \Sigma$). Definiamo un'applicazione $\nu \colon \Sigma \to [0, +\infty]$ con*

$$\nu(E) = \mu(g^{-1}(E)) \qquad (E \in \Sigma).$$

Allora ν è una misura e $\nu(C) = \mu(\Omega)$. Inoltre, vale la seguente regola di cambio di variabile. Data una funzione $f \colon C \to \overline{\mathbb{R}}$ che sia Σ-misurabile,

$$\int_C f \, d\nu \text{ esiste} \quad \Leftrightarrow \quad \int_\Omega (f \circ g) \, d\mu \text{ esiste}$$

e, in tal caso, i due integrali coincidono. In particolare, $f \in L_1(\nu)$ se e solo se $f \circ g \in L_1(\mu)$.

Proof. Verificare che ν è una misura e che $\nu(C) = \mu(\Omega)$, è immediato.
Ora, per ogni $E \in \Sigma$, abbiamo $\int_C \chi_E \, d\nu = \nu(E) = \mu(g^{-1}(E)) = \int_\Omega \chi_{g^{-1}(E)} \, d\mu = \int_\Omega (\chi_E \circ g) \, d\mu$. Di conseguenza, $\int_C s \, d\nu = \int_\Omega (s \circ g) \, d\mu$ per ogni funzione semplice Σ-misurabile $s \geq 0$ su C. Sia ora $f \colon C \to [0, +\infty]$ una funzione Σ-misurabile. Esiste una successione di funzioni semplici nonnegative $\{s_n\}$ tale che $s_n \nearrow f$ puntualmente su C. Siccome $s_n \circ g \nearrow f \circ g$ puntualmente su Ω, per il teorema della convergenza monotona,

$$\int_C f \, d\nu = \lim_n \int_C s_n \, d\nu = \lim_n \int_\Omega (s_n \circ g) \, d\mu = \int_\Omega (f \circ g) \, d\mu.$$

Ora, considerando le parti positiva e negativa, il resto della dimostrazione è immediato. □

2.2. **Disuguaglianza integrale di Jensen – versione bis.**

Jensen2 **Theorem 2.2** (Jensen II)**.** *Siano $C \subset \mathbb{R}^d$ un insieme convesso, Σ una σ-algebra su C contenente $\mathcal{B}(C)$, $f\colon C \to (-\infty,+\infty]$ una funzione convessa Σ-misurabile. Siano inoltre $(\Omega,\mathcal{A},\mu)$ uno spazio di misura finita non banale, $g\colon \Omega \to C$ una funzione ($\mathcal{A}$-Σ)–misurabile tale che $g \in L_1(\mu)$. Allora $\frac{1}{\mu(\Omega)}\int_\Omega g\,d\mu \in C$, l'integrale $\int_\Omega (f\circ g)\,d\mu$ esiste ed appartiene a $(-\infty,+\infty]$, e*

$$f\left(\frac{1}{\mu(\Omega)}\int_\Omega g\,d\mu\right) \le \frac{1}{\mu(\Omega)}\int_\Omega (f\circ g)\,d\mu\,.$$

Proof. La misura $\tilde{\mu} := \frac{1}{\mu(\Omega)}\,\mu$ è una misura di probabilità, alla quale possiamo applicare il push forward (Proposizione 2.1) per ottenere una misura di probabilità ν su C:

$$\nu(E) = \tfrac{1}{\mu(\Omega)}\,\mu(g^{-1}(E)) \qquad (E \in \Sigma).$$

Ora tutto segue dal Teorema 1.5: $\|\cdot\| \in L_1(\nu)$ (in quanto $\|g\| \in L_1(\mu)$), $\frac{1}{\mu(\Omega)}\int_\Omega g\,d\mu = x_\nu \in C$, $\frac{1}{\mu(\Omega)}\int_\Omega (f\circ g)\,d\mu = \int_C f\,d\nu$ esiste, e infine la disuguaglianza di Jensen $f\left(\int_C x\,d\nu(x)\right) \le \int_C f\,d\nu$ diventa la disuguaglianza finale. □

Dipartimento di Matematica, Università degli Studi, Via C. Saldini 50, 20133 Milano, Italy

E-mail address: libor.vesely@unimi.it