CORSO DI LAUREA IN MATEMATICA
Università di Pisa

Studio delle analogie fra somma e minimo di variabili aleatorie

Tesi di Laurea Triennale

CANDIDATO:
Giovanni Acerbi

RELATORE:
Dott. **Dario Trevisan**

Anno Accademico 2020/2021

Indice

Introduzione

Siano X e Y due variabili aleatorie indipendenti, con densità f e g, a valori in $\mathbb{R}^n$: allora è definita la somma $X+Y$, che ha densità data dalla convoluzione fra le due densità,

$$f * g(x) := \int_{\mathbb{R}^n} f(y)g(x-y)\ dy.$$

Sia l'operazione di somma che l'operazione di convoluzione hanno proprietà rilevanti, e ad esse sono associate importanti disuguaglianze, come la disuguaglianza di riarrangiamento o la disuguaglianza di Young con costanti ottimali. A noi interesserà vedere quali di queste proprietà e disuguaglianze varranno per un'altra operazione fra variabili aleatorie, l'operazione di minimo, che indicheremo con $\wedge$, e l'operazione che induce fra le densità, cioè

$$(f \triangle g)(x) := f(x) \int_x^\infty g(t)\ dt + g(x) \int_x^\infty f(t)\ dt.$$

In particolare, la disuguaglianza di riarrangiamento varrà senza alcuna modifica, mentre bisognerà modificare la disuguaglianza di Young per farla valere in questo caso, in quanto si trovano facilmente contresempi alla disuguaglianza originale. Inoltre, riusciremo a formulare un teorema limite centrale, e a mostrare che due quantità rilevanti nella teoria dell'informazione, l'entropia differenziale di Shannon e l'informazione di Fisher, si comportano bene per questo teorema limite centrale. Infine, faremo un paio di osservazioni collegate al trasporto ottimo di massa.

Capitolo 1

Somma e convoluzione

1.1 Prime proprietà

Nel seguito, X e Y saranno variabili aleatorie a valori in $\mathbb{R}^n$, f e g saranno funzioni da $\mathbb{R}^n$ a $\mathbb{R}$ misurabili, e μ la misura di Lebesgue su $\mathbb{R}^n$

La convoluzione, come operazione, è bilineare e associativa, e se f e g sono densità di probabilità, anche $f*g$ la è. Inoltre, una proprietà importante che servirà anche nel seguito riguarda la "convergenza delle riscalate":

Definizione 1.1 *Sia f una funzione da $\mathbb{R}^n$ a $\mathbb{R}$, e sia $\delta > 0$. La* **riscalata** *di f di parametro δ è la funzione $\sigma_\delta f : \mathbb{R}^n \to \mathbb{R}$ definita da*

$$\sigma_\delta f(x) := \frac{1}{\delta^n} f\Big(\frac{x}{\delta}\Big).$$

Nel caso in cui f sia la densità di una variabile aleatoria X, $\sigma_\delta f$ non è nient'altro che la densità di δX. Il seguente risultato è ben noto:

Proposizione 1.2 (Convergenza delle riscalate) *Sia f una funzione in $L^p(\mathbb{R}^n)$, e g una funzione in $L^1(\mathbb{R}^n)$. Se poniamo $m := \int_{\mathbb{R}^n} g(x)dx$, allora*

$$f * (\sigma_\delta g) \overset{L^p}{\to} m \cdot f$$

per $\delta \downarrow 0$.

In particolare, se X e Y hanno densità, allora $X + \delta Y$ converge a X in L^1 per $\delta \downarrow 0$.

L'ultima proprietà che andiamo a considerare in questo paragrafo è l'esistenza di un semigruppo associato sia alla somma che alla convoluzione. Per rendere la sua definizione più agevole, chiamiamo

$$g_t(x) := \frac{1}{\sqrt{2\pi t}} e^{\frac{-|x|^2}{2t}},$$

e abusando di notazione chiamiamo $\mathcal{N}(\mu, \sigma^2)$ una variabile aleatoria normale di media μ e varianza σ^2, ovvero una variabile con densità data dalla formula

$$g(x) = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{\frac{-|x-\mu|^2}{2\sigma^2}}.$$

Ricordiamo che, se $\mathcal{N}(\mu_1, \sigma_1^2)$ e $\mathcal{N}(\mu_2, \sigma_2^2)$ sono indipendenti, allora la loro somma ha legge $\mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.

Definizione 1.3 (Semigruppo del calore) *Il semigruppo di operatori $(S_t)_{t\in\mathbb{R}^+}$ definito da $S_t f := f * g_t$ è detto* **semigruppo del calore***, ed è effettivamente un semigruppo per quanto detto sopra, cioè vale $S_{t_1+t_2} = S_{t_1} \circ S_{t_2}$.*

Due proprietà importanti di questo semigruppo sono riassunte nel seguente lemma:

Lemma 1.4 *Per $t \downarrow 0$ vale $S_t f \to f$, ovvero il semigruppo del calore è un semigruppo continuo*
Vale inoltre $\partial_t S_t f = \frac{1}{2}\Delta S_t f$, dove Δ è il laplaciano.

L'esistenza di un semigruppo offre, fra molte altre cose, una strada aggiuntiva per alcune dimostrazioni: in particolare, vedremo una dimostrazione della disuguaglianza di Young con costanti ottimali che sfrutta la monotonia di una certa quantità lungo le traiettorie del semigruppo.

1.2 Disuguaglianza di riarrangiamento

Lo scopo di questa sezione è dimostrare la disuguaglianza di riarrangiamento in versione integrale. Procediamo con ordine, definendo il riarrangiamento di una funzione; ci restringeremo a una determinata classe di funzioni, quelle che vanno a zero all'infinito.

Definizione 1.5 *Data f su $\mathbb{R}^n$, diciamo che f* **va a zero all'infinito** *se $\mu\{|f| \geq a\} < \infty$ per ogni a reale positivo.*

Notiamo che basta essere in qualche L^p per $p \geq 1$, quindi è una condizione che non sarà necessario specificare nella maggior parte degli enunciati

Definizione 1.6 *Data f su $\mathbb{R}^n$ a valori non negativi che va a zero all'infinito, la sua* **riarrangiata** *è l'unica funzione (a meno di uguaglianze quasi ovunque) f^* tale che i suoi insiemi di livello siano delle palle centrate in zero, e abbiano la stessa misura degli insiemi di livello di f per lo stesso valore, cioè*

$$\mu\{f \geq a\} = \mu\{f^* \geq a\}.$$

Dato A insieme di R^n a misura finita, il suo **riarrangiato** *A^* è la palla centrata in zero avente la stessa misura di A.*

Segue dalla definizione che f^* è radiale e radialmente decrescente. Inoltre ha la stessa integrabilità di f. Andiamo ora ad enunciare la disuguaglianza di riarrangiamento per la convoluzione

Teorema 1.7 (Riesz) *Date f, g e h su $\mathbb{R}^n$ a valori non negativi che vanno a zero all'infinito, indichiamo*

$$I(f, g, h) := \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} h(x)f(y)g(x-y)\ dx\ dy = \int_{\mathbb{R}^n} h(x)(f * g)(x)\ dx.$$

Allora vale

$$I(f, g, h) \leq I(f^*, g^*, h^*).$$

Dimostreremo solo il caso in cui $n = 1$, perché il caso generale non ci sarà utile ai fini del confronto con l'operazione di minimo. È possibile trovare la dimostrazione completa su [1], da dove abbiamo preso anche la seguente dimostrazione. Premettiamo un lemma:

Lemma 1.8 (Layer cake representation) *Sia f una funzione misurabile non negativa su uno spazio di misura (Ω, μ). Allora*

$$f(x) = \int_0^\infty I_{\{f>t\}}(x)\ dt.$$

In particolare, possiamo scrivere $f = \int_0^\infty I_{\{f>t\}}\ dt$. Visto che se A è un insieme di $\mathbb{R}^n$ allora $\Big(I_A\Big)^* = I_{A^*}$, vale

$$f^* = \int I_{\{f>t\}^*}\ dt = \int I^*_{\{f>t\}}\ dt.$$

Procediamo a dimostrare la disuguaglianza di riarrangiamento in dimensione uno:

$\mathcal{D}im$: grazie al lemma, possiamo dimostrare la disuguaglianza nel caso in cui f, g e h siano delle indicatrici: infatti, supponendo che la disuguaglianza valga per le indicatrici, avremmo

$$I(f,g,h) = \int_{\mathbb{R}}\int_{\mathbb{R}} h(x)f(y)g(x-y)\ dx\ dy =$$

$$= \int_{\mathbb{R}}\int_{\mathbb{R}}\int_0^\infty\int_0^\infty\int_0^\infty I_{\{h>t\}}(x)I_{\{f>t'\}}(y)I_{\{g>t''\}}(x-y)\ dt''\ dt'\ dt\ dx\ dy =$$

$$= \int_0^\infty\int_0^\infty\int_0^\infty\int_{\mathbb{R}}\int_{\mathbb{R}} I_{\{h>t\}}(x)I_{\{f>t'\}}(y)I_{\{g>t''\}}(x-y)\ dt''\ dt'\ dt\ dx\ dy \leq$$

$$\leq \int_0^\infty\int_0^\infty\int_0^\infty\int_{\mathbb{R}}\int_{\mathbb{R}} I^*_{\{h>t\}}(x)I^*_{\{f>t'\}}(y)I^*_{\{g>t''\}}(x-y)\ dt''\ dt'\ dt\ dx\ dy =$$

$$= I(f^*, g^*, h^*).$$

Quindi, siano F, G e H tre insiemi. Indicheremo con le stesse lettere anche le corrispettive funzioni indicatrici. Per regolarità esterna della misura, esistono degli F_k aperti a misura finita che decrescono a F, e similmente dei G_k e degli H_k. Per convergenza dominata, $I(F_k, G_k, H_k)$ tende a $I(F, G, H)$, e similmente $I(F_k^*, G_k^*, H_k^*)$ tende a $I(F^*, G^*, H^*)$. Allora ci basta mostrare la disuguaglianza nel caso in cui F, G e H siano tre aperti a misura finita. Ma ogni aperto di $\mathbb{R}$ è un'unione disgiunta numerabile di intervalli (si vede facilmente ragionando sulle componenti connesse). Se $F = \bigsqcup_{k=1}^\infty I_k$, chiamiamo $F_m := \bigsqcup_{k=1}^m I_k$ e ciclici. Allora $F_m \uparrow F$, perciò per convergenza monotona $I(F_m, G_m, H_m)$ tende a $I(F, G, H)$, e similmente per le riarrangiate. Dunque ci basta mostrare la tesi per unioni disgiunte finite di intervalli per avere la tesi.

Scriviamo allora

$$F(x) = \sum_{j=1}^{l} f_j(x - a_j),$$

dove le f_j sono indicatrici di intervalli centrati, e gli a_j dei reali qualsiasi. Similmente, $G(x) = \sum_{j=1}^l g_j(x - b_j)$ e $H(x) = \sum_{j=1}^l h_j(x - c_j)$.
$I(F, G, H)$ è una somma di termini del tipo

$$I_{pqr} := \int_{\mathbb{R}}\int_{\mathbb{R}} h_p(x - c_p)f_q(y - a_q)g_r(x - y - b_r)\ dx\ dy.$$

Quello che vorremmo mostrare, è che avvicinando tutti gli intervalli per formarne uno unico andiamo a migliorare I. A tal proposito, consideriamo F_t, G_t e H_t dove ogni termine $f_j(x-a_j)$ è stato rimpiazzato da $f_j(x-ta_j)$, $0 \leq t \leq 1$, e similmente per le altre due. Allora

$$I_{pqr}(t) = \int_{\mathbb{R}} \int_{\mathbb{R}} h_p(x-tc_p) f_q(y-ta_q) g_r(x-y-tb_r)\ dx\ dy =$$

$$= \int_{\mathbb{R}} \int_{\mathbb{R}} h_p(x+t(a_q+b_r-c_p)) f_q(y) g_r(x-y)\ dx\ dy =$$

$$\int_{\mathbb{R}} h_p(x-(a_q+b_r-c_p)t) u_{qr}(x)\ dx\ dy,$$

dove $u_{qr} = f_q * g_r$ è una funzione già riarrangiata (quindi simmetrica e decrescente su $\mathbb{R}^+$). In particolare, per t che decresce il termine da integrare cresce puntualmente.
Questo ragionamento non conclude del tutto: infatti, dobbiamo fermare questo procedimento ogni qual volta due intervalli si "toccano" e unire i due intervalli in questione per formarne uno solo. Visto che il numero di intervalli è finito, dopo un numero finito di passi otteniamo un unico intervallo (centrato) che corrisponde a F^*, e grazie alla crescenza di ogni passo otteniamo la tesi. $\square$

1.3 Disuguaglianza di Young con costanti ottimali

Diamo subito l'enunciato, che non necessita di definizioni preliminari:

Teorema 1.9 (Disuguaglianza di Young) *Dati $p, q, r \geq 1$ tali che $\frac{1}{p} + \frac{1}{q} = 1 + \frac{1}{r}$, e date $f \in L^p(\mathbb{R}^N)$, $g \in L^q(\mathbb{R}^N)$, vale*

$$\|f * g\|_r \leq C_{p,q,r,N} \|f\|_p \|g\|_q,$$

con $C_{p,q,r,N}$ costante ottimale ottenuta con f e g opportune densità gaussiane. Esplicitamente,

$$C_{p,q,r,N} = \left(\frac{C_p C_q}{C_r} \right)^N,$$

dove $C_t = \sqrt{\frac{t^{1/t}}{(t')^{1/t'}}}$ con t' esponente coniugato di t.

La dimostrazione sarà lunga e sfrutterà molto l'esistenza del semigruppo del calore, ma otterremo un risultato più generale di quello appena enunciato. È possibile trovare questa dimostrazione in [2].
Dato n intero positivo, siano $p_1, \ldots, p_n \geq 1$ dei reali tali che

$$\sum_{i=1}^{n} \frac{1}{p_i} = n - 1.$$

Siano $f_1, \ldots, f_n$ delle funzioni non negative in $L^1(\mathbb{R}^N)$, e siano $u_j(x,t)$ le soluzioni dell'equazione del calore con coefficienti di diffusione κ_j

$$\frac{\partial u_j(x,t)}{\partial t} = \kappa_j \Delta u_j(x,t)$$

con dato inziale f_j (cioè $u_j(x,t)$ tende puntualmente a $f_j(x)$ per $t \downarrow 0$). Grazie al Lemma 1.4 vale

$$u_j = f_j * g_{2\kappa_j t}.$$

Cosa andremo a fare? Dimostreremo che esiste una scelta (unica) dei coefficienti di diffusione per cui la funzione

$$w(x,t) := u_1^{1/p_1} * \ldots * u_n^{1/p_n}(x,t)$$

è monotona crescente, con stazionarietà solo se i dati iniziali sono gaussiane opportune. Poi mostreremo che un'altra quantità si comporta monotonamente nel tempo, e da lì sarà evidente la disuguaglianza di Young per come l'abbiamo enunciata.

Lemma 1.10 *Con le condizioni imposte sopra, se $\kappa_j = (p_j p_j')^{-1}$ per ogni j, allora w è crescente nel tempo a partire da*

$$w(x,0) = f_1^{1/p_1} * \ldots * f_n^{1/p_n}(x).$$

In più, w è costante nel tempo se e solo se ogni f_j è multipla di una densità gaussiana di varianza $cN\kappa_j$, con c indipendente da j.

Non lo dimostreremo, in quanto è molto tecnico e rischia di sviare dall'argomento centrale della sezione

Una conseguenza di questo lemma, è che la quantità

$$\Psi(t) := \sup_x w(x,t) = \sup_x u_1^{1/p_1} * \ldots * u_n^{1/p_n}(x,t)$$

è crescente. Calcolando il limite di questa quantità e scegliendo bene n e le κ_j, avremo la disuguaglianza di Young con costanti ottimali, ma procediamo con ordine: mostriamo che Ψ, come funzionale, è *invariante per dilatazione*, ovvero se prendiamo una funzione $\lambda(t)$ positiva, allora la Ψ generata dalle u_j è la stessa della Ψ generata dalle $\sigma_{\lambda(t)}u_j$. Infatti, la $w'(x,t)$ del caso riscalato è uguale a

$$w'(\lambda(t)x,t) = \lambda(t)^{N(n-1)} \cdot \frac{1}{\lambda(t)^{N\sum 1/p_j}} w(x,t) = w(x,t),$$

dove il primo fattore esce cambiando variabile nell'integrale della formula della convoluzione, e l'ultima uguaglianza segue dalla condizione $\sum_j 1/p_j = n-1$, da cui l'invarianza per dilatazione di Ψ. Siamo ora pronti a enunciare e dimostrare il teorema chiave:

Teorema 1.11 *Sotto le ipotesi sui p_j e sui κ_j del teorema precedente, Ψ cresce dal valore*

$$\Psi(0) = \sup_x f_1^{1/p_1} * \ldots * f_n^{1/p_n}(x)$$

al valore limite

$$\lim_{t\to\infty} \Psi(t) = \prod_{j=1}^{n} C_{p_j}^N \|f_j\|_1^{1/p_j},$$

dove le costanti C_p sono quelle definite nell'enunciato della disuguaglianza di Young. In più, Ψ è costante nel tempo se e solo se f_j è multipla di una densità gaussiana con varianza $cN\kappa_j$, con c indipendente da j.

$\mathcal{D}im$: la crescenza di Ψ e la sua costanza solo nel caso specificato arrivano dal lemma precedente, dunque ci basta mostrare che il limite per $t \to \infty$ è quello indicato, e avremo la tesi.

Scaliamo con la funzione $\lambda(t) = \sqrt{\frac{1}{1+2t}}$, cosicché le riscalate siano

$$U_j(x,t) = \sqrt{1+2t}^N u(x\sqrt{1+2t}, t).$$

Usando la formula esplicita della soluzione u_j ottenuta nel Lemma 1.4, cioè $u_j = f_j * g_{2\kappa_j t}$, otteniamo

$$U_j = \left(\sigma_{\sqrt{\frac{1}{1+2t}}} f\right) * g_{\kappa\frac{2t}{1+2t}} \to m(f) g_\kappa,$$

dove il limite è quantomeno in L^1. Passando al limite, otteniamo dunque

$$\lim_{t\to\infty}\Psi(t) = \prod_{j=1}^{n} \|f_j\|_1^{1/p_j} \sup_x g_{\kappa_1}^{1/p_1} * \ldots * g_{\kappa_n}^{1/p_n}(x).$$

Visto che

$$g_\kappa^{1/p} = C_p^N \cdot (2\pi)^{(2p'/N)^{-1}} g_{1/p'},$$

sfruttando il fatto che $\sum_j 1/p'_j = 1$ otteniamo infine

$$\sup_x g_{\kappa_1}^{1/p_1} * \ldots * g_{\kappa_n}^{1/p_n}(x) = (2\pi)^{-N/2} \prod_{j=1}^{n} C_{p_j}^N,$$

da cui la tesi. □

Per ottenere la disuguaglianza di Young, consideriamo il caso $n = 3$: chiamando le tre funzioni per semplicità f, g e h, la disuguaglianza che abbiamo ottenuto ci dice che

$$\int_{\mathbb{R}^N} h(x)(f*g)(x)\ dx\ dy \le (C_p C_q C_{r'})^N \|f\|_p \|g\|_q \|h\|_{r'},$$

dove $1/p + 1/q + 1/r' = 2$. Per $h = (f*g)^{r-1}$ otteniamo esattamente

$$\|f*g\|_r^r \le (C_p C_q C_{r'})^N \|f\|_p \|g\|_q \|f*g\|_r^{r-1},$$

cioè la disuguaglianza di Young con esponenti ottimali notando che $C_{r'} = (C_r)^{-1}$. Si può generalizzare a un numero arbitrario di funzioni, e la dimostrazione è sostanzialmente la stessa.

Capitolo 2

L'operazione di minimo fra v.a.

In questo capitolo, dopo aver delineato le proprietà immediate del minimo e dell'operatore $\triangle$ che stiamo per definire, dimostreremo la disuguaglianza di riarrangiamento e vedremo che una variante della disuguaglianza di Young vale per quest'operazione. Nel prossimo capitolo passeremo a un teorema limite centrale, e vedremo un paio di quantità che si comportano bene per tale limite. Infine, vedremo un pochino di cose collegate al trasporto ottimo.

2.1 Prime proprietà

Nel seguito, X e Y saranno v.a. indipendenti a valori non negativi con densità f e g rispettivamente. Indichiamo con F, G rispettivamente le loro funzioni di ripartizione $F(x) = P\{X \leq x\}$. Ricordiamo che vale $f = F'$ quasi certamente.

Per prima cosa calcoliamo la funzione di ripartizione di $X \wedge Y$: visto che $\{X \wedge Y \geq x\} = \{X \geq x\} \cap \{Y \geq x\}$, vale

$$P\{X \wedge Y \leq x\} = 1 - P\{X \wedge Y \geq x\} = 1 - P\{X \geq x\} \cdot P\{Y \geq x\} =$$

$$= 1 - (1 - F(x))(1 - G(x)) = F(x) + G(x) - F(x)G(x).$$

In particolare, la sua derivata (che è la densità di $X \wedge Y$) vale proprio

$$f(x) + g(x) - f(x)G(x) - g(x)F(x) = f(x)(1 - G(x)) + g(x)(1 - F(x)) =$$

$$= f(x) \int_x^\infty g(t)\ dt + g(x) \int_x^\infty f(t)\ dt =: (f \triangle g)(x).$$

Questo operatore è associativo e bilineare in f e g, ma solo nella formulazione integrale. Inoltre, se f e g sono decrescenti, anche $f \triangle g$ la è.

Come per la somma, abbiamo un risultato di convergenza delle riscalate, motivato dal fatto che $X \wedge (kY)$, per k grande, è all'incirca X.

Proposizione 2.1 (Convergenza delle riscalate) *Siano f e g due funzioni in $L^1(\mathbb{R}^+)$. Se poniamo $m := \int_{\mathbb{R}^+} g(x)dx$, allora*

$$f\triangle(\sigma_k g) \xrightarrow{L^1} m \cdot f$$

per $k \to \infty$.

$\mathcal{D}im$: facciamo il conto:

$$\|f\triangle(\sigma_k g) - m \cdot f\|_1 =$$

$$= \left| \int_{\mathbb{R}^+} \left(f(x) \int_{x/k}^{\infty} g(t)\ dt - m \cdot f(x) + \frac{1}{k} g(x/k) \int_x^{\infty} f(t)\ dt \right)\ dx \right| \leq$$

$$\leq \int_{\mathbb{R}^+} \left| f(x) \int_{x/k}^{\infty} g(t)\ dt - m \cdot f(x) \right|\ dx + \int_{\mathbb{R}^+} \left| g(x) \int_{kx}^{\infty} f(t)\ dt \right|\ dx$$

Il primo termine tende a 0 per convergenza dominata, e anche il secondo sempre per lo stesso motivo. $\square$

Inoltre, come per la somma e la convoluzione, anche queste operazioni sono associate a un semigruppo. Per semplificare la notazione, chiamiamo

$$\gamma_\lambda(x) := \lambda e^{-\lambda x},$$

che è la densità di una v.a. esponenziale di parametro λ. Una v.a. esponenziale di parametro λ verrà invece indicata con $Exp(\lambda)$

Lemma 2.2 *La famiglia di operatori $(T_t)_{t\in\mathbb{R}^+}$ definito da $T_t f := f\triangle\gamma_t$ è un semigruppo.*

$\mathcal{D}im$: per associatività di $\triangle$, ci basta mostrare che $\gamma_{t+s} = \gamma_t\triangle\gamma_s$. Ma questo è vero: notando che $\int_x^\infty \gamma_\lambda = 1 - e^{-\lambda x}$, abbiamo

$$\gamma_t\triangle\gamma_s(x) = te^{-tx}(e^{-sx}) + se^{-sx}(e^{-tx}) = (t+s)e^{-(t+s)x},$$

da cui la tesi. $\square$

Anche in questo caso si tratta di un semigruppo continuo (segue dalla convergenza delle riscalate) che risolve un'equazione differenziale, cioè

$$\partial_t u(t,x) = -xu(t,x) + \int_x^\infty u(t,y)\ d\mu(y) = \partial_x \left(x \int_x^\infty u(t,y)\ d\mu(y) \right).$$

Questo conclude i preliminari su quest'operazione.

2.2 Disuguaglianza di riarrangiamento

Consideriamo prima il caso discreto. Il caso continuo seguirà discretizzando e passando al limite, ma procediamo in ordine, ridefinendo le cose nel caso discreto. Di entrambi questi risultati non è stato trovato alcun cenno in letteratura, e sono pertanto da considerarsi nuovi.

Definizione 2.3 *Date $f, g \in \ell^1$ a valori non negativi, poniamo*

$$f\triangle g(n) := f(n)\sum_{k\geq n} g(k) + g(n)\sum_{k\geq n} f(k) - f(n)g(n)$$

È immediato verificare che se f e g sono decrescenti, allora anche $f\triangle g$ ha questa proprietà. Questo sarà centrale nella dimostrazione della disuguaglianza di riarrangiamento.

Definizione 2.4 *Data $f \in \ell^1$, chiamiamo f^* l'unico elemento decrescente di $\ell^1(\mathbb{N})$ con stessa misura degli insiemi di livello di f. Una sua espressione è*

$$\int_0^\infty I_{\{f>t\}^*}\, dt,$$

dove con A^* si intende l'insieme $\{0, \ldots, \#A - 1\}$.

Teorema 2.5 (Disuguaglianza di riarrangiamento discreta) *Se f, g e h sono tre elementi di ℓ^1 a valori non negativi, allora*

$$\sum_n h(n) \cdot f\triangle g(n) \leq \sum_n h^*(n) \cdot f^*\triangle g^*(n).$$

Le somme in questione hanno sempre senso in $[0, \infty]$.

$\mathcal{D}im$: Usando il Lemma 1.8 possiamo ricondurci al caso in cui f, g e h sono delle indicatrici. Infatti, riguardando la dimostrazione di questo passaggio nel caso della convoluzione, ci accorgiamo che l'unica cosa che serve è che l'operatore sia bilineare. Visto che siamo nel discreto, possiamo verificarlo solo per indicatrici finite.

Siano dunque $f = I_A$, $g = I_B$ e $h = I_C$. Possiamo decomporre in modo unico $A = A_0 \cup A_1$, e similmente gli altri, di modo tale che A_0 sia della forma $\{0, 1, \ldots, a\}$ e A_1 sia "staccato "da A_0, cioè $\inf A_1 \geq a + 2$. A_0 può essere eventualmente vuoto, ma in tal caso $0 \notin A_1$. Poniamo $A' := A_0 \cup (A_1 - 1)$. Quello che noi vorremmo mostrare è che
$\sum_n I_{C'}(n) \cdot I_{A'}\triangle I_{B'}(n) \geq \sum_n I_C(n) \cdot I_A\triangle I_B(n)$. Se ci riusciamo, abbiamo finito perché in un numero finito di applicazioni di questo arriviamo a I_{A*} e corrispettivi, dunque alla tesi. Prima di procedere col conto, osserviamo

che $I_A(n) = I_{A_0}(n) + I_{A_1}(n)$ e $I_{A'}(n) = I_{A_0}(n) + I_{A_1}(n+1)$. Inoltre, se chiamiamo $\tau f(n) := f(n+1)$, abbiamo che

$$(\tau f)\triangle g(n) = f(n+1)\sum_{k\geq n} g(k) + g(n)\sum_{k\geq n+1} f(k) - f(n+1)g(n).$$

In particolare,

$$(\tau f)\triangle g(n) - f\triangle g(n) = f(n+1)\sum_{k\geq n+1} -f(n)\sum_{k\geq n} g(k).$$

Siamo pronti: per linearità,

$$\begin{aligned}
&\sum_n I_{C'}(n) I_{A'}\triangle I_{B'}(n) = \\
&\qquad = \sum_n I_{C_0}(n) I_{A_0}\triangle I_{B_0}(n) + \sum_n I_{C_1-1}(n) I_{A_0}\triangle I_{B_0}(n) \\
&\qquad + \sum_n I_{C_0}(n) I_{A_1-1}\triangle I_{B_0}(n) + \sum_n I_{C_1-1}(n) I_{A_1-1}\triangle I_{B_0}(n) \\
&\qquad + \sum_n I_{C_0}(n) I_{A_0}\triangle I_{B_1-1}(n) + \sum_n I_{C_1-1}(n) I_{A_0}\triangle I_{B_1-1}(n) \\
&\qquad + \sum_n I_{C_0}(n) I_{A_1-1}\triangle I_{B_1-1}(n) + \sum_n I_{C_1-1}(n) I_{A_1-1}\triangle I_{B_1-1}(n).
\end{aligned}$$

Consideriamo i termini uno a uno, tenendo presente che vogliamo mostrare che sono più grandi della loro controparte nella somma senza primi. Il primo è uguale in entrambe.
Per il secondo, visto che I_{A_0} e I_{B_0} sono decrescenti, anche $I_{A_0}\triangle I_{B_0}$ è decrescente, pertanto

$$\sum_n I_{C_1-1}(n) I_{A_0}\triangle I_{B_0}(n) \geq \sum_n I_{C_1}(n+1) I_{A_0}\triangle I_{B_0}(n+1) =$$

$$= \sum_n I_{C_1}(n) I_{A_0}\triangle I_{B_0}(n),$$

dove l'ultimo passaggio deriva dall'aggiunta di un termine nullo (infatti $I_{C_1}(0) = 0$). Quest'aggiunta di termini nulli verrà fatta di frequente, quindi dove non necessario non la spiegherò.
Per il terzo (e analogamente per il quinto), ci basta mostrare che

$$\sum_n I_{C_0}(n)(I_{A_1-1} - I_{A_1})\triangle I_{B_0}(n) \geq 0$$

Per quanto detto prima di iniziare il conto, vale però

$$\sum_n I_{C_0}(n)(I_{A_1-1} - I_{A_1})\triangle I_{B_0}(n) =$$

$$\sum_n I_{C_0}(n)I_{A_1}(n+1)\sum_{k\geq n+1} I_{B_0}(k) - \sum_n I_{C_0}(n)I_{A_1}(n)\sum_{k\geq n} I_{B_0}(k) \geq$$

$$\sum_n I_{C_0}(n+1)I_{A_1}(n+1)\sum_{k\geq n+1} I_{B_0}(k) - \sum_n I_{C_0}(n)I_{A_1}(n)\sum_{k\geq n} I_{B_0}(k) = 0$$

sempre per decrescenza di I_{C_0}.
Per il quarto (e analogamente per il sesto), si procede in modo simile:

$$\sum_n I_{C_1}(n+1)I_{A_1-1}\triangle I_{B_0}(n) - \sum_n I_{C_1}(n)I_{A_1}\triangle I_{B_0}(n) =$$

$$\sum_n I_{C_1}(n+1)I_{A_1}(n+1)\sum_{k\geq n+1} I_{B_0}(k) - \sum_n I_{C_1}(n)I_{A_1}(n)\sum_{k\geq n} I_{B_0}(k) \geq$$

$$\sum_n I_{C_1}(n)I_{A_1}(n)\sum_{k\geq n} I_{B_0}(k) - \sum_n I_{C_1}(n)I_{A_1}(n)\sum_{k\geq n} I_{B_0}(k) = 0.$$

Per il settimo e l'ottavo si procede allo stesso modo. La tesi è dunque dimostrata. □

Notiamo che l'idea dimostrativa è molto simile a quella del caso della convoluzione: unire gli insiemi fino a ricongiungerli in uno solo.

La dimostrazione è facilmente generalizzabile al caso di più funzioni, cioè si riesce a dimostrare che date $f_1, \ldots, f_m, h$, vale

$$\sum_n h(n) \cdot f_1 \triangle f_2 \triangle \ldots \triangle f_m(n) \leq \sum_n h^*(n) \cdot f_1^* \triangle f_2^* \triangle \ldots \triangle f_m^*(n).$$

Infatti, dopo aver definito

$$f_1 \triangle f_2 \triangle \ldots \triangle f_m(n) := \sum_{i=1}^m f_i(n) \prod_{j\neq i} \Big(\sum_{k=n}^\infty f_j(k) \Big),$$

la restrizione alle indicatrici è ancora valida (chiamiamo gli insiemi C e A^i). Inoltre, osservando che l'operazione $\triangle$ è associativa e commutativa, data una qualunque funzione σ da $\{0, \ldots, m\}$ in $\{0, 1\}$ vale:

$$\sum_n I_{C'_{\sigma(0)}}(n)\big(\triangle_{i=1}^m I_{A^{i\prime}_{\sigma(i)}}\big)(n) =$$

$$= \sum_n I_{C'_{\sigma(0)}}(n)\Big(\big(\triangle_{\sigma(i)=1} I_{A_1^i - 1}\big)\triangle\big(\triangle_{\sigma(i)=0} I_{A_0^i}\big)\Big)(n) \geq$$

$$\sum_n I_{C'_{\sigma(0)}}(n)\Big(\big(\triangle_{\sigma(i)=1} I_{A_1^i}\big)\triangle\big(\triangle_{\sigma(i)=0} I_{A_0^i}\big)\Big)(n) =$$

$$= \sum_n I_{C_{\sigma(0)}}(n)\big(\triangle_{i=1}^m I_{A^i_{\sigma(i)}}\big)(n).$$

Infatti, l'unica proprietà che abbiamo usato è che le indicatrici con pedice 0 sono decrescenti (cosa che succede anche quando ne triangoliamo un paio) e che le indicatrici con pedice uno hanno valore nullo in 0.

2.3 Disuguaglianza di riarrangiamento, caso continuo

Stavolta, tutte le funzioni stanno in $L^1(\mathbb{R}^+)$. Le definizioni sono simili, cioè f^* è definita nello stesso modo di prima, e $f\triangle g$ è definita da

$$f\triangle g(x) = f(x)\int_x^\infty g(s)ds + g(x)\int_x^\infty f(s)ds.$$

Teorema 2.6 (Disuguaglianza di riarrangiamento continua) *Se $f, g, h \in L^1(\mathbb{R}^+)$ sono a valori non negativi, allora*

$$\int h(x)\cdot f\triangle g(x) \leq \int h^*(x)\cdot f^*\triangle g^*(x).$$

$\mathcal{D}im$: Osserviamo che, similmente al caso discreto, possiamo ridurci a dimostrare la tesi per f, g e h indicatrici. Il conto è lo stesso sostituendo la somma con l'integrale.
Inoltre, dati A, A', B, B', C, C', vale

$$\int (I_C(x) - I'_C(x))\cdot(I_A - I'_A)\triangle(I_B - I'_B)(x) \leq |C - C'|\cdot(|A - A'| + |B + B'|),$$

questo perché una disuguaglianza grezza sull'operatore $\triangle$ è $\|f\triangle g\|_\infty \leq \|f\|_\infty\|g\|_1 + \|f\|_1\|g\|_\infty$. Similmente, anche con gli insiemi riarrangiati c'è una disuguaglianza analoga. Da questo segue che possiamo mostrare la disuguaglianza di riarrangiamento continua solo per una determinata classe di indicatrici, che approssimi tutte le indicatrici nel senso della disuguaglianza di sopra. Questa classe è quella delle indicatrici degli insiemi chiusi e limitati tali che esiste un divisore razionale comune di tutti gli estremi. Ad esempio, $[\frac{1}{2}, 1]\cup[\frac{7}{4}, \frac{9}{4}]$ va bene, mentre $\bigcup_{n\in\mathbb{N}}[n, n + \frac{1}{2^n}]$ no.
In questa situazione, possiamo applicare il caso discreto. Infatti, se $f = I_A$, $g = I_B$ e $h = I_C$, allora se k è il massimo comun divisore relativo agli estremi di questi, pongo

$$f'(n) := \int_{\frac{n}{k}}^{\frac{n+1}{k}} f(x)dx,$$

e analoghe sono le definizioni di g' e h'. Tutte e 3 risultano essere delle indicatrici, a meno di una costante moltiplicativa, e si ha sia

$$f'\triangle g'(n) = \int_{\frac{n}{k}}^{\frac{n+1}{k}} f\triangle g(x)dx,$$

sia di conseguenza

$$\sum_n h'(n)\cdot f'\triangle g'(n) = \int h\cdot f\triangle g.$$

Il discorso è analogo per $(f')^*$ e per gli altri, notando tra l'altro che $(f')^* = (f^*)'$. Allora applicando la disuguaglianza di riarrangiamento nel caso discreto, otteniamo la tesi, cioè la disuguaglianza di riarrangiamento nel continuo. $\square$

Anche questo si può generalizzare a un qualsiasi numero finito di variabili.

2.4 Disuguaglianza di Young, introduzione

Come vedremo fra poco, la disuguaglianza di Young presa com'è non risulterà vera. Il restante tempo sarà dedicato quindi a cercare formulazioni diverse, e a dimostrarle o trovare controesempi.

Proposizione 2.7 *Dati $p, r > 1$, non esiste nessuna costante $c(\lambda, p, r)$ tale che per ogni $f \in L^1 \cap L^r \cap L^p$ valga*

$$\|f \triangle \gamma_\lambda\|_r \leq c(\lambda, p, r)\|f\|_p.$$

In particolare, non esiste nessuna costante c indipendente da f e g tale che

$$\|f \triangle g\|_r \leq c\|f\|_p\|g\|_q.$$

$\mathcal{D}im$: poniamo $f_n := \frac{1}{n} I_{[a,a+n]}$, con a un numero reale abbastanza grande. Allora $f_n \triangle \gamma_\lambda$ è quasi uguale a γ_λ (è uguale fino ad a, quantomeno), perciò la sua norma r è più grande di una costante positiva indipendente da n. Ciononostante, per $n \to \infty$, la norma p di f_n tende a 0, da cui la tesi. Per ottenere il secondo enunciato, basta sostituire $g = \gamma_\lambda$. $\square$

Anche nel caso $r > 1, p = 1$ si ha lo stesso risultato: basta infatti porre $f_n := n I_{[0, \frac{1}{n}]}$ per avere un controesempio, mentre il caso $p = r = 1$ è banalmente vero con costante ottimale $c = 1$. Questo controesempio funziona anche per una classe di disuguaglianze più ampia.

Proposizione 2.8 *Dati $p, r > 1$, non esiste nessuna costante $c(\lambda, p, r)$ e nessun esponente $\alpha(\lambda, p, r) \in (0, 1)$ tali che per ogni $f \in L^1 \cap L^r \cap L^p$ valga*

$$\|f \triangle \gamma_\lambda\|_r \leq c\|f\|_1^\alpha \cdot \|f\|_p^{1-\alpha}.$$

Vedremo fra due sezioni che è vera una disuguaglianza simile nel discreto.

2.5 Disuguaglianza di Young: caso continuo con misure generiche

Un'altra possibilità (errata) nel caso continuo è che valga la disuguaglianza di Young con misure esponenziali, o in generale con misure non concentrate su un compatto (e non nulle).

Sia dunque ν una misura su $\mathbb{R}^+$, che non sia concentrata su un insieme compatto.

Proposizione 2.9 *Dati $p, r > 1$, non esiste nessuna costante $c(\lambda, p, r)$ tale che per ogni $f \in L^1(\nu) \cap L^p(\nu) \cap L^r(\nu)$ valga*

$$\left(\int (f \triangle \gamma_\lambda(x))^r d\nu(x) \right)^{\frac{1}{r}} \leq c \cdot \left(\int (f(x))^p d\nu(x) \right)^{\frac{1}{p}}$$

$\mathcal{D}im$: i controesempi sfruttano la stessa idea precedente, sono solo più complicati da scrivere. Il fatto è sempre che, potendo concentrare f molto distante da 0, se valesse la disuguaglianza davvero, si avrebbe che imponendo $\int f(x)d\nu(x) = 1$ si limiterebbe dal basso $\int f(x)^p d\nu$.

Per quanto riguarda il controesempio, sia $a \in \mathbb{R}^+$ tale che $\nu([0,a]) > 0$ e $\nu([a,\infty]) > 0$. Sia f_n costantemente $1/n$ da a a un certo b_n e nulla fuori, con il b_n scelto in modo che f_n sia una densità di probabilità. Allora la norma $L^p(\nu)$ di f_n va come $1/n^{p-1}$, che tende a 0 per $n \to \infty$, mentre la norma r del primo pezzo è più di

$$\left(\int_0^a \gamma_\lambda^r(x) d\nu(x) \right)^{1/r},$$

che non dipende da f, da cui la tesi. $\square$

La condizione che la misura non sia concentrata su un compatto può essere ulteriormente rilassata, chiedendo che esista un a tale che $\nu[0,a[$ e $\nu]a,\infty[$ siano entrambi strettamente positivi (anche infiniti eventualmente)

2.6 Disuguaglianza di Young: stima con somma. Le esponenziali sono ottimali?

Vista l'espressione esplicita di $\triangle$, si può pensare di stimare non con il prodotto delle norme, ma con una somma di prodotti di norme. Questa scelta funziona, quindi saremo anche interessati a vedere se le esponenziali potranno o meno ottimizzare la disuguaglianza.

Proposizione 2.10 *Per $1 \leq r \leq p$, esiste una costante $c(r,p)$ tale che date f e g in $L^1 \cap L^p \cap L^r$ valga*

$$\|f \triangle g\|_r \leq c\left(\|f\|_p \|g\|_1 + \|f\|_1 \|g\|_p\right).$$

In particolare, esistono due costanti c_1, c_2 tali che

$$\|f \triangle \gamma_\lambda\|_r \leq c_1 \|f\|_p + c_2 \|f\|_1$$

$\mathcal{D}im$: per $p = r$, dalla formula che definisce $f \triangle g$ si ricava facilmente

$$\|f \triangle g\|_p \leq \|f\|_p \|g\|_1 + \|f\|_1 \|g\|_p,$$

La tesi per $1 \leq r \leq p$ si ricava per interpolazione. $\square$

Le costanti c_1, c_2 così ottenute risultano essere

$$c_1 = \lambda \|e^{-\lambda x}\|_r = \lambda \left(\frac{1}{\lambda r} \right)^{\frac{1}{r}} \qquad c_2 = \|e^{-\lambda x}\|_{\frac{pr}{p-r}} = \left(\frac{p-r}{\lambda pr} \right)^{\frac{p-r}{pr}}$$

Passiamo alla possibile ottimalità. Consideriamo $f = \gamma_\mu$: allora $f\triangle\gamma_\lambda = \gamma_{\lambda+\mu}$. Per prima cosa facciamo il caso $p = r$. La disuguaglianza è allora

$$\frac{(\lambda+\mu)^{\frac{1}{p'}}}{p^{\frac{1}{p}}} \leq c_1 + c_2 \frac{\mu^{\frac{1}{p'}}}{p^{\frac{1}{p}}}.$$

Vedendo questo, l'unica cosa che ci sembra sensato dimostrare è che le esponenziali massimizzino la disuguaglianza a norma p fissata. Un buon punto di partenza è mostrare che risolvono l'equazione di Eulero-Lagrange del problema

$$\int_0^\infty e^{-\lambda px}(f + \lambda(1-F))^p$$

a norma 1 e p di f fissate, dove F è l'integrale fra 0 e x di f. Se vario f di εh, la condizione sulla norma 1 equivale a porre h a media nulla, mentre la condizione sulla norma p fissata ha come derivata direzionale

$$\frac{d}{d\varepsilon}\int (f+\varepsilon h)^p = p\int hf^{p-1}.$$

La derivata direzionale del funzionale stesso, invece, è

$$\frac{d}{d\varepsilon}\int e^{-\lambda px}(f+\varepsilon h+\lambda(1-F-\varepsilon H))^p = p\int e^{-\lambda px}(f+\lambda(1-F))^{p-1}(h-\lambda H).$$

Allora sostituendo $f = \gamma_\mu$, la derivata del funzionale diventa (a meno di costanti moltiplicative)

$$\int e^{-\lambda px}e^{-\mu(p-1)x}(h-\lambda H) = \int e^{-(\lambda+\mu)(p-1)x}\frac{d}{dx}(He^{-\lambda x}) =$$

$$= \int e^{-(\lambda+\mu)(p-1)x}He^{-\lambda x}.$$

Integriamo per parti anche il vincolo, ottenendo $\int e^{-\mu(p-1)x}H$. Dopo qualche passaggio, ci accorgiamo che per i moltiplicatori di Lagrange nessuna esponenziale γ_μ è punto critico.

Guardiamo pertanto il caso $r < p$: riscriviamo la derivata direzionale di $\|f\triangle\gamma_\lambda\|_r^r$ per chiarezza:

$$\int e^{-\lambda rx}(f+\lambda(1-F))^{r-1}(h-\lambda H) = \int \left(e^{-\lambda(r-1)x}(f+\lambda(1-F))^{r-1}\right)\frac{d}{dx}\left(e^{-\lambda x}H\right),$$

che per $f = \gamma_\mu$ diventa, dopo aver integrato per parti,

$$\int e^{-(\lambda+\mu)(r-1)x}e^{-\lambda x}H.$$

Applicando i moltiplicatori di Lagrange, se f fosse punto critico dovrebbe valere

$$\int e^{-(\lambda+\mu)(r-1)x} e^{-\lambda x} H + c e^{-\mu(p-1)x} H = 0$$

$$e^{-(\lambda r+\mu(r-1))x} = c e^{-\mu(p-1)x}$$

$$\lambda r + \mu(r-1) = \mu(p-1)$$

$$\mu = \mu(\lambda, p, r) = \frac{\lambda r}{p-r}.$$

Facciamo qualche osservazione: abbiamo ottenuto che, fissati λ, p ed r esiste un'unico punto critico esponenziale a norma p fissata. Il parametro del punto critico tende a ∞ per $r \to p$, che è in linea col fatto che per $p = r$ non ci sono punti critici.
In corrispondenza di tale μ, la norma r di $f \triangle \gamma_\lambda$ è

$$\|f \triangle \gamma_\lambda\|_r = \frac{\left(\frac{\lambda p}{p-r}\right)^{\frac{1}{r'}}}{r^{\frac{1}{r}}},$$

mentre la norma p di f è

$$\|f\|_p = \frac{\left(\frac{\lambda r}{p-r}\right)^{\frac{1}{p'}}}{p^{\frac{1}{p}}}.$$

È aperto il problema di determinare se questo punto critico sia davvero un massimo o meno.

2.7 Disuguaglianza di Young, caso discreto

Ora lavoriamo in $\mathbb{N}$. Per sostituire le esponenziali, scegliamo le geometriche: poniamo dunque $\gamma_\lambda(n) := \lambda(1-\lambda)^n$, per $\lambda \in (0,1)$. Ricordiamo che nel discreto vale $\|f\|_p \leq \|f\|_1$, con uguaglianza solo se f è concentrata in un unico punto Allora, per quanto detto sopra, si riesce ad ottenere una disuguaglianza pura

$$\|f \triangle \gamma_\lambda\|_p \leq c(\lambda, p)\|f\|_1.$$

Vale anche di più, e riassumiamo il tutto nel seguente enunciato:

Proposizione 2.11 *Date $f, g \in L^1(\mathbb{N})$ non negative, vale*

$$\|f \triangle g\|_p \leq \|f\|_1 \cdot \|g\|_1.$$

$\mathcal{D}im$: $\|f \triangle g\|_1 = \|f\|_1 \cdot \|g\|_1$, da cui

$$\|f \triangle g\|_p \leq \|f \triangle g\|_1 = \|f\|_1 \cdot \|g\|_1. \quad \square$$

Questa costante è ovviamente ottimale, e anzi è ottimale anche fissando g (infatti, basta scegliere $f := I_{\{0\}}$). In particolare, il caso di uguaglianza è ottenuto con una delle due geometrica, seppure si tratti di un caso limite.

Capitolo 3

TLC per l'operazione di minimo

Lo scopo di questa sezione, in analogia con il caso gaussiano, è dimostrare che il limite del minimo di n variabili i.i.d. è vicino ad una esponenziale, in un certo senso. Poi introdurremo due quantità, importanti nella teoria dell'informazione, che passano al limite nel senso del TLC.

Iniziamo ricordando l'enunciato del teorema limite centrale per la somma:

Teorema 3.1 (TLC) *Date X_i v.a. i.i.d. centrate, con varianza σ^2 finita, allora la successione*

$$S_n := \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i$$

converge in legge a una v.a. gaussiana $\mathcal{N}(0, \sigma^2)$.

Un risultato analogo vale per l'operazione di minimo:

Teorema 3.2 (TLC minimo) *Date X_i a valori non negativi, i.i.d con densità f continua in 0 e non nulla in 0, allora*

$$M_n := n(X_1 \wedge X_2 \wedge \ldots \wedge X_n)$$

converge in legge a una v.a. esponenziale $Exp(f(0))$.

$\mathcal{D}im$: la densità di $X_1 \wedge X_2 \wedge \ldots \wedge X_n$ è

$$nf(x)\big(1 - F(x)\big)^{n-1},$$

pertanto la densità di $n(X_1 \wedge X_2 \wedge \ldots \wedge X_n)$ è

$$f\left(\frac{x}{n}\right)\left(1 - F\left(\frac{x}{n}\right)\right)^{n-1}.$$

Per $n \to \infty$, $\frac{x}{n}$ va a 0, perciò

$$f\left(\frac{x}{n}\right) \to f(0),$$

e sviluppando al prim'ordine F, si ha

$$\left(1 - F\left(\frac{x}{n}\right)\right)^{n-1} \approx \left(1 - f(0)\frac{x}{n}\right)^{n} \to e^{-f(0)x},$$

da cui la tesi (c'è convergenza puntuale, uniforme sui compatti, e questo basta ad assicurare la convergenza debole, cioè la convergenza in legge). □

È ancora da stabilire se le ipotesi sulla densità siano ottimali o meno, ma per i nostri scopi sono sufficienti.

3.1 Massimizzazione dell'entropia

Andiamo a definire una quantità che passa al limite per entrambi i nostri TLC. Questa quantità è stata introdotta per la prima volta da Shannon in [3], l'articolo che ha sostanzialmente fatto nascere la teoria dell'informazione

Definizione 3.3 (Entropia differenziale di Shannon) *Data una v.a. X con densità f, la sua* **entropia differenziale di Shannon** *è definita da*

$$H(X) := -\int f(x)\log f(x)dx = H(f).$$

Le densità su $\mathbb{R}$ che massimizzano l'entropia a media e varianza fissate sono le gaussiane. In analogia con questo, possiamo mostrare che le densità su $\mathbb{R}^+$ che la massimizzano a media fissata sono proprio le esponenziali.

Teorema 3.4 *Fra le densità di probabilità con media fissata $\frac{1}{\lambda}$, quella che massimizza l'entropia è la densità esponenziale γ_λ, in corrispondenza della quale si ha $H(\gamma_\lambda) = 1 - \log\lambda$.*

$\mathcal{D}im$: prese f, g densità di probabilità qualsiasi, vale

$$\int f(x)\log\frac{f(x)}{g(x)}dx = -\int f(x)\log\frac{g(x)}{f(x)}dx \geq -\log\left(\int g(x)dx\right) = 0$$

per la disuguaglianza di Jensen (questa quantità viene detta **distanza di Kullback-Leibler** fra f e g). Allora

$$0 \leq \int f(x)\log\frac{f(x)}{\gamma_\lambda(x)}dx = -H(f) - \int f(x)\log\lambda dx + \int \lambda x f(x)dx =$$
$$= -H(f) - \log\lambda + 1 = H(\gamma_\lambda) - H(f),$$

da cui la tesi. □

Questa dimostrazione funziona allo stesso modo per mostrare quanto detto in precedenza sulle gaussiane, cioè che sono le massimizzatrici a varianza fissata.

3.2 Convergenza dell'entropia per TLC

In analogia con quanto succede con la somma, vogliamo mostrare che l'entropia passa al limite per TLC.

Teorema 3.5 *Se ho una successione di v.a. i.i.d. X_i centrate a entropia finita, allora*
$$H(S_n) \to H(\mathcal{N}(0, Var(X)).$$
In più, la successione $nH(S_n)$ è super-additiva (una successione a_n è super-additiva se $a_{n+m} \geq a_n + a_m$), e in particolare la convergenza è monotona sulla sottosuccessione $n_k = 2^k$.

La dimostrazione di questo si trova in [4]. È importante notare che la super-additività deriva dalla disuguaglianza

$$H(\sqrt{\lambda}X + \sqrt{1-\lambda}Y) \geq \lambda H(X) + (1-\lambda)H(Y),$$

che è dimostrabile tramite la disuguaglianza di Young con costante ottimale come fatto in [5].

È stato dimostrato in [6] che la convergenza è monotona in generale, come congetturato originariamente da Shannon, e non solo sulla particolare sottosuccessione delle potenze di due.

In analogia con questo, possiamo mostrare che anche col minimo succede questa cosa, ma la convergenza non risulterà avere belle proprietà, almeno nei primi termini.

Teorema 3.6 *Data una successione di v.a. i.i.d. $(X_i)_{i\geq 1}$ non negative, a entropia finita e con densità f continua in 0 e non nulla in 0, allora*

$$H(M_n) \to H(Exp(f(0))) = 1 - \log f(0).$$

$\mathcal{D}im$: per quanto detto due paragrafi fa, la v.a. M_n ha densità

$$f\left(\frac{x}{n}\right)\left(1 - F\left(\frac{x}{n}\right)\right)^{n-1},$$

pertanto la sua entropia è

$$\int_0^\infty -f\left(\frac{x}{n}\right)\left(1 - F\left(\frac{x}{n}\right)\right)^{n-1} \log f\left(\frac{x}{n}\right)dx +$$
$$+ \int_0^\infty -(n-1)f\left(\frac{x}{n}\right)\left(1 - F\left(\frac{x}{n}\right)\right)^{n-1} \log\left(1 - F\left(\frac{x}{n}\right)\right).$$

Puntualmente, il primo termine tende a $-f(0)e^{-f(0)x}\log f(0)$, che integrato fa $-\log f(0)$. Il secondo invece tende a $f(0)^2 xe^{-f(0)x}$, che integrato fa 1. Non riusciremo a trovare una dominazione, quindi questo è solo una motivazione a continuare. Cambiando variabile otteniamo:

$$\int -nf(1-F)^{n-1}\log f - n(n-1)f(1-F)^{n-1}\log(1-F).$$

Sia $r \in \mathbb{N}$ tale che f sia compresa fra $f(0)-\varepsilon$ e $f(0)+\varepsilon$ su $[0,r]$. Visto che nx^{n-1} tende a 0 su $[0,1[$, con la convergenza che è uniforme sui compatti, e visto che $f\log f \in L^1$, il primo addendo ha la coda dominata tendente a 0 (la coda è quella da r in poi). Per quanto riguarda l'inizio, l'integrale fra 0 e r è compreso fra l'integrale fra 0 e r di $-nf(1-F)^{n-1}\log(f(0)-\varepsilon)$ e quello di $-nf(1-F)^{n-1}\log(f(0)+\varepsilon)$. Facciamo solo il primo visto che l'altro è analogo:

$$\int_0^r -nf(1-F)^{n-1}\log(f(0)-\varepsilon) = \log(f(0)-\varepsilon)\int_0^r \frac{d}{dx}(1-F)^n =$$

$$= \log(f(0)-\varepsilon)\cdot((1-F(r))^n - 1) \to -\log(f(0)-\varepsilon);$$

pertanto, è vero che l'integrale del primo pezzo tende a $-\log f(0)$ per la generalità di ε.
Per il secondo pezzo, integriamo per parti (ci occuperemo dopo della legittimità dei passaggi):

$$\int -n(n-1)f(1-F)^{n-1}\log(1-F) = \int (n-1)\frac{d}{dx}(1-F)^n\log(1-F) =$$

$$= (n-1)(1-F)^n\log(1-F)|_0^\infty + \int (n-1)f(1-F)^{n-1} =$$

$$= -\frac{n-1}{n}(1-F)^n|_0^\infty = \frac{n-1}{n} \to 1,$$

da cui si avrebbe la tesi in toto. Per poter integrare per parti, basta che entrambe le funzioni da integrare siano integrabili (e derivabili, che è chiaro), ma questo è vero perché f sta in L^1 e $(1-F)^{n-1}\log(1-F)$ sta in L^∞ per $n > 1$. $\square$

In generale, questa convergenza non sarà monotona: infatti, ad $H(X)$ fissato, $f(0)$ ci determina se la convergenza sarà crescente o decrescente (se per assurdo fosse sempre monotona), ma noi possiamo cambiare $f(0)$ e un suo piccolo intorno senza cambiare significativamente l'entropia di X, e quindi la direzione di convergenza, che ci porta ad un assurdo.

3.3 Informazione di Fisher

Vista la forma che ha il TLC col minimo, ci chiediamo se esista una caratterizzazione variazionale dell'entropia a densità in 0 fissata. Questa caratterizzazione è data dall'informazione di Fisher. Prima di dimostrare che passa al limite per TLC, ci fermeremo a confutare alcune congetture, e a introdurre un semigruppo che potrebbe essere utile in relazione all'operazione di minimo.

Definizione 3.7 (Informazione di Fisher) *Data una v.a. X a valori in $\mathbb{R}^n$ con densità f strettamente positiva ovunque e C^1, definiamo la sua* **informazione di Fisher** *come*

$$I(X) = I(f) := \int_{\mathbb{R}^n} \frac{|\nabla f|^2}{f} dx.$$

Nel caso del minimo, ci sarà f' anziché ∇f, e l'integrale sarà su $\mathbb{R}^+$. Inoltre,

$$I(f) = \int_0^\infty |\frac{d}{dx}(\log f(x))|^2 \cdot f(x) dx = 4 \int_0^\infty |\frac{d}{dx}\sqrt{f(x)}|^2 dx$$

Proposizione 3.8 *Fra le densità C^1 con $f(0) = f_0$ fissato, quella che minimizza l'informazione di Fisher è la densità esponenziale γ_{f_0}.*

$\mathcal{D}im$: sia $g := \sqrt{f}$. Il problema variazionale corrispondente è

$$\int |g'|^2 dx + c \int g^2,$$

le cui equazioni di Eulero-Lagrange con moltiplicatori sono $g'' = cg$, da cui g è esponenziale, perciò anche f è un'esponenziale, e il parametro è proprio il valore iniziale. È un minimo per teoremi standard di calcolo delle variazioni (è convesso nella derivata, e c è fissato). $\square$

Per completezza, riportiamo che l'informazione di Fisher si comporta bene per riscalamento, cioè $I(\lambda X) = \frac{I(X)}{\lambda^2}$. In particolare, visto che $I(Exp(1)) = 1$,vale $I(Exp(\lambda)) = \lambda^2$.

3.4 Congetture sull'informazione di Fisher e semigruppo

Fra le prime cose che si possono pensare di dimostrare, ma che risulta falsa, c'è la seguente:

Proposizione 3.9 *Al variare di X, Y v.a. indipendenti con $I(X), I(Y)$ fissate, la quantità $I(X \wedge Y)$* **non** *è minima per X e Y esponenziali.*

$\mathcal{C}$ontresempio: se per assurdo fosse minima per variabili esponenziali, allora notando che $I(Exp(f(0))) = f(0)^2$, si avrebbe anche la seguente disuguaglianza:

$$\sqrt{I(X \wedge Y)} \geq \sqrt{I(X)} + \sqrt{I(Y)}.$$

Ma questa è falsa: infatti, possiamo concentrare X e Y su due insiemi diversi (ad esempio su $[0,1]$ e su $[2,3]$), di modo che il minimo fra le due sia proprio X, e questo porta alla falsità della tesi. È possibile fare tutto formalmente, scegliendo opportunamente una successione di variabili aleatorie. Descriviamo brevemente come fare per completezza: per approssimare la densità f di Y, che supponiamo a valori in $[1,2]$ e C^1, creiamo g in questo modo. Fissiamo $a \in [0,1]$, $b > 0$ tali che $af + b$ sia con integrale più piccolo di $1 - b$. Poniamo $g \equiv b$ su $[0,1]$, $g \equiv af + b$ su $[1,2]$ e esponenziale (con un piccolo raccordo C^∞) su $[2, \infty[$ di modo che g risulti essere una densità di probabilità C^1. È chiaro che, per $a \to 1$, $b \to 0$, l'informazione di Fisher di f e g sarà la stessa, e con questa approssimazione la dimostrazione diventa formale. $\square$

Ora introduciamo un semigruppo che potrebbe aiutare ad ottenere nuovi risultati: poniamo

$$U_s(X) := e^s X \wedge Exp(f(0)(1 - e^{-s}))$$

per le v.a. X con densità continua in zero, e $U_s(f_X)$ la densità di $U_s(X)$. È effettivamente un semigruppo, poiché notando che $U_s(f)(0) = f(0)$,

$$\begin{aligned} U_s(U_{s'}(X)) &= e^s e^{s'} X \wedge e^s Exp(f(0)(1 - e^{-s'})) \wedge Exp(f(0)(1 - e^{-s})) = \\ &= e^{s+s'} X \wedge Exp(f(0)(1 - e^{-s} + e^{-s} - e^{-s-s'})), \end{aligned}$$

da cui la tesi. Inoltre, questo semigruppo è continuo e $U_s(X) \to Exp(f(0))$ per $s \to \infty$. Va notato che non è un semigruppo lineare, ma vale $U_s(f \triangle g) = (U_s f) \triangle (U_s g)$.

Questo semigruppo ha la particolarità di preservare $f(0)$ lungo le traiettorie (è il motivo per cui è stato introdotto), e inoltre le sue traiettorie collegano X con $Exp(f(0))$, quindi è naturale chiedersi se ad esempio l'informazione di Fisher decresce lungo queste traiettorie. Questo risulta essere falso, come vedremo fra poco, ma prima abbiamo bisogno di dare una definizione.

Definizione 3.10 (Generatore) *Il* **generatore** *di un semigruppo continuo U_s da uno spazio di Banach B in sè è l'operatore non limitato $A : D(A) \to B$ definito come*

$$Ax = -\lim_{s \downarrow 0} \frac{U_s x - x}{s},$$

con $D(A)$ insieme dove questo limite esiste.

Il generatore del semigruppo appena introdotto è

$$Af(x) = -f'(x) - f(x) - xf(0)f(x) + f(0)(1 - F(x)).$$

Una cosa che ci si può chiedere a riguardo è se la funzione $s \mapsto I(U_s(f))$ sia decrescente, o addirittura convessa. Anche questa risulta essere falsa.

Proposizione 3.11 *La funzione* $s \mapsto I(U_s(f))$ **non** *è decrescente. In particolare, non è nemmeno convessa.*

$\mathcal{C}$ontresempio: per le funzioni che sceglieremo sarà possibile usare il limite sotto il segno di integrale. Pertanto,

$$\frac{\partial}{\partial s} I(U_s(f))|_{s=0} = \int 2\frac{f' \cdot (Af)'}{f} - \frac{f'^2 \cdot Af}{f^2} =$$

$$= \int \frac{-2ff'^2 - 2ff'f'' - 4f(0)f^2f' - 2xf(0)ff'^2}{f^2} +$$

$$+ \frac{f'^3 + ff'^2 + xf(0)ff'^2 - f(0)f'^2(1-F)}{f^2} =$$

$$= \int \frac{-ff'^2 - 2ff'f'' - 4f(0)f^2f' - xf(0)ff'^2 + f'^3 - f(0)f'^2(1-F)}{f^2}$$

dove abbiamo usato che $(Af)' = -f'' - f' - 2f(0)f - xf(0)f'$ se f è almeno C^2 a tratti.
Fra questi termini, il primo, il quarto e il sesto sono sicuramente negativi. Inoltre, se f è decrescente anche il quinto è negativo, e il secondo segue il segno di f''. In particolare, all'infinito solo il secondo e il terzo saranno positivi di solito. Se proviamo a riscalare di λ, otteniamo

$$\frac{\partial}{\partial s} I(T_s(\frac{1}{\lambda}f(\frac{x}{\lambda}))|_{s=0} =$$

$$= \int -\frac{1}{\lambda^2}\frac{ff'^2 + 4f(0)f^2f' + f(0)f'^2(1-F)}{f^2} - \frac{1}{\lambda^3}\frac{2ff'f'' + xf(0)ff'^2 - f'^3}{f^2}.$$

Allora ci basta mostrare che uno di questi due addendi ha integrale negativo per mostrare la falsità della tesi (nel caso del primo, basta fare il limite per $\lambda \to \infty$, nel caso del secondo per $\lambda \to 0$). Scegliamo il secondo, in cui non compare F.
Cerchiamo il controesempio fra le funzioni C^2 che fino a un certo x_0 sono lineari, e da lì in poi sono esponenziali. In particolare, fissato $p(x_0) = p := \frac{x_0^2}{2} + x_0 + 1 > 0$, sia $f(x)$ uguale a $\frac{1+x_0}{p} - \frac{x}{p}$ fra 0 e x_0, uguale a $\frac{e^{x_0 - x}}{p}$ da x_0 in poi. È facile verificare che f è continua, C^1, C^2 a tratti ed è una densità di probabilità. Inoltre, tornando al calcolo della derivata dell'informazione di Fisher, l'integrale fino a x_0 è negativo, quindi deve essere per forza positivo l'integrale dopo x_0. Ma questo vale, dopo aver traslato di x_0,

$$\int \frac{2pe^{-x} - (x + x_0)(1 + x_0)e^{-x} - pe^{-x}}{p^2} = \frac{1}{p^2}(p - (1 + x_0)^2),$$

che ha lo stesso segno di $p - (1 + x_0)^2 = -\frac{x_0^2}{2} - x_0$, che per x_0 grande è negativo, da cui la tesi. $\square$

3.5 Convergenza Fisher per TLC

Come l'entropia, anche l'informazione di Fisher si comporta bene sia col nostro teorema del limite centrale, sia con quello standard della somma.

Teorema 3.12 *Se ho una successione di v.a. i.i.d. X_i centrate a informazione di Fisher finita, allora*

$$I(S_n) \to I(\mathcal{N}(0, Var(X)).$$

In più, la successione $nH(I_n)$ è sub-additiva, e in particolare la convergenza è monotona sulla sottosuccessione $n_k = 2^k$.

La dimostrazione di questo si trova in [4].
Analogamente, vale il seguente:

Teorema 3.13 *Sia X_i una successione di v.a. i.i.d. a informazione di Fisher finita e con densità $f \in C^1$ non nulla in 0. Se f è limitata, allora*

$$I(M_n) \to I(Exp(f(0))) = f(0)^2.$$

$\mathcal{D}im$: ricordiamo ancora una volta che la densità di M_n è

$$f\left(\frac{x}{n}\right)\left(1 - F\left(\frac{x}{n}\right)\right)^{n-1},$$

la cui derivata prima, omettendo l'argomento di tutte le funzioni che sarà sempre x/n, è

$$\frac{1}{n}f'(1-F)^{n-1} + \frac{n-1}{n}f^2(1-F)^{n-2}.$$

Calcoliamo a parte anche il suo quadrato, sempre omettendo l'argomento:

$$\frac{1}{n^2}(f')^2(1-F)^{2n-2} - 2\frac{n-1}{n^2}f^2 f'(1-F)^{2n-3} + \left(\frac{n-1}{n}\right)^2 f^4(1-F)^{2n-4}.$$

Dunque vale, tenendo come argomento sottinteso x/n,

$$I(M_n) = \int \frac{1}{n^2}\frac{(f')^2}{f}(1-F)^{n-1} - 2\frac{n-1}{n^2}ff'(1-F)^{n-2} + \left(\frac{n-1}{n}\right)^2 f^3(1-F)^{n-3},$$

perciò dopo aver cambiato variabile otteniamo

$$\int \frac{1}{n}\frac{(f')^2}{f}(1-F)^{n-1} - 2\frac{n-1}{n}ff'(1-F)^{n-2} + \frac{(n-1)^2}{n}f^3(1-F)^{n-3}.$$

Il primo pezzo converge a 0 (serve che f'^2/f sia integrabile, cioè informazione finita), quindi guardiamo il secondo. Integrando per parti,

$$\int 2ff'(1-F)^{n-2} = \int \frac{d}{dx}(f^2)(1-F)^{n-2} =$$

$$= f^2(1-F)^{n-2}|_0^\infty + (n-2)\int f^3(1-F)^{n-3} = -f^2(0) + (n-2)\int f^3(1-F)^{n-3}.$$

Questi passaggi sono leciti perché, visto che f è limitata, hanno senso almeno due termini su tre, quindi anche il terzo ha senso (ovvero $\int ff'(1-F)^{n-2}$). L'integrale che compare, dopo essere stato moltiplicato per $\frac{n-1}{n}$ e sommato al terzo pezzo, diventa

$$\int \frac{1}{n} f^3(1-F)^{n-3},$$

che tende a 0 esattamente come il primo pezzo. Pertanto, l'informazione di Fisher di M_n ha lo stesso limite di $\frac{n-1}{n}f^2(0)$, che tende a $f^2(0) = I(Exp(f(0)))$, da cui la tesi. $\square$

Come si evince dalla dimostrazione, bastano ipotesi più deboli su f (limitata all'infinito e in L^3, ad esempio). Questo conclude la sezione sul TLC.

3.6 Esponenziali e trasporto ottimo di massa

Il trasporto ottimo di massa è una branca della matematica che sta godendo di ampio interesse, attualmente. Ci sono due cose che vorremmo guardare: la prima è studiare il costo di trasporto $c(x, y) := x \wedge y$. La seconda, mostrare che rispetto alla distanza di Kantorovich-Wasserstein sia la convoluzione, sia $\triangle$ sono delle contrazioni.

Facciamo dunque una breve introduzione al trasporto ottimo. Tutto quello che verrà detto si può trovare in [7], con le dimostrazioni e con un percorso storicamente accurato.

Nella formulazione moderna del trasporto ottimo, dovuta a Kantorovich, sono dati due spazi A, B e una funzione $c : A \times B \to [0, \infty]$ detta "costo di trasporto", che rappresenta quanto costa portare qualcosa dal punto x al punto y. Sono date anche due misure di probabilità μ, ν su A e su B rispettivamente, che rappresentano la distribuzione iniziale di massa. Quello che si vuole fare è minimizzare

$$\int_{A \times B} c(x, y) d\pi$$

al variare di π nelle misure di probabilità su $A \times B$ tali che $P_{A\sharp}\pi = \mu$, $P_{B\sharp}\pi = \nu$. Il risultato principale è il seguente:

Teorema 3.14 *Se A e B sono spazi polacchi (metrizzabili a base numerabile) e $c : A \times B \to [0, \infty]$ è un costo semicontinuo inferiormente, allora il minimo è ottenuto per qualche misura π.*

In particolare, se $A = B = \mathbb{R}^+$, e $c(x, y) = x \wedge y$, il minimo del problema di trasporto associato è ottenuto. Possiamo formulare il problema di trasporto in un modo diverso, che ci tornerà comodo per quello che vorremo dire. Anziché lavorare con le misure sul prodotto, lavoriamo con le v.a. , e in particolare minimizziamo

$$\mathbb{E}[c(X, Y)]$$

al variare di X, Y variabili aleatorie a valori in A, B rispettivamente, con leggi μ, ν rispettivamente. È facile verificare che queste due formulazioni sono equivalenti.

Chiamiamo $\mathcal{C}(\mu, \nu)$ il valore ottenuto minimizzando il costo di trasporto con leggi μ, ν rispettivamente, e chiamiamo $\mu \triangle \nu$ la legge del minimo fra le v.a. indipendenti X e Y aventi leggi rispettivamente μ, ν. Allora vale:

Proposizione 3.15 *Date μ, ν e η, vale*

$$\mathcal{C}(\mu, \nu) \geq \mathcal{C}(\mu, \nu \triangle \eta).$$

$\mathcal{D}im$: siano X, Y le variabili che minimizzano il costo con μ e ν. A meno di allargare lo spazio su cui è definito Y (cosa che è possibile lasciando invariata la legge di Y), possiamo supporre che esista nello stesso spazio una v.a. Z, indipendente da Y, con legge η. Allora $Y \wedge Z$ ha legge $\nu \triangle \eta$ per definizione, e vale

$$\mathcal{C}(\mu, \nu \triangle \eta) \leq E[X \wedge (Y \wedge Z)] \leq E[X \wedge Y] = \mathcal{C}(\mu, \nu),$$

da cui la tesi. □

Possiamo applicare questo risultato al semigruppo U_s: definiamo $U_s\mu$ come la legge di U_sX con X v.a. con legge μ. Allora se $s > s'$, vale

$$\mathcal{C}(U_s\mu, \nu) = \mathcal{C}((e^{s-s'}e^{s'}X) \wedge Exp(f(0)(1 - e^{-s})), \nu) =$$

$$= \mathcal{C}(e^{s-s'}U_{s'}\mu \triangle Exp(f(0)(1 - e^{s'-s})), \nu) \leq \mathcal{C}(e^{s-s'}U_{s'}\mu, \nu).$$

Visto che $\mathcal{C}(a\mu, a\nu) = a\mathcal{C}(\mu, \nu)$, dove la moltiplicazione è sempre da intendersi relativa alle leggi, vale

$$\mathcal{C}(U_sX, U_sY) \leq e^{s-s'}\mathcal{C}(U_{s'}X, U_{s'}Y).$$

Possiamo anche dedurne un importante corollario sul costo di trasporto:

Corollario 3.16 *Siano $(X_i)_{i\geq 1}$ e $(Y_i)_{i\geq 1}$ due successioni di v.a. i.i.d. indipendenti. Siano $M_n^X := nX_1 \wedge \ldots \wedge X_n$, $M_n^Y := nY_1 \wedge \ldots \wedge Y_n$. Allora la successione $n \mapsto \mathcal{C}(M_n^X, M_n^Y)$è sub-additiva. In particolare, sulla sottosuccessione delle potenze di 2 è monotona.*

$\mathcal{D}im$: chiamiamo $\bar{M}_n^X := X_1 \wedge \ldots \wedge X_n = M_n^X/n$, e similmente $\bar{M}_n^Y$. Dati $n, m \geq 1$, vale

$$\mathcal{C}(\bar{M}_n^X \wedge (X_{n+1} \wedge \ldots \wedge X_{n+m}), \bar{M}_n^Y \wedge (Y_{n+1} \wedge \ldots \wedge Y_{n+m})) \leq$$

$$\leq \mathcal{C}(\bar{M}_n^X, \bar{M}_n^Y \wedge (Y_{n+1} \wedge \ldots \wedge Y_{n+m})) \leq \mathcal{C}(\bar{M}_n^X, \bar{M}_n^Y).$$

In particolare,

$$\mathcal{C}(\bar{M}_{n+m}^X, \bar{M}_{n+m}^Y) \leq \min\{\mathcal{C}(\bar{M}_n^X, \bar{M}_n^Y), \mathcal{C}(\bar{M}_m^X, \bar{M}_m^Y)\} \leq$$

$$\leq \frac{n}{n+m}\mathcal{C}(\bar{M}_n^X, \bar{M}_n^Y) + \frac{m}{n+m}\mathcal{C}(\bar{M}_m^X, \bar{M}_m^Y),$$

ovvero la successione $n \mapsto n\mathcal{C}(\bar{M}_n^X, \bar{M}_n^Y)$ è sub-additiva. Visto che

$$n\mathcal{C}(\bar{M}_n^X, \bar{M}_n^Y) = \mathcal{C}(M_n^X, M_n^Y),$$

abbiamo la tesi. □

Una proprietà simile vale per il costo di trasporto $c(x, y) := |x - y|^2$ con il TLC della somma. Tuttavia, come mostrato in [8], in generale la convergenza non sarà monotona sull'intera successione in questo secondo caso, quindi ci aspettiamo che neanche per il minimo la convergenza sarà monotona.

3.7 Distanza di Kantorovich-Wasserstein

La domanda alla quale questa distanza risponde è la seguente: la convergenza stretta di misure è metrizzabile? Sì, se ci restringiamo alle misure a media finita. Questa distanza nasce dalla teoria del trasporto ottimo tramite la dualità.

Definizione 3.17 (Distanza di Kantorovich-Wasserstein) *Date* $\mu, \nu \in P_1(\mathbb{R})$*, cioè misure di probabilità con momento primo finito, la loro* **distanza di Kantorovich-Wasserstein** *è*

$$W_1(\mu, \nu) := \sup_{f \in Lip_1(\mathbb{R})} \left\{ \int f d\mu - \int f d\nu \right\}.$$

L'operazione di convoluzione è una contrazione rispetto a questa distanza, nel senso dell'enunciato seguente:

Proposizione 3.18 *Data* η *misura di probabilità con densità* ρ*, e* $\mu, \nu \in P_1(\mathbb{R})$*, allora vale*

$$W_1(\mu * \eta, \nu * \eta) \leq W_1(\mu, \nu),$$

dove la convoluzione fra due misure è intesa, in analogia con la sezione precedente, come operazione sulle leggi.

$\mathcal{D}im$: presa f 1-lipschitziana, vale

$$\int f(x) d(\mu * \eta)(x) = \int f(x) \int \mu(dy) \rho(x-y) dx =$$

$$= \int \mu(dy) \Big(\int f(x) \rho(x-y) dx \Big).$$

Il secondo integrale, dopo il cambio di variabile $z = x - y$, diventa

$$\int f(z+y) \rho(z) dz,$$

che è una funzione 1-lipschitziana in y. In particolare,

$$W_1(\mu * \eta, \nu * \eta) = \sup_{f \in Lip_1(\mathbb{R})} \left\{ \int f d(\mu * \eta) - \int f d(\nu * \eta) \right\} =$$

$$= \sup_{f \in Lip_1(\mathbb{R})} \left\{ \int \mu(dy) \Big(\int f(x) \rho(x-y) dx \Big) - \int \nu(dy) \Big(\int f(x) \rho(x-y) dx \Big) \right\} \leq$$

$$\leq \sup_{f \in Lip_1(\mathbb{R})} \left\{ \int f d\mu - \int f d\nu \right\} = W_1(\mu, \nu),$$

cioè la tesi. □

La distanza W_1 può essere estesa a distanza fra variabili aleatorie (al posto che misure di probabilità) con momento primo senza difficoltà. Inoltre, $W_1(\mu,\nu)$ corrisponde alla soluzione del problema di trasporto ottimo con costo $c(x,y) := |x-y|$, grazie al teorema di dualità.

Anche il semigruppo del minimo è una contrazione rispetto a questa distanza:

Proposizione 3.19 *Data η misura di probabilità con densità ρ, e $\mu,\nu \in P_1(\mathbb{R}^+)$, allora vale*

$$W_1(\mu\triangle\eta, \nu\triangle\eta) \leq W_1(\mu,\nu)$$

$\mathcal{D}im$: usiamo la formulazione con le 1-lipschitziane. Procederemo per approssimazione, supponendo che μ e ν abbiano densità g_1, g_2: allora, presa f 1-lipschitziana, e indicando con R la funzione di ripartizione di ρ, vale

$$\int f(x)d\mu\triangle\eta(x) = \int f(x)g_1(x)(1-R(x))dx + \int f(x)\rho(x)(1-G_1(x))dx.$$

Integriamo per parti il secondo, chiamando $F_\rho(x) := \int_0^x f(t)\rho(t)dt$:

$$\int f\rho(1-G_1)dx = F_\rho(1-G_1)|_0^\infty + \int g_1(x)F_\rho dx.$$

Allora

$$\int f(x)d\mu\triangle\eta(x) = \int \Big(f(x)(1-R(x)) + F_\rho(x)\Big)d\mu(x),$$

e in particolare se $\psi := f(1-R) + F_\rho$ è 1-lipschitziana abbiamo la tesi. Notiamo che ci basta mostrarlo localmente. Fissato x_0, scegliamo un suo intorno della forma $I_\delta := [x_0-\delta, x_0+\delta]$. Allora per ogni x in I_δ vale

$$\psi(x) - \psi(x_0) = (f(x) - f(x_0))\int_x^\infty \rho(t)dt + \int_{x_0}^x (f(t)-f(x_0))\rho(t)dt,$$

dunque prendendo il valore assoluto risulta

$$|\psi(x)-\psi(x_0)| \leq |x-x_0|\int_{x_0-\delta}^\infty \rho(t)dt + |x-x_0|\int_{I_\delta}\rho(t)dt =$$

$$= |x-x_0|\bigg(\int_{x_0-\delta}^\infty \rho(t)dt + \int_{I_\delta}\rho(t)dt\bigg).$$

Se possiamo scegliere δ in modo che il valore tra parentesi sia più piccolo di 1, abbiamo finito. Se $R(x_0)$ è non nullo, è sempre possibile. Invece, fra gli x_0 con $R(x_0)=0$, ne esiste uno solo tale che $\int_{I_\delta}\rho(t)dt > 0$ per ogni $\delta > 0$. Negli altri, basta scegliere δ piccolo in modo che l'integrale sia nullo. Dunque, al di fuori di un punto, abbiamo mostrato che ψ è 1-lipschitziana. Ma una funzione continua e 1-lipschitziana su tutto il dominio tranne un punto è 1-lipschitziana, quindi si può concludere come nel caso della convoluzione.

Il procedimento di approssimazione è facile: infatti, ogni misura può essere approssimata (nel senso della distanza di Kantorovich-Wasserstein) con misure assolutamente continue. Inoltre, $\mu \mapsto \mu \triangle \rho$ è continua nella topologia della convergenza stretta, e in particolare se $\mu_n \to \mu$, allora $\mu_n \triangle \rho \to \mu \triangle \rho$ (convergono puntualmente le funzioni di ripartizione, ad esempio). La tesi segue approssimando μ e ν e usando la disuguaglianza triangolare. □

Notiamo infine che è cruciale il fatto che l'operazione di minimo sia fatta per variabili indipendenti: infatti, è facile trovare un controesempio, che scriveremo in termini di v.a. per comodità. Basta prendere, ad esempio, sullo spazio $\Omega = [0, 2]$ con probabilità uniforme le v.a. $X = I_{[1,2]}$, $Y = I_{[0,1]}$ e $Z = 2I_{[0,1]} + \frac{1}{2}I_{[1,2]}$.

Bibliografia

[1] E.H. Lieb, M. Loss, and American Mathematical Society. *Analysis*. Crm Proceedings & Lecture Notes. American Mathematical Society, 2001.

[2] Giuseppe Toscani. Heat equation and convolution inequalities, 2013.

[3] Claude E. Shannon. A mathematical theory of communication. *Bell Syst. Tech. J.*, 27(3):379–423, 1948.

[4] Andrew R. Barron. Entropy and the central limit theorem. *The Annals of Probability*, 14(1):336–342, 1986.

[5] A. Dembo, T.M. Cover, and J.A. Thomas. Information theoretic inequalities. *IEEE Transactions on Information Theory*, 37(6):1501–1518, 1991.

[6] Shiri Artstein, Keith Ball, Franck Barthe, and Assaf Naor. Solution of shannon's problem on the monotonicity of entropy. *Artstein, S. and Ball, K.M. and Barthe, F. and Naor, A. (2004) Solution of Shannon's problem on the monotonicity of entropy. Journal of the American Mathematical Society, 17 (4). pp. 975-982. ISSN 08940347*, 17, 10 2004.

[7] Luigi Ambrosio, Klaus Deckelnick, Gerhard Dziuk, Masayasu Mimura, Vsevolod Solonnikov, and H. Soner. *Lecture Notes on Optimal Transport Problems*. Springer, 01 2003.

[8] Walter Schachermayer, Uwe Schmock, and Josef Teichmann. Non-monotone convergence in the quadratic wasserstein distance. In *Séminaire de Probabilités XLII*, pages 131–136. Springer, 2009.