



UNIVERSITÀ DI PISA

DIPARTIMENTO DI MATEMATICA
TESI DI LAUREA TRIENNALE

**Un'introduzione
alla statistica asintotica
attraverso una generalizzazione
del Teorema Limite Centrale**

CANDIDATO:
Carminè Frascella

RELATORE:
Prof. Maurizio Pratelli

ANNO ACCADEMICO 2014/2015

*A papà,
mamma,
Irene,
e Maria Ausiliatrice.*

Indice

Introduzione e Ringraziamenti	7
1 Prime definizioni riguardo la statistica asintotica	9
1.1 Esperimenti e confronto di rischi	9
1.2 Una distanza tra esperimenti	14
1.3 Rappresentazione canonica di Blackwell	16
1.4 Un'osservazione sulla convergenza di misure	24
2 Contiguità e Trasformata di Hellinger	27
2.1 Contiguità	27
2.2 Distanza di Hellinger	36
2.3 Trasformata di Hellinger	38
3 Shift gaussiani e esperimenti poissoniani	41
3.1 Shift gaussiani	41
3.2 Esperimenti poissoniani	48
4 Alcuni risultati preliminari	55
4.1 La procedura di Lindeberg	55
4.2 Lo split di Lévy	58
4.3 La disuguaglianza di simmetrizzazione di Lévy	60
4.4 Condizioni per la shift-compattezza	61
4.5 Differenziabilità secondo Peano e secondo Fréchet	65
5 Un Teorema Limite Centrale per “array” infinitesimali	67
5.1 Teorema Limite Centrale: una prima generalizzazione	67
5.2 I tre lemmi di Le Cam e una seconda generalizzazione	72
Bibliografia	81

Introduzione e Ringraziamenti

L'obiettivo principale di questa tesi è quello di introdurre alcuni concetti e risultati di grande rilevanza nell'ambito della statistica asintotica, ossia quel settore della statistica, sviluppatasi nella seconda metà del XX secolo, che studia e formalizza metodi per approssimare famiglie di misure (ossia esperimenti statistici) mediante altre famiglie di misure, matematicamente più trattabili. Un esperimento statistico parametrico, infatti, non è altro che una famiglia di misure di probabilità su un certo spazio, che denoteremo quasi sempre con Ω , indicizzate da un insieme Θ di parametri, di cardinalità arbitraria. Proseguendo nella lettura, poi, l'attenzione si sposta su determinate classi di variabili aleatorie: viene presentata, quindi, una generalizzazione del Teorema Limite Centrale, e mediante questa si ritorna poi, nelle sezioni conclusive, ad analizzare approssimazioni e limiti riguardanti successioni di esperimenti statistici.

Nel primo capitolo vengono semplicemente date le definizioni fondamentali, che serviranno in tutto l'elaborato, riguardanti in generale la statistica decisionale: viene quindi data la nozione di esperimento statistico, e vengono introdotte le nozioni di funzione di costo e di rischio da essi indotte. Viene poi definita una distanza tra esperimenti, utile per rendere formale il concetto di convergenza di una successione di esperimenti ad un dato esperimento limite, e poi si introduce la rappresentazione canonica di Blackwell, per esprimere in maniera equivalente e matematicamente più trattabile la distanza tra esperimenti. Si definisce quindi la norma dual-Lipschitziana, una norma di misure che induce, al pari della più nota norma di Lévy-Prohorov, la convergenza in legge, e si conclude con alcune osservazioni, per l'appunto, sulla convergenza stretta di una successione di misure definite sulla retta reale.

Nel secondo capitolo viene invece introdotta, mediante alcuni teoremi, la nozione di contiguità tra due successioni di probabilità, in varie formulazioni tutte tra loro equivalenti: essa può essere vista come una versione asintotica della definizione di assoluta continuità tra due misure. Nelle sezioni finali, invece, vengono introdotte la distanza di Hellinger, l'affinità e la trasformata di Hellinger, particolarmente funzionali quando si hanno diversi esperimenti statistici effettuati in sequenza, ognuno indipendentemente dai risultati ottenuti effettuando i precedenti.

Nel terzo capitolo si focalizza l'attenzione su due classi di esperimenti statistici, gli esperimenti gaussiani e quelli poissoniani: i primi, decisamente più usati per via del Teorema Limite Centrale; i secondi, meno trattabili dal punto di vista computazionale, ma più importanti per i nostri fini. In entrambi i casi, attraverso alcune proposizioni, vengono analizzate le proprietà di cui godono queste due classi di esperimenti.

Nel quarto capitolo vengono presentati alcuni risultati, dovuti in gran parte a J.W.Lindeberg e P.Lévy, interessanti di per sè ma utili per presentare poi i risultati principali di questa tesi, nel capitolo seguente. In particolare, viene presentata una dimostrazione elementare di una prima versione del Teorema Limite Centrale, nella sezione iniziale del capitolo. Viene poi introdotta la nozione di shift-compattezza per una successione di misure di probabilità sulla retta, la quale è una generalizzazione della relativa compattezza (equivalente alla condizione di tensione, per il teorema di Prohorov), e vengono mostrati alcuni interessanti risultati a riguardo. Si conclude, quindi, presentando la nozione di differenziabilità di una funzione in un punto secondo Peano e secondo Frechét.

Nel quinto e ultimo capitolo si vedono, infine, i risultati principali di questa tesi: viene quindi enunciata e dimostrata una versione generalizzata del Teorema Limite Centrale, e poi tale risultato viene ancora generalizzato mediante l'uso di tre lemmi, tecnici ma dimostrabili elementarmente, che J.Hajek ha denominato "i tre lemmi di Le Cam". In particolare, nelle versioni più note del Teorema Limite Centrale, vi è sempre l'ipotesi di identica distribuzione delle variabili, ipotesi che qui viene sostituita con altre molto più deboli. Si conclude quindi vedendo esplicitamente alcuni risultati, riguardanti successioni di esperimenti statistici binari (ossia tali che $\Theta = \{0, 1\}$), sfruttando il fatto che essi si possono studiare mediante i rapporti di verosimiglianza, che sono variabili aleatorie, e usando quindi i teoremi visti precedentemente.

Desidero ringraziare anticipatamente innanzitutto la mia famiglia, ma in generale chiunque mi abbia aiutato a raggiungere questo traguardo così importante: fare una lista sarebbe impossibile, oltre che inutile. Un sentito ringraziamento al mio relatore, il Professor Maurizio Pratelli, sempre disponibile in tutto il periodo che ho impiegato per scrivere questa tesi. Un ringraziamento speciale anche alla mia amica Graziana, che ha accettato senza esitazione di leggerla per valutarne le parti testuali. Mi scuso sin da ora per eventuali sviste e imprecisioni, sicuramente presenti in questo elaborato, assumendome tutte le responsabilità.

Carmine Frascella

Capitolo 1

Prime definizioni riguardo la statistica asintotica

In questo capitolo verranno introdotte alcune definizioni molto importanti nell'ambito della statistica, che useremo anche nei capitoli successivi.

Vedremo in particolare cosa è un **esperimento statistico**, con la definizione data da D.Blackwell nel 1951, e introdurremo una distanza tra esperimenti: useremo tale distanza, poi, per definire il concetto di convergenza di una successione di esperimenti ad un dato esperimento limite. Daremo varie caratterizzazioni di questa distanza, alcune delle quali ci aiuteranno dal punto di vista prettamente computazionale: a prima vista, infatti, essa apparirà difficilmente calcolabile.

Concluderemo con alcune osservazioni sulla convergenza di misure definite su $(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$.

1.1 Esperimenti e confronto di rischi

Iniziamo introducendo il concetto di **problema statistico decisionale**, rifacendoci a A.Wald.

Supponiamo che uno statistico osservi un fenomeno, la cui legge di probabilità dipende da un parametro $\theta \in \Theta$, con lo scopo di effettuare una scelta $z \in Z$ (che quindi è l'insieme delle possibili scelte, o decisioni).

La scelta dell'azione z comporta come conseguenza un **costo** $W(\theta, z) \in]-\infty, +\infty]$, dunque è ben definita una **funzione di costo** W , definita in $\Theta \times Z$ a valori in $]-\infty, +\infty]$.

Descriviamo i singoli passi:

- Si osserva un dato $x \in \Omega$, ove Ω è munito della probabilità P_θ , con θ ignoto;
- Sulla scorta di tale dato, si considera una misura di probabilità ρ_x su (Z, \mathcal{T}) ;
- Si considera un punto $z \in Z$ a caso, con (Z, \mathcal{T}, ρ_x) spazio probabilizzato.

Sotto P_θ , la scelta di z comporta una perdita $W(\theta, z) = W_\theta(z)$. Il costo medio dopo aver osservato $x \in \Omega$ è dunque:

$$W_{\theta,x}^m \stackrel{\text{def}}{=} \int_Z W_\theta(z) \rho_x(dz)$$

Definizione 1.1.1. Si definisce **costo medio**:

$$W_\theta^m \stackrel{\text{def}}{=} \int_\Omega \left[\int_Z W_\theta(z) \rho_x(dz) \right] P_\theta(dx)$$

Definizione 1.1.2. Si definisce **procedura decisionale randomizzata** la funzione ρ , definita in Ω , tale che $\rho(x) = \rho_x$.

Definizione 1.1.3. Data una regola decisionale ρ , si definisce **rischio** della regola decisionale ρ , per la funzione costo W e il valore θ , la speranza del costo medio, sotto P_θ , calcolata in Ω , ossia il costo medio:

$$R(\theta, \rho) \stackrel{\text{def}}{=} \int_\Omega \left[\int_Z W_\theta(z) \rho_x(dz) \right] P_\theta(dx)$$

Scriveremo, in maniera più concisa:

$$R(\theta, \rho) = W_\theta \rho P_\theta$$

Dobbiamo però assicurarci che tutto ciò abbia senso. Supponiamo dunque che:

- Per ogni $\theta \in \Theta$, valga:

$$\inf_{z \in Z} W_\theta(z) > -\infty,$$

per non avere problemi nelle somme di funzioni di costo;

- Fissato $\theta \in \Theta$, al variare di $x \in \Omega$, le probabilità ρ_x siano definite su una σ -algebra \mathcal{D} di Z , tale che la funzione $z \rightarrow W_\theta(z)$ sia misurabile;
- Dato $D \in \mathcal{D}$, la funzione $x \rightarrow \rho_x(D)$ sia \mathcal{A} -misurabile (ove \mathcal{A} è una σ -algebra di Ω , si veda la definizione di esperimento).

Abbiamo dunque una funzione ρ , definita su $\Omega \times \mathcal{D}$, che rispetta le due condizioni dette sopra: una funzione siffatta è detta anche **nucleo di transizione** o **nucleo Markoviano**, dato che funzioni del genere compaiono nella teoria delle catene di Markov.

Osservazione 1. Analizziamo il ruolo della σ -algebra \mathcal{A} : la terza condizione riflette in qualche modo il fatto che la regola decisionale dev'essere formulata con ciò che si ha a disposizione. Non è improbabile che si abbiano poche informazioni rispetto a quelle immaginabili, per mancanza di memoria o per motivi di privacy, ad esempio. Si potrebbe per esempio pensare di non registrare un vettore $\mathbf{X} = (X_1, \dots, X_n)$, ma di registrare solamente, per motivi di privacy o per aumentare la velocità di trasmissione dei dati, la media empirica:

$$\bar{\mathbf{X}} = \frac{1}{n} \cdot \sum_{j=1}^n X_j$$

In tal caso, la σ -algebra potrebbe essere $\mathcal{B}(\mathbb{R}^n)$, ma in realtà è la σ -algebra generata, in \mathbb{R}^n , da $\bar{\mathbf{X}}$.

Osservazione 2. Su alcuni testi si può trovare una definizione di regola decisionale differente (quella **non-randomizzata**).

Essa viene definita come una funzione $\delta : \Omega \rightarrow Z$; la funzione costo è la funzione C , definita in Ω a valori in $[0, +\infty]$, definita da $C_\theta(x) = C(\theta, \delta(x))$ (in particolare, tale

funzione risulta una variabile aleatoria non negativa definita su Ω). Anche la definizione di rischio è differente:

$$R(\theta, \delta) = E^\theta[C(\theta, \delta)] = \int_{\Omega} C(\theta, \delta) P_\theta(dx)$$

Tuttavia, la definizione che abbiamo dato noi (quella randomizzata) conserva una proprietà di convessità. Se α^0 e α^1 sono due regole decisionali randomizzate, allora per ogni $r \in [0, 1]$ la seguente è una regola decisionale randomizzata:

$$\alpha^r \stackrel{\text{def}}{=} (1-r)\alpha^0 + r\alpha^1$$

Tale proprietà di convessità è essenziale per alcuni risultati, pertanto scegliamo la definizione (1.1.2).

Prima di continuare, diamo una definizione fondamentale, che verrà largamente usata in questo capitolo e nei successivi.

Definizione 1.1.4. Si definisce **esperimento statistico** una famiglia di misure di probabilità $\mathcal{E} = \{P_\theta \mid \theta \in \Theta\}$ definite su una σ -algebra \mathcal{A} di sottoinsiemi di un insieme Ω ; l'insieme Θ è spesso detto **spazio dei parametri**.

Consideriamo dunque un esperimento $\mathcal{E} = \{P_\theta \mid \theta \in \Theta\}$ e una funzione di costo W definita su $\Theta \times Z$.

Nel caso di osservazioni effettuate a più riprese, le regole relative al proseguimento o all'interruzione dell'esperimento saranno sempre decise a priori. Possiamo ad esempio considerare una successione di variabili aleatorie reali indipendenti $(X_n)_{n \in \mathbb{N}^+}$, tutte aventi densità uniforme sull'intervallo $[\theta, \theta + 1]$, con $\theta \in \mathbb{R}$. Dato $0 < \varepsilon < 1$, possiamo stabilire che l'esperimento si fermerà non appena varrà:

$$\max_{j \leq n} X_j - \min_{j \leq n} X_j \geq 1 - \varepsilon$$

Fissato $\varepsilon \in [0, 1]$, l'esperimento è in accordo con quanto stabilito prima.

È chiaro che differenti valori di ε danno esperimenti diversi. Data una funzione di costo, tali esperimenti differenti danno funzioni di rischio differenti: è logico pensare che la qualità di una regola decisionale dipenda dalla funzione di rischio che induce. Anticipiamo subito che la scelta della funzione di costo non è indifferente, anzi va eseguita con cura, in base al problema statistico che vogliamo trattare: a tale scelta è correlata la scelta dell'insieme delle decisioni Z .

Esempio. Dato $\mathcal{E} = \{P_\theta \mid \theta \in \Theta\}$, con $\Theta = \mathbb{R}$, con:

$$P_\theta = f_\theta \cdot \lambda,$$

con λ misura di Lebesgue 1-dimensionale, e:

$$f_\theta(x) = \frac{1}{\pi(1+(x-\theta)^2)}$$

funzioni di costo come $W_1(\theta, z) = |z - \theta|$ o $W_2(\theta, z) = |z - \theta|^2$ sono utili per stimare il parametro θ , mentre una funzione di costo del genere:

$$W_3(\theta, z) = \begin{cases} 0, & z\theta \geq 0 \\ 1, & z\theta < 0 \end{cases}$$

è più utile per un test unilatero. In tal caso, è conveniente scegliere un insieme delle decisioni Z formato da due elementi, $+$ e $-$, tali che:

$$\begin{cases} W_3(\theta, +) = 0, & \theta \geq 0 \\ W_3(\theta, +) = 1, & \theta < 0 \\ W_3(\theta, -) = 1, & \theta > 0 \\ W_3(\theta, -) = 0, & \theta \leq 0 \end{cases}$$

Dovrebbe essere ora chiaro, con quest'esempio, che la qualità di una regola decisionale dipende dalla funzione di rischio che induce, una volta fissata la funzione di costo. Dato allora un esperimento \mathcal{E} e una funzione di costo W , definiamo l'insieme seguente $\mathfrak{R}(\mathcal{E}, W)$ dicendo che una funzione $r : \Theta \rightarrow]-\infty, +\infty]$ appartiene a $\mathfrak{R}(\mathcal{E}, W)$ se e solo se, per definizione, esiste una regola decisionale ρ su \mathcal{E} tale che la funzione di rischio soddisfi:

$$W_\theta \rho P_\theta = \int_{\Omega} \left[\int_Z W_\theta(z) \rho_x(dz) \right] P_\theta(dx) \leq r(\theta),$$

per ogni $\theta \in \Theta$.

Osservazione 3. Includere funzioni che maggiorano le funzioni di rischio nell'insieme appena definito serve solo a esprimere determinati risultati in maniera più semplice.

Analizziamo ora qualche proprietà dell'insieme appena definito. Per non cadere nel banale, supporremo Z e Θ non vuoti.

Proposizione 1.1.1. L'insieme $\mathfrak{R}(\mathcal{E}, W)$ non è vuoto.

Proposizione 1.1.2. L'insieme $\mathfrak{R}(\mathcal{E}, W)$ è convesso.

Dimostrazione. Siano $r_1, r_2 \in \mathfrak{R}(\mathcal{E}, W)$, e sia $\alpha \in [0, 1]$. Dobbiamo semplicemente dimostrare che $r_\alpha \in \mathfrak{R}(\mathcal{E}, W)$, ove:

$$r_\alpha(\theta) \stackrel{\text{def}}{=} (1 - \alpha)r_1(\theta) + \alpha r_2(\theta)$$

Per ipotesi, siano ρ_1, ρ_2 regole decisionali tali che, per ogni $\theta \in \Theta$:

- $W_\theta \rho_1 P_\theta \leq r_1(\theta)$;
- $W_\theta \rho_2 P_\theta \leq r_2(\theta)$.

Consideriamo allora la regola decisionale:

$$\rho_\alpha \stackrel{\text{def}}{=} (1 - \alpha)\rho_1 + \alpha\rho_2,$$

Si ottiene allora:

$$\begin{aligned} W_\theta \rho_\alpha P_\theta &= W_\theta [(1 - \alpha)\rho_1 + \alpha\rho_2] P_\theta = (1 - \alpha) W_\theta \rho_1 P_\theta + \alpha W_\theta \rho_2 P_\theta \leq \\ &\leq (1 - \alpha)r_1(\theta) + \alpha r_2(\theta) = r_\alpha(\theta), \end{aligned}$$

da cui la tesi. □

Proposizione 1.1.3. Se $r \in \mathfrak{R}(\mathcal{E}, W)$ e g è una funzione limitata e non negativa, allora $r + g \in \mathfrak{R}(\mathcal{E}, W)$.

Prima di continuare, un paio di considerazioni sulla notazione: indicheremo con $\overline{\mathfrak{R}}(\mathcal{E}, W)$ la chiusura di $\mathfrak{R}(\mathcal{E}, W)$ rispetto alla convergenza puntuale, nello spazio delle funzioni definite in Θ a valori in $[-\infty, +\infty]$. Inoltre indicheremo con π , salvo avvisi contrari, misure di probabilità su Θ aventi supporto finito, ossia tali che esiste un insieme $F \subseteq \Theta$ di cardinalità finita tale che $\pi(\Theta - F) = 0$: esse verranno utilizzate come leggi a priori su Θ (consideriamo questo sotto caso per semplificare le dimostrazioni, in quanto d'ora in avanti Θ avrà quasi sempre cardinalità finita).

Definizione 1.1.5. Dati un esperimento \mathcal{E} (come nella definizione (1.1.4)), una funzione di costo W e una probabilità π a supporto finito, si definisce **rischio bayesiano**:

$$\chi(\pi, \mathcal{E}, W) \stackrel{\text{def}}{=} \inf_{r \in \mathfrak{R}(\mathcal{E}, W)} \left\{ \int_{\Theta} r(\theta) \pi(d\theta) \right\}$$

Il senso di questa definizione è chiaro: dati un esperimento e una funzione di costo, si vuole cercare una procedura decisionale che minimizzi la funzione di rischio indotta.

Teorema 1.1.4. Sia Θ un insieme arbitrario, e sia \mathfrak{R} un insieme di funzioni da Θ in $]-\infty, +\infty]$. Supponiamo che:

- Valgano le ipotesi minime, descritte precedentemente, perchè tutto sia ben definito;
- La chiusura $\overline{\mathfrak{R}}$ di \mathfrak{R} rispetto alla convergenza puntuale, nell'insieme delle funzioni da Θ in $]-\infty, +\infty]$, sia costituita da funzioni a valori in $]-\infty, +\infty]$.

Sia poi $f : \Theta \rightarrow]-\infty, +\infty]$ una funzione. Allora $f \in \overline{\mathfrak{R}}$ se e solo se, per ogni probabilità π con supporto finito su Θ , vale:

$$\chi(\pi) = \inf_{r \in \mathfrak{R}} \left\{ \int_{\Theta} r(\theta) \pi(d\theta) \right\} \leq \int_{\Theta} f(\theta) \pi(d\theta)$$

Dimostrazione. Si ha, per $f \in \overline{\mathfrak{R}}$:

$$\chi(\pi) = \inf_{r \in \mathfrak{R}} \left\{ \int_{\Theta} r(\theta) \pi(d\theta) \right\} = \inf_{r \in \overline{\mathfrak{R}}} \left\{ \int_{\Theta} r(\theta) \pi(d\theta) \right\} \leq \int_{\Theta} f(\theta) \pi(d\theta),$$

da cui la condizione è sicuramente necessaria.

L'altra implicazione usa la compattezza di determinati insiemi, e il fatto che una famiglia di insiemi compatti gode della proprietà dell'intersezione finita: trascuriamo però i dettagli, e concludiamo la dimostrazione. \square

Dal seguente teorema si ottiene, valutando un caso particolare, la seguente osservazione.

Osservazione 4. Supponiamo che $f : \Theta \rightarrow]-\infty, +\infty]$ sia una costante b . Allora, dal teorema (1.1.4), $f \in \overline{\mathfrak{R}}$ se e solo se, per ogni misura π a supporto finito su Θ , vale $\chi(\pi) \leq b$. Si ottiene allora la seguente uguaglianza:

$$\sup_{\pi} \inf_{r \in \mathfrak{R}} \int_{\Theta} r(\theta) \pi(d\theta) = \inf_{r \in \mathfrak{R}} \sup_{\theta \in \Theta} r(\theta)$$

Infatti:

- Per ogni $r \in \overline{\mathfrak{R}}$, $b_r = \sup_{\theta \in \Theta} r(\theta) \in \overline{\mathfrak{R}}$, da cui:

$$\sup_{\pi} \inf_{r \in \overline{\mathfrak{R}}} \int_{\Theta} r(\theta) \pi(d\theta) \leq \inf_{r \in \overline{\mathfrak{R}}} \sup_{\theta \in \Theta} r(\theta) ,$$

o equivalentemente:

$$\sup_{\pi} \inf_{r \in \overline{\mathfrak{R}}} \int_{\Theta} r(\theta) \pi(d\theta) \leq \sup_{\pi} \inf_{r \in \overline{\mathfrak{R}}} \int_{\Theta} \left[\sup_{\theta \in \Theta} r(\theta) \right] \pi(d\theta) = \inf_{r \in \overline{\mathfrak{R}}} \sup_{\theta \in \Theta} r(\theta) ;$$

- Fissata $r \in \overline{\mathfrak{R}}$, e fissato $\varepsilon > 0$, esiste $\theta_0 \in \Theta$ tale che $r(\theta_0) \geq \sup_{\theta \in \Theta} r(\theta) - \varepsilon$, da cui, considerando la misura π_{θ_0} tale che $\pi_{\theta_0}(\theta_0) = 1$, si ha:

$$\int_{\Theta} r(\theta) \pi_{\theta_0}(d\theta) = r(\theta_0) \geq \sup_{\theta \in \Theta} r(\theta) - \varepsilon ,$$

e la tesi si ottiene per l'arbitrarietà di ε .

1.2 Una distanza tra esperimenti

Osservazione 5. Le funzioni di costo, introdotte nella sezione (1.1), risultano a valori in $] - \infty, +\infty]$: ciò può risultare utile in determinati contesti, tuttavia ora supporremo che ogni funzione di costo W sia a valori in $[0, 1]$.

Supponiamo di effettuare un esperimento a scelta (ma non entrambi) tra $\mathcal{E} = \{P_{\theta} \mid \theta \in \Theta\}$ e $\mathcal{F} = \{Q_{\theta} \mid \theta \in \Theta\}$, e supponiamo inoltre che Θ non risenta della nostra scelta sull'esperimento da effettuare. Può accadere che \mathcal{E} sia definito su una σ -algebra \mathcal{A} di un insieme Ω , mentre \mathcal{F} sia definito su una σ -algebra \mathcal{B} di un insieme Ω' , magari diverso da Ω .

Definizione 1.2.1. Si definisce **deficit** di \mathcal{E} rispetto a \mathcal{F} , e si indica con $\delta(\mathcal{E}, \mathcal{F})$, il minimo numero $\delta \in [0, 1]$ tale che, presa una qualsiasi funzione W (tale che $0 \leq W \leq 1$), per ogni $r_2 \in \mathfrak{R}(\mathcal{F}, W)$ esiste $r_1 \in \overline{\mathfrak{R}}(\mathcal{E}, W)$ tale che, per ogni $\theta \in \Theta$:

$$r_1(\theta) \leq r_2(\theta) + \delta$$

Equivalentemente, $\delta(\mathcal{E}, \mathcal{F})$ è il minimo numero $\delta \in [0, 1]$ tale che, per ogni W tale che $0 \leq W \leq 1$:

$$\mathfrak{R}(\mathcal{F}, W) + \delta \subseteq \overline{\mathfrak{R}}(\mathcal{E}, W) ,$$

ove ovviamente:

$$\mathfrak{R}(\mathcal{F}, W) + \delta \stackrel{\text{def}}{=} \{r + \delta \mid r \in \mathfrak{R}(\mathcal{F}, W)\}$$

In effetti, dobbiamo provare che tale minimo esiste: ciò deriva, in sostanza, dal fatto che $\overline{\mathfrak{R}}(\mathcal{E}, W)$ è compatto.

Definizione 1.2.2. Si definisce **distanza** tra gli esperimenti \mathcal{E}, \mathcal{F} :

$$\Delta(\mathcal{E}, \mathcal{F}) \stackrel{\text{def}}{=} (\delta(\mathcal{E}, \mathcal{F}) \vee \delta(\mathcal{F}, \mathcal{E}))$$

Osservazione 6. Si noti che non è richiesto che \mathcal{E} e \mathcal{F} siano famiglie di misure definite sullo stesso spazio, ma solo che essi abbiano un comune spazio dei parametri.

La distanza Δ appena definita non induce una metrica, ma una **pseudometrica**. In particolare:

- Ovviamente vale, per due esperimenti \mathcal{E} e \mathcal{F} , che $\Delta(\mathcal{E}, \mathcal{F}) = \Delta(\mathcal{F}, \mathcal{E})$;
- Dati tre esperimenti \mathcal{E} , \mathcal{F} e \mathcal{G} , vale la disuguaglianza triangolare:

$$\Delta(\mathcal{E}, \mathcal{G}) \leq \Delta(\mathcal{E}, \mathcal{F}) + \Delta(\mathcal{F}, \mathcal{G})$$

Non è invece detto che, se $\Delta(\mathcal{E}, \mathcal{F}) = 0$, allora \mathcal{E} e \mathcal{F} sono lo stesso esperimento.

Definizione 1.2.3. Due esperimenti \mathcal{E}, \mathcal{F} tali che $\Delta(\mathcal{E}, \mathcal{F}) = 0$ si dicono **equivalenti**, o **dello stesso tipo**.

Osservazione 7. Dato Θ , la classe degli esperimenti su Θ è una classe propria, mentre la classe dei tipi di esperimenti (nel senso dato poc'anzi) è un insieme, e la distanza definita in tale insieme quoziente (nel modo ovvio) è una metrica su tale insieme.

Merita un'occhio di riguardo il caso in cui $\delta(\mathcal{E}, \mathcal{F}) = 0$: in tal caso, diremo che \mathcal{F} dà meno informazioni rispetto a \mathcal{E} o, equivalentemente, che \mathcal{E} è migliore di \mathcal{F} , o che dà più informazioni di \mathcal{F} .

Osservazione 8. Il deficit $\delta(\mathcal{E}, \mathcal{F})$ appena definito, sebbene abbia un significato statistico profondo, non sembra facile da calcolare, nella maggior parte dei casi: è utile allora, in certi casi, cercare limitazioni superiori o inferiori.

Esempio. Consideriamo una popolazione di N persone (N molto elevato): vogliamo effettuare un esperimento per stimare la percentuale di persone che passerà il Natale all'estero. Consideriamo allora gli esperimenti $\mathcal{E}_N, \mathcal{F}_N$ (la scelta dell'esperimento, è banale, non influenza tale percentuale), ove:

- L'esperimento \mathcal{E}_N consiste nel selezionare 1000 persone in blocco, in maniera casuale, e intervistarle;
- L'esperimento \mathcal{F}_N consiste nel selezionare casualmente e intervistare una persona, e ripetere l'operazione per 1000 volte.

È palese che nel secondo caso, a differenza del primo, può capitare che una certa persona venga intervistata più volte. È ragionevole pensare (e si può dimostrare) che valga $\delta(\mathcal{E}_N, \mathcal{F}_N) = 0$, mentre il deficit opposto è più difficile da calcolare. Tuttavia, sicuramente esso è correlato alla probabilità di intervistare una persona più volte, da cui si ha che $\lim_{N \rightarrow +\infty} \delta(\mathcal{F}_N, \mathcal{E}_N) = 0$. Gli esperimenti $\mathcal{E}_N, \mathcal{F}_N$ si dicono quindi **asintoticamente equivalenti**.

Esempio. Siano (nell'esempio che segue sarà $\Theta = \mathbb{R}$):

- $\mathcal{E}_n = \{P_{\theta,n} \mid \theta \in \Theta\}$ tale che $P_{\theta,n}$ è la legge di un campione di taglia n e legge di Cauchy, la quale è definita su \mathbb{R} dalla densità (rispetto alla misura di Lebesgue):

$$f_{\theta}(x) = \frac{1}{\pi[1 + (x - \theta)^2]} ;$$

- $\mathcal{F}_n = \{Q_{\theta,n} \mid \theta \in \Theta\}$ tale che $Q_{\theta,n}$ è la legge di un campione di taglia n e legge gaussiana $N(\theta, 2)$.

Si può dimostrare che:

$$\lim_{n \rightarrow +\infty} \Delta(\mathcal{E}_n, \mathcal{F}_n) = 0 ,$$

con velocità dell'ordine di $1/\sqrt{n}$. Anche gli esperimenti $\mathcal{E}_n, \mathcal{F}_n$, pertanto, sono asintoticamente equivalenti: si possono studiare dunque i rischi su \mathcal{E}_n , con un certo grado di approssimazione, analizzando \mathcal{F}_n .

Concludiamo la sezione enunciando un paio di corollari delle definizioni date finora, e del teorema (1.1.4). In questo caso, \mathcal{E} e \mathcal{F} sono due esperimenti arbitrari, aventi un comune insieme dei parametri Θ .

Corollario 1.2.1. Il deficit $\delta(\mathcal{E}, \mathcal{F})$ è il minimo $\delta \in [0, 1]$ tale che per ogni funzione di costo W , con $0 \leq W \leq 1$, si ha:

$$\overline{\mathfrak{R}}(\mathcal{F}, W) + \delta \subseteq \overline{\mathfrak{R}}(\mathcal{E}, W)$$

Corollario 1.2.2. Il deficit $\delta(\mathcal{E}, \mathcal{F})$ è il minimo $\delta \in [0, 1]$ tale che i rischi bayesiani soddisfano la relazione:

$$\chi(\pi, W, \mathcal{E}) \leq \chi(\pi, W, \mathcal{F}) + \delta ,$$

per ogni funzione di costo W con $0 \leq W \leq 1$ e per ogni probabilità π con supporto finito su Θ . Dunque vale la seguente uguaglianza, in termini di rischi bayesiani:

$$\Delta(\mathcal{E}, \mathcal{F}) = \sup_{\pi, W} |\chi(\pi, W, \mathcal{E}) - \chi(\pi, W, \mathcal{F})|$$

1.3 Rappresentazione canonica di Blackwell

D'ora in poi, salvo avvisi contrari, Θ sarà un insieme di cardinalità k : ci concentreremo infatti su esperimenti aventi un insieme dei parametri Θ di cardinalità finita.

Consideriamo dunque un esperimento $\mathcal{E} = \{P_\theta \mid \theta \in \Theta\} = \{P_{\theta_1}, \dots, P_{\theta_k}\}$, ove per ogni $i = 1, \dots, k$ P_{θ_i} è una probabilità su (Ω, \mathcal{A}) . Sia:

$$S = \sum_{i=1}^k P_{\theta_i}$$

Si ha chiaramente che S domina ogni misura P_{θ_i} , da cui, per il noto teorema di Radon-Nikodym, per ogni $i = 1, \dots, k$ esiste $f_i = dP_{\theta_i}/dS$.

Valutando in $x \in \Omega$, si ottiene un vettore di rapporti di verosimiglianza, a valori in \mathbb{R}^k :

$$\mathbf{v}(x) \stackrel{\text{def}}{=} [f_1(x) \quad \dots \quad f_k(x)]^t ,$$

che definisce una funzione $\mathbf{v} : \Omega \rightarrow \mathbb{R}^k$.

Per ogni $x \in \Omega$ valgono le seguenti condizioni:

- Per ogni $i = 1, \dots, k$ vale $f_i(x) \geq 0$;
- $\sum_{i=1}^k f_i(x) = 1$.

Allora, per ogni $x \in \Omega$, vale $\mathbf{v}(x) \in U(\Theta)$, ove:

$$U(\Theta) \stackrel{\text{def}}{=} \left\{ \mathbf{u} \in \mathbb{R}^k \mid u_i \geq 0 \text{ per ogni } i = 1, \dots, k, \sum_{i=1}^k u_i = 1 \right\}$$

è detto **simplexso standard** (di dimensione $(k - 1)$).

Definizione 1.3.1. Si definisce **misura canonica di Blackwell** $m_{\mathcal{E}}$ di \mathcal{E} la misura immagine di S tramite \mathbf{v} (definita, ovviamente, su $U(\Theta)$).

Prima di continuare, premettiamo subito che è più conveniente parlare di ricavi più che di costi: diamo, allora, la seguente definizione.

Definizione 1.3.2. Data una funzione di costo W , con $0 \leq W \leq 1$, si definisce **funzione di ricavo** associata:

$$V \stackrel{\text{def}}{=} 1 - W$$

Vorremmo allora massimizzare, fissata una legge a priori π su Θ , il cosiddetto **guadagno bayesiano**, al variare delle regole decisionali ρ :

$$\Gamma(\pi, \mathcal{E}, V) \stackrel{\text{def}}{=} \sup_{\rho} \int_{\Theta} \int_{\Omega} \int_Z V_{\theta}(z) \rho_x(dz) P_{\theta}(dx) \pi(d\theta) \quad (1.1)$$

Essendo π discreta (visto che Θ è finito), è ben definito, per $i = 1, \dots, k$, il valore $\pi_i = \pi(\theta_i)$. Si ottiene allora la seguente scrittura equivalente, scambiando gli operatori di sommatoria (finita) e integrale, e considerando le densità f_i , per $i = 1, \dots, k$:

$$\begin{aligned} \Gamma(\pi, \mathcal{E}, V) &= \sup_{\rho} \sum_{i=1}^k \pi_i \int_{\Omega} \int_Z V_{\theta_i}(z) \rho_x(dz) P_{\theta_i}(dx) = \\ &= \sup_{\rho} \sum_{i=1}^k \pi_i \int_{\Omega} \left(\int_Z V_{\theta_i}(z) \rho_x(dz) \right) f_i(x) S(dx) = \\ &= \int_{\Omega} \left(\sup_{\rho} \int_Z \left[\sum_{i=1}^k \pi_i V_{\theta_i}(z) f_i(x) \right] \rho_x(dz) \right) S(dx) \end{aligned}$$

Chiaramente, tale quantità verrà massimizzata se per ogni $x \in \Omega$ definiremo una quantità $\mathbf{z}(x)$ tale che:

$$\sup_{z \in Z} \left[\sum_{i=1}^k \pi_i V_{\theta_i}(z) f_i(x) \right] = \sum_{i=1}^k \pi_i V_{\theta_i}(\mathbf{z}(x)) f_i(x)$$

La quantità tra parentesi quadre dipende da x solo attraverso $\mathbf{v}(x)$, essendo $f_i(x) = \mathbf{v}_i(x)$ per ogni $i = 1, \dots, k$. Potremmo dunque massimizzare, una volta fissato $\mathbf{u} \in U(\Theta)$, la quantità:

$$\sum_{i=1}^k \pi_i V_{\theta_i}(z) u_i$$

Definiamo dunque la funzione $\phi : U(\Theta) \rightarrow \mathbb{R}$ ponendo:

$$\phi(\mathbf{u}) = \sup_{z \in Z} \left[\sum_{i=1}^k \pi_i V_{\theta_i}(z) u_i \right]$$

Osservazione 9. Al variare di π e di V , le funzioni $\phi(\mathbf{u})$ formano una classe, che denoteremo con Φ_1 , e che analizzeremo più avanti. Vedremo che le funzioni in Φ_1 sono misurabili e Lipschitziane: ciò avrà diverse conseguenze.

Se in particolare $\mathbf{u} \in U(\Theta)$ è della forma $\mathbf{u} = \mathbf{v}(x)$, con $x \in \Omega$, e in più $\phi(\mathbf{u})$ è ottenuto per $z_0 \in Z$, si definisce allora:

$$\mathbf{z}(x) = (\phi \circ \mathbf{v})(x) \stackrel{\text{def}}{=} z_0$$

Può però capitare che per qualche $\mathbf{u} = \mathbf{v}(x) \in U(\Theta)$, con $x \in \Omega$, non esista un z_0 che realizza il massimo (ossia $\phi(\mathbf{v}(x))$). In tal caso, possiamo comunque avvicinarci all'estremo superiore con un grado di precisione arbitrario, da cui in ogni caso:

$$\sup_{\rho} \int_Z \left[\sum_{i=1}^k \pi_i V_{\theta_i}(z) f_i(x) \right] \rho_x(dz) = (\phi \circ \mathbf{v})(x),$$

Il ricavo bayesiano, inizialmente definito dalla formula (1.1), è dunque:

$$\Gamma(\pi, \mathcal{E}, V) = \int_{\Omega} (\phi \circ \mathbf{v})(x) S(dx),$$

che può essere scritto in maniera equivalente usando la misura canonica di Blackwell $m_{\mathcal{E}}$ nel modo che segue:

$$\Gamma(\pi, \mathcal{E}, V) = \int_{U(\Theta)} \phi(\mathbf{u}) m_{\mathcal{E}}(\mathbf{u}) \quad (1.2)$$

Una funzione come ϕ è una funzione della forma:

$$\phi(\mathbf{u}) = \sup_{z \in Z} \left[\sum_{i=1}^k c_i(z) u_i \right],$$

ove i coefficienti $c_i(z)$, per $i = 1, \dots, k$, sono tali che:

- Per ogni $i = 1, \dots, k$, vale $c_i(z) \geq 0$ per ogni $z \in Z$;
- $\sum_{i=1}^k \sup_{z \in Z} [c_i(z)] \leq 1$.

Ciò è sostanzialmente dovuto al fatto che la funzione di ricavo V è compresa tra 0 e 1.

Viceversa, una funzione della forma:

$$\alpha(\mathbf{u}) = \sup_{z \in Z} \left[\sum_{i=1}^k c_i(z) u_i \right],$$

ove i coefficienti soddisfano le due condizioni esposte sopra, corrisponde a un certo problema statistico, con una certa misura π su Θ e una certa funzione di ricavo V . Definendo infatti:

$$s \stackrel{\text{def}}{=} \sum_{i=1}^k \sup_{z \in Z} [c_i(z)],$$

si hanno due casi:

- Vale $s = 0$. In questo caso la situazione è banale, e in particolare $V \equiv 0$;
- Vale $s > 0$ (il caso interessante). In tal caso definiamo, per $i = 1, \dots, k$:

$$\pi_i = \frac{1}{s} \sup_{z \in Z} [c_i(z)] ,$$

quindi definiamo V :

$$V_{\theta_i}(z) = \begin{cases} c_i(z)/\pi_i , & \pi_i > 0 \\ 0 , & \pi_i = 0 \end{cases}$$

Cominciamo ora ad esporre qualche lemma preparatorio. Innanzitutto, definiamo su $U(\Theta)$ la seguente distanza:

$$\|\mathbf{u} - \mathbf{u}'\| \stackrel{\text{def}}{=} \max_{i=1, \dots, k} |u_i - u'_i| \quad (1.3)$$

Lemma 1.3.1. La classe Φ_1 è un insieme convesso di funzioni convesse su $U(\Theta)$.

Se $\phi_1, \phi_2 \in \Phi_1$, allora:

$$\frac{1}{2}(\phi_1 \vee \phi_2) \in \Phi_1$$

Inoltre ogni $\phi \in \Phi_1$ è 1-Lipschitziana:

$$|\phi(\mathbf{u}) - \phi(\mathbf{u}')| \leq \|\mathbf{u} - \mathbf{u}'\|$$

Dimostrazione. Sia $\phi \in \Phi_1$. Allora, per ogni $\mathbf{u} \in U(\Theta)$:

$$\phi(\mathbf{u}) = \sup_{z \in Z} \left[\sum_{i=1}^k c_i(z) u_i \right] ,$$

dove:

- Per ogni $i = 1, \dots, k$, vale $c_i(z) \geq 0$ per ogni $z \in Z$;
- Vale:

$$\sum_{i=1}^k \sup_{z \in Z} [c_i(z)] \leq 1$$

Allora, dati $\mathbf{u}, \mathbf{u}' \in U(\Theta)$, se consideriamo la funzione così definita:

$$\xi(\mathbf{u}) = \sum_{i=1}^k c_i(z) u_i ,$$

allora

$$\begin{aligned} |\xi(\mathbf{u}) - \xi(\mathbf{u}')| &= \left| \sum_{i=1}^k c_i(z) [u_i - u'_i] \right| \leq \\ &\leq \sum_{i=1}^k c_i(z) |u_i - u'_i| \leq \sum_{i=1}^k c_i(z) \|\mathbf{u} - \mathbf{u}'\| \leq \|\mathbf{u} - \mathbf{u}'\| , \end{aligned}$$

dunque ξ è 1-Lipschitziana. Tale condizione viene conservata se consideriamo l'estremo superiore, puntualmente su Z : concludiamo che $\phi \in \Phi_1$ è 1-Lipschitziana.

Dimostrare a questo punto che Φ_1 è convesso, che una funzione $\phi \in \Phi_1$ è convessa, e che date $\phi_1, \phi_2 \in \Phi_1$, allora $\frac{1}{2}(\phi_1 \vee \phi_2) \in \Phi_1$ consta di verifiche molto semplici: trascuriamo perciò i dettagli. \square

Osservazione 10. Il fatto che $\phi \in \Phi_1$ è 1-Lipschitziana rende ϕ misurabile, da cui $\phi \circ \mathbf{v}$ è sempre misurabile.

Abbiamo già visto che, per un esperimento \mathcal{E} , con Θ finito (dunque $\mathcal{E} = \{P_i \mid i = 1, \dots, k\}$), il ricavo bayesiano si scrive nella forma vista nella formula (1.2), usando la misura di Blackwell $m_{\mathcal{E}}$ di \mathcal{E} .

Definizione 1.3.3. Una qualsiasi misura m su $U(\Theta)$ tale che, per $i = 1, \dots, k$:

$$\int_{U(\Theta)} \mathbf{u}_i dm(\mathbf{u}) = 1$$

definisce un esperimento, considerando le misure seguenti, per $i = 1, \dots, k$:

$$Q_i(d\mathbf{u}) = u_i m(d\mathbf{u})$$

In particolare, $m_{\mathcal{E}}$ definisce un esperimento $\mathcal{F} = \{Q_i \mid i = 1, \dots, k\}$, con:

$$Q_i(d\mathbf{u}) = u_i m_{\mathcal{E}}(d\mathbf{u}),$$

detto **rappresentazione canonica di Blackwell** di \mathcal{E} .

Essa ha alcune interessanti applicazioni, e consente di caratterizzare in maniera alternativa la distanza tra due esperimenti \mathcal{E} e \mathcal{F} , come dimostra il seguente lemma, che enunciamo solamente.

Lemma 1.3.2. La distanza tra due esperimenti \mathcal{E} e \mathcal{G} , aventi lo stesso spazio dei parametri Θ , di cardinalità finita, è:

$$\Delta(\mathcal{E}, \mathcal{G}) = \sup_{\phi \in \Phi_1} \left| \int_{U(\Theta)} \phi(\mathbf{u}) [m_{\mathcal{E}}(d\mathbf{u}) - m_{\mathcal{G}}(d\mathbf{u})] \right|$$

In particolare, se due esperimenti hanno la stessa misura di Blackwell, allora essi sono equivalenti, e per ogni funzione di costo W , si ha:

$$\mathfrak{R}(\mathcal{E}, W) = \mathfrak{R}(\mathcal{G}, W)$$

Ciò vale, in particolare, per un esperimento \mathcal{E} e per la sua rappresentazione canonica \mathcal{F} .

Vediamo ora qualche interessante disuguaglianza verificata dalla distanza Δ . Ricordiamo che la distanza che abbiamo definito su $U(\Theta)$ è quella definita dalla formula (1.3): essa può essere usata per definire una distanza tra misure portate da $U(\Theta)$, come nel caso della cosiddetta norma dual-Lipschitziana.

Definizione 1.3.4. Se μ è una misura finita con segno portata da $U(\Theta)$, si definisce **norma dual-Lipschitziana** di μ :

$$\|\mu\|_D \stackrel{\text{def}}{=} \sup_f \left| \int_{U(\Theta)} f d\mu \right|,$$

ove l'estremo superiore è calcolato tra le funzioni f 1-Lipschitziane e tali che $\|f\|_{\infty} \leq 1$.

Tale norma definisce, in modo naturale, una distanza tra due misure μ_1 e μ_2 , data da $\|\mu_1 - \mu_2\|_D$.

Osservazione 11. È opportuno con confondersi con la norma L^1 di una misura:

$$\|\mu\|_1 = \sup_f \left| \int_{U(\Theta)} f \, d\mu \right|,$$

ove l'estremo superiore è calcolato tra le funzioni f misurabili con $\|f\|_\infty \leq 1$. È evidente che valga $\|\mu\|_D \leq \|\mu\|_1$, ma non è detto che valga l'uguaglianza.

Consideriamo ora due esperimenti \mathcal{E} e \mathcal{F} e le loro misure di Blackwell $m_{\mathcal{E}}$ e $m_{\mathcal{F}}$. Dato che ogni $\phi \in \Phi_1$ risulta 1-Lipschitziana, si ha banalmente:

$$\Delta(\mathcal{E}, \mathcal{F}) \leq \|m_{\mathcal{E}} - m_{\mathcal{F}}\|_D$$

Per trovare una migliore relazione tra i due membri di questa disuguaglianza, vediamo ora due lemmi.

Lemma 1.3.3. Una misura canonica di Blackwell è univocamente determinata dai suoi valori su Φ_1 .

Dimostrazione. Sia Φ l'insieme delle funzioni del tipo $a \cdot \phi$, con $\phi \in \Phi_1$, e $a \geq 0$ reale; sia poi \mathcal{H} lo spazio delle differenze del tipo $\phi_1 - \phi_2$, con $\phi_1, \phi_2 \in \Phi$.

Discende dal lemma (1.3.1) che Φ è convesso, e inoltre, se $\phi_1, \phi_2 \in \Phi$, allora $(\phi_1 \vee \phi_2) \in \Phi$; risulta inoltre che \mathcal{H} è un reticolo vettoriale. Una misura di Blackwell m definisce pertanto un funzionale lineare e positivo μ su Φ_1 , quindi su Φ e su \mathcal{H} , per linearità.

Il funzionale μ è σ -liscio (a tal proposito si legga l'osservazione (12)). Infatti le funzioni in \mathcal{H} sono Lipschitziane, dunque continue. Usando il Lemma del Dini, allora, se $(\phi_n)_{n \in \mathbb{N}^+} \searrow 0$ puntualmente su $U(\Theta)$, allora $(\phi_n)_{n \in \mathbb{N}^+} \searrow 0$ uniformemente: essendo $\|\mu\|_D < +\infty$, esso risulta allora σ -liscio.

Possiamo allora estendere μ , con una procedura dovuta a P.J.Daniell (a tal proposito si legga l'osservazione (13)), allo spazio delle funzioni integrabili rispetto a tale misura (nel senso di Daniell).

Notiamo che le funzioni costanti appartengono a \mathcal{H} , dato che:

$$\sum_{i=1}^k u_i = 1$$

Allora tutte le funzioni limitate e misurabili rispetto alla σ -algebra generata dalle funzioni in \mathcal{H} sono integrabili nel senso di Daniell. L'estensione di Daniell, dunque, esiste: essa è unica, per le proprietà di linearità, positività, e σ -lisciezza (non entriamo nei dettagli). \square

Osservazione 12. Un funzionale si dice σ -liscio se, presa una successione $(\phi_n)_{n \in \mathbb{N}^+} \in \mathcal{H}$ tale che $(\phi_n)_{n \in \mathbb{N}^+} \searrow 0$ puntualmente su $U(\Theta)$, allora:

$$[\mu(\phi_n)]_{n \in \mathbb{N}^+} \searrow 0$$

Osservazione 13. Analizziamo ora la procedura di Daniell per estendere μ .

Iniziamo considerando un reticolo vettoriale \mathcal{H} . Definiamo la **classe superiore** $U_{\mathcal{H}}$ di funzioni, dicendo che $g \in U_{\mathcal{H}}$ se e solo se esiste $(\phi_n)_{n \in \mathbb{N}^+} \in \mathcal{H}$ tale che $g = \sup_{n \in \mathbb{N}^+} \phi_n$.

Definiamo poi, in maniera duale, la **classe inferiore** $L_{\mathcal{H}}$, dicendo che $g \in L_{\mathcal{H}}$ se e solo se $-g \in U_{\mathcal{H}}$.

Estendiamo allora μ a $U_{\mathcal{H}}$ definendo, per $g \in U_{\mathcal{H}}$:

$$\bar{\mu}(g) \stackrel{\text{def}}{=} \sup_{\substack{h \in \mathcal{H} \\ h \leq g}} \mu(h),$$

e a questo punto, per $f \in L_{\mathcal{H}}$ definiamo:

$$\underline{\mu}(f) \stackrel{\text{def}}{=} -\bar{\mu}(-f)$$

Chiamiamo allora **integrabile** nel senso di Daniell una funzione β tale che per ogni $\varepsilon > 0$ esistono due funzioni $f \in L_{\mathcal{H}}$ e $g \in U_{\mathcal{H}}$, con $f \leq \beta \leq g$, tali che:

- $\bar{\mu}(g) < +\infty$;
- $\bar{\mu}(g) - \underline{\mu}(f) < \varepsilon$.

Il valore $\tilde{\mu}(\beta) \in [\underline{\mu}(f), \bar{\mu}(g)]$ è dunque univocamente determinato, e si ottiene pertanto l'estensione $\tilde{\mu}$ voluta. Il fatto che tale estensione sia σ -liscia, lineare e positiva non è immediato, perciò tralasciamo i dettagli.

Osservazione 14. Le misure di Blackwell risultano σ -finite anche se le probabilità P_{θ} non lo sono: tale situazione si presenta quando si ha a che fare con processi stocastici.

Lemma 1.3.4. Le misure di Blackwell formano un insieme compatto, rispetto alla metrica indotta dalla norma dual-Lipschitziana $\|\cdot\|_D$.

Dimostrazione. Sia $\varepsilon > 0$, e sia $M \in]0, +\infty[$. Chiamiamo Λ_M l'insieme delle funzioni f su $U(\Theta)$ che risultano M -Lipschitziane, e tali che $\|f\|_{\infty} \leq M$. Dato che $U(\Theta)$ è totalmente limitato, possiamo determinare un insieme $F_M \subseteq U(\Theta)$ di cardinalità finita (sia $F_M = \{\mathbf{x}_1, \dots, \mathbf{x}_n\}$) tale che, per $\mathbf{u} \in U(\Theta)$, sia:

$$d(\mathbf{u}, F_M) \leq \frac{\varepsilon}{M^2}$$

È chiaro ora che, se conosciamo $f \in \Lambda_M$ su F_M , allora conosciamo f su tutto $U(\Theta)$, a meno di un errore uniformemente minore di $\frac{\varepsilon}{M^2} \cdot M = \frac{\varepsilon}{M}$.

Possiamo dunque approssimare Λ_M con una sottofamiglia finita $\Xi_M \subseteq \Lambda_M$, a meno di un errore minore di $\frac{\varepsilon}{M}$, in norma uniforme. Possiamo supporre, ora, che le funzioni coordinate appartengano a Ξ_M .

Una qualsiasi misura di Blackwell può essere valutata su Ξ_M : in particolare, per una funzione coordinata u_i , vale $\mu(u_i) = 1$. Al variare di μ , fissata $f \in \Xi_M$, si ottiene allora un insieme compatto di \mathbb{R} : data allora una successione di misure di Blackwell $(\mu_n)_{n \in \mathbb{N}^+}$, si può estrarre una sottosuccessione $(\mu_{n_k})_{k \in \mathbb{N}^+}$ convergente, nel senso che esiste un valore reale $a(f)$ tale che:

$$\lim_{k \rightarrow +\infty} \mu_{n_k}(f) = \lim_{k \rightarrow +\infty} \int_{U(\Theta)} f d\mu_{n_k} = a(f)$$

Essendo $|\Xi_M| < +\infty$, il processo di estrazione può essere ripetuto, e in un numero finito di passi si ottiene una successione $(\mu_h)_{h \in \mathbb{N}^+}$ tale che:

$$\lim_{h \rightarrow +\infty} \int_{U(\Theta)} f d\mu_h = a(f)$$

per ogni $f \in \Xi_M$. Con un procedimento diagonale (non diamo i dettagli), data $(\mu_n)_{n \in \mathbb{N}^+}$, si può estrarre una sottosuccessione $(\mu_{n_k})_{k \in \mathbb{N}^+}$ tale che, per ogni $M \in \mathbb{N}^+$ e ogni $f \in \Xi_M$, esiste $a(f) \in \mathbb{R}$ tale che:

$$\lim_{k \rightarrow +\infty} \int_{U(\Theta)} f d\mu_{n_k} = a(f)$$

Tale successione di valori converge anche nel caso in cui $f \in \Lambda$ è solamente Lipschitziana, dato che l'insieme numerabile seguente:

$$D \stackrel{\text{def}}{=} \bigcup_{M \in \mathbb{N}^+} \Xi_M$$

è denso nell'insieme Λ delle funzioni Lipschitziane: $\overline{D} = \Lambda$.

Osserviamo che, per una funzione coordinata $u_i \in \Xi_M$, con M qualsiasi, vale banalmente $a(u_i) = 1$.

Essendo a (ossia il funzionale che mappa f in $a(f)$) un limite di funzionali lineari e positivi, esso è lineare e positivo. Per il lemma del Dini, allora, a è σ -liscio su Λ , quindi della forma:

$$a(f) = \int_{U(\Theta)} f d\mu ,$$

con μ misura di Blackwell opportuna.

Resta da dimostrare che la convergenza sia uniforme su Λ_1 : di sicuro lo è sull'insieme finito Ξ_1 . In più:

$$\limsup_{k \rightarrow +\infty} \|\mu_{n_k} - \mu\|_D \leq \varepsilon ,$$

e la tesi segue dall'arbitrarietà di $\varepsilon > 0$. □

Siamo ora pronti per esporre il principale risultato di questo capitolo.

Teorema 1.3.5. Sia Θ finito, di cardinalità k . Esiste allora una funzione ω_k , definita su $[0, 1]$ e infinitesima per $t \rightarrow 0$, tale che:

$$\Delta(\mathcal{E}, \mathcal{F}) \leq \|m_{\mathcal{E}} - m_{\mathcal{F}}\|_D \leq \omega_k(\Delta(\mathcal{E}, \mathcal{F})) ,$$

per ogni coppia di esperimenti indicizzati in Θ .

Dimostrazione. Abbiamo già visto che vale:

$$\Delta(\mathcal{E}, \mathcal{F}) = \sup_{\phi \in \Phi_1} \left| \int_{U(\Theta)} \phi dm_{\mathcal{E}} - \int_{U(\Theta)} \phi dm_{\mathcal{F}} \right| \leq \|m_{\mathcal{E}} - m_{\mathcal{F}}\|_D$$

Definiamo allora:

$$\omega_k(t) \stackrel{\text{def}}{=} \sup_{\substack{\mathcal{E}, \mathcal{F} \\ \Delta(\mathcal{E}, \mathcal{F}) \leq t}} (\|m_{\mathcal{E}} - m_{\mathcal{F}}\|_D)$$

È banale notare che ω_k è una funzione crescente. Supponiamo per assurdo che esista $\varepsilon > 0$ tale che, per $t > 0$, vale $\omega_k(t) \geq 2\varepsilon$.

Esiste allora una successione di coppie di esperimenti $(\mathcal{E}_n, \mathcal{F}_n)_{n \in \mathbb{N}^+}$ tale che, per ogni $n \in \mathbb{N}^+$:

- $\Delta(\mathcal{E}, \mathcal{F}) \leq \frac{1}{n}$;
- $\|m_{\mathcal{E}_n} - m_{\mathcal{F}_n}\|_D \geq \varepsilon$.

Per il lemma (1.3.4), esistono μ_1, μ_2 misure di Blackwell tali che:

$$\lim_{n \rightarrow +\infty} \|m_{\mathcal{E}_n} - \mu_1\|_D = 0, \quad \lim_{n \rightarrow +\infty} \|m_{\mathcal{F}_n} - \mu_2\|_D = 0$$

Deve però valere anche $\|\mu_1 - \mu_2\|_D \geq \varepsilon$, per quanto ipotizzato.

Dall'ipotesi $\Delta(\mathcal{E}, \mathcal{F}) \leq \frac{1}{n}$, per $n \in \mathbb{N}^+$, si ha che μ_1 e μ_2 corrispondono allo stesso esperimento: per il lemma (1.3.3), allora, $\mu_1 = \mu_2$, e ciò è assurdo. \square

Corollario 1.3.6. Fissato Θ di cardinalità finita, gli esperimenti indicizzati in Θ formano un insieme compatto rispetto alla distanza Δ .

1.4 Un'osservazione sulla convergenza di misure

Sia ancora Θ finito, con $|\Theta| = k$. Dato un esperimento $\mathcal{E} = \{P_\theta \mid \theta \in \Theta\}$, per definizione la misura di Blackwell $m_{\mathcal{E}}$ è la misura immagine su $U(\Theta)$ ottenuta mediante \mathbf{v} a partire da:

$$S = \sum_{\theta \in \Theta} P_\theta$$

Si ha $\mathbf{v}(x) = [f_1(x), \dots, f_k(x)]^t$, con $f_i = dP_{\theta_i}/dS$. Nella trasformazione, la densità f_i diventa la funzione coordinata u_i , da cui, sotto P_{θ_i} , la distribuzione di \mathbf{v} diventa la misura $u_i \cdot m_{\mathcal{E}}$.

In molti problemi statistici, potrebbe essere utile avere altri rapporti di verosimiglianza. Ad esempio, fissando un particolare $s \in \Theta$, vorremmo considerare le verosimiglianze della forma dP_θ/dP_s , al variare di θ in Θ . Come sappiamo dal teorema di Radon-Nikodym, la densità dP_θ/dP_s non è ovunque definita. Chiameremo allora dP_θ/dP_s la densità di P_θ rispetto a P_s moltiplicata per l'indicatrice dell'insieme dove P_s domina P_θ .

Sul semplice $U(\Theta)$ tale densità diventa uguale al rapporto u_θ/u_s , definito ove $u_s > 0$: ciò deriva direttamente dalle relazioni $dP_\theta = f_\theta \cdot dS$ e $dP_s = f_s \cdot dS$.

Restringiamoci ora al caso di misure di probabilità definite su spazi finito-dimensionali (come \mathbb{R}^k): in questo caso ci sono diverse definizioni di convergenza, tutte topologicamente equivalenti alla cosiddetta **convergenza stretta**, ossia la convergenza degli integrali delle funzioni continue e limitate.

Si può ad esempio dare la seguente definizione equivalente. Data una successione di probabilità $(P_n)_{n \in \mathbb{N}^+}$ e una probabilità P_0 , si ha che la successione converge vagamente, o debolmente, o strettamente, a P_0 , se, per ogni funzione g continua e limitata:

$$\lim_{n \rightarrow +\infty} \int_{\mathbb{R}^k} g dP_n = \int_{\mathbb{R}^k} g dP_0$$

Si può dimostrare (non lo faremo) che tale convergenza è equivalente alla convergenza della distanza dual-Lipschitziana a 0:

$$\lim_{n \rightarrow +\infty} \|P_n - P_0\|_D = 0$$

Osservazione 15. Se $(X_n)_{n \in \mathbb{N}^+}$ è una successione di variabili aleatorie ad esempio reali, con $X_n \sim P_n$ per ogni $n \in \mathbb{N}^+$, e X_0 è una variabile aleatoria reale di legge P_0 , allora è facile convincersi che $(X_n)_{n \in \mathbb{N}^+}$ converge in legge a X_0 se e solo se $\|P_n - P_0\|_D$ tende a 0.

Introduciamo sin da ora una distanza che ci servirà nei prossimi capitoli. Dato lo spazio metrico $(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$ munito della distanza euclidea d , definiamo, per un insieme boreliano $A \subseteq \mathbb{R}^k$ e per $\varepsilon \in [0, 1]$ (nella definizione che vedremo, risulta chiaro che non ha senso considerare valori di ε maggiori di 1):

$$A^\varepsilon \stackrel{\text{def}}{=} \bigcup_{x \in A} \overline{B}(x, \varepsilon),$$

ossia:

$$A^\varepsilon \stackrel{\text{def}}{=} \{x \in \mathbb{R}^k \mid d(x, A) \leq \varepsilon\}$$

Definizione 1.4.1. Date due misure μ e ν , si definisce **distanza di Lévy-Prohorov** tra μ e ν la seguente quantità:

$$\pi(\mu, \nu) \stackrel{\text{def}}{=} \inf \{ \varepsilon \in [0, 1] \mid \mu(B) \leq \nu(B^\varepsilon) + \varepsilon \text{ per ogni } A \in \mathcal{B}(\mathbb{R}^k) \}$$

Osservazione 16. Essendo $(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$ uno spazio separabile e completo (se munito della distanza euclidea d), si può dimostrare che lo spazio delle probabilità su tale spazio è separabile e completo.

Si può dimostrare che la convergenza in norma dual-Lipschitziana è equivalente alla convergenza secondo la metrica di Lévy-Prohorov:

$$\lim_{n \rightarrow +\infty} \pi(P_n, P_0) = 0$$

Ci sono altre definizioni equivalenti di convergenza, una delle quali coinvolge, ad esempio, le funzioni di ripartizione: questa è una delle definizioni più vecchie e meno flessibili a disposizione.

Tali diverse definizioni risultano più o meno comode a seconda dei casi, e risultano equivalenti sotto le opportune ipotesi minime, per poter definire le varie distanze (ad esempio, per definire la distanza di Lévy-Prohorov è necessario che lo spazio sia munito di una distanza).

Lemma 1.4.1. Sia $(\mathcal{E}_n)_{n \in \mathbb{N}^+}$ una successione di esperimenti, ove, per $n \in \mathbb{N}^+$, si ha:

$$\mathcal{E}_n = \{P_{\theta, n} \mid \theta \in \Theta\},$$

ove Θ è un comune spazio dei parametri, di cardinalità finita k .

Sia poi \mathcal{F} un esperimento, che chiameremo “esperimento limite”:

$$\mathcal{F} = \{Q_\theta \mid \theta \in \Theta\}$$

Sono fatti equivalenti:

1. Le misure di Blackwell $(m_{\mathcal{E}_n})_{n \in \mathbb{N}^+}$ convergono a $m_{\mathcal{F}}$ nel senso ordinario, ossia quello della distanza dual-Lipschitziana;
2. La successione di variabili aleatorie vettoriali $(\mathbf{v}_n)_{n \in \mathbb{N}^+}$, ove:

$$\mathbf{v}_n \stackrel{\text{def}}{=} [dP_{1,n}/dS_n \ \dots \ dP_{k,n}/dS_n]^t ,$$

ove per ogni $n \in \mathbb{N}^+$ \mathbf{v}_n è una variabile aleatoria vettoriale definita su $(\Omega_n, \mathcal{A}_n, S_n)$, converge in legge al vettore:

$$\mathbf{w} \stackrel{\text{def}}{=} [dQ_1/dS' \ \dots \ dQ_k/dS']^t ,$$

variabile aleatoria vettoriale definita su $(\Omega, \mathcal{A}, S')$;

3. Fissato $s \in \Theta$, la successione di variabili aleatorie vettoriali $(\mathbf{v}_n^s)_{n \in \mathbb{N}^+}$, ove:

$$\mathbf{v}_n^s \stackrel{\text{def}}{=} [dP_{1,n}/dP_{s,n} \ \dots \ dP_{k,n}/dP_{s,n}]^t ,$$

ove per ogni $n \in \mathbb{N}^+$ \mathbf{v}_n^s è una variabile aleatoria vettoriale definita su $(\Omega_n, \mathcal{A}_n, P_{s,n})$, converge in legge al vettore:

$$\mathbf{w}^s \stackrel{\text{def}}{=} [dQ_1/dQ_s \ \dots \ dQ_k/dQ_s]^t ,$$

variabile aleatoria vettoriale definita su $(\Omega, \mathcal{A}, Q_s)$.

Questo lemma risulterà molto utile nei prossimi capitoli, in particolare nell'ultimo: è ora chiaro che per studiare i limiti di successioni di esperimenti, conviene (se Θ ha cardinalità finita) passare all'analisi dei rapporti di verosimiglianza.

Capitolo 2

Contiguità e Trasformata di Hellinger

Nel capitolo 1 abbiamo visto come, nel caso in cui Θ è finito, la convergenza di esperimenti nel senso della distanza dual-Lipschitziana è equivalente alla convergenza in distribuzione dei rapporti di verosimiglianza.

In questo capitolo introdurremo la nozione di **contiguità**, che è in sostanza una forma asintotica dell'assoluta continuità: vedremo varie caratterizzazioni della contiguità, attraverso un teorema.

Più avanti, poi, introdurremo la **trasformata di Hellinger**, utile per studiare osservazioni indipendenti, in analogia alle funzioni caratteristiche nel caso di somme di variabili aleatorie indipendenti.

2.1 Contiguità

La nozione di contiguità, introdotta da L.Le Cam e J.Esary nel 1956, è stata largamente usata negli anni seguenti, per le sue numerose applicazioni, da diversi autori, che si sono cimentati in particolar modo nel dimostrare la contiguità di determinate successioni di probabilità sotto varie ipotesi.

Supponiamo che sia $\Theta = \{0, 1\}$. Consideriamo una successione $((X_n, \mathcal{A}_n))_{n \in \mathbb{N}^+}$ di spazi misurabili (muniti ciascuno di una σ -algebra), e sia $(\mathcal{E}_n)_{n \in \mathbb{N}^+}$ una successione di esperimenti binari, ove per $n \in \mathbb{N}^+$ $\mathcal{E}_n = \{P_{0,n}, P_{1,n}\}$ è definito su (X_n, \mathcal{A}_n) .

Introdurremo ora una particolare trasformazione, che ci permetterà di considerare, per ogni $n \in \mathbb{N}^+$, l'intervallo $[0, 1]$ invece di X_n .

Sia:

$$S_n \stackrel{\text{def}}{=} P_{0,n} + P_{1,n},$$

e sia, per $i = 0, 1$, $f_{i,n} = dP_{i,n}/dS_n$. Tali densità sono ben definite S_n -quasi ovunque, in quanto per ogni $n \in \mathbb{N}^+$ si ha $P_{0,n} \ll S_n$, $P_{1,n} \ll S_n$.

L'immagine di S_n mediante $f_{1,n}$ è una misura μ_n su $[0, 1]$, la quale è la somma delle probabilità immagine mediante $f_{1,n}$ di $P_{0,n}$ e $P_{1,n}$.

È evidente che:

- L'immagine di $P_{1,n}$ mediante $f_{1,n}$ ha densità u_n rispetto a μ_n , con $0 \leq u_n \leq 1$, e vale chiaramente:

$$\int_0^1 u_n d\mu_n = \int_{X_n} dP_{1,n} = \int_{X_n} f_{1,n} dS_n = 1 ;$$

- L'immagine di $P_{0,n}$ mediante $f_{1,n}$ ha densità $1-u_n$ rispetto a μ_n , e vale chiaramente:

$$\int_0^1 1 - u_n d\mu_n = \int_{X_n} dP_{0,n} = \int_{X_n} f_{0,n} dS_n = 1 .$$

In statistica, soprattutto quando si ha a che fare con osservazioni indipendenti, si usa passare ai logaritmi dei rapporti di verosimiglianza.

Consideriamo la seguente funzione $\lambda : [0, 1] \rightarrow \tilde{\mathbb{R}}$ definita da:

$$\lambda(x) = \log \left(\frac{x}{1-x} \right),$$

ove $\lambda(0) = -\infty$, $\lambda(1) = +\infty$. Tale funzione è strettamente crescente.

La misura immagine $\lambda(\mu_n)$ è una misura M_n su $\tilde{\mathbb{R}}$. Essa è la distribuzione di $\log(f_{1,n}/f_{0,n})$ sotto S_n . Si hanno dunque le seguenti densità:

$$p_0(z) = \frac{1}{1 - e^z}, \quad p_1(z) = \frac{e^z}{1 - e^z},$$

con $z \in \tilde{\mathbb{R}}$, e le usuali convenzioni per $z = \pm\infty$. La funzione $\log(u/(1-u))$ può infatti assumere i valori $-\infty$ e $+\infty$, se definita in $[0, 1]$.

Operando il pull-back del rapporto $u/(1-u)$, si ottiene, sullo spazio (X_n, \mathcal{A}_n) , il rapporto $f_{1,n}/f_{0,n}$: questo può essere considerata una definizione, per il rapporto di verosimiglianza di $P_{1,n}$ su $P_{0,n}$, e il suo logaritmo può assumere i valori $-\infty$ e $+\infty$. Questo è un inconveniente, quando si vanno a sommare i logaritmi dei rapporti di verosimiglianza relativi a osservazioni indipendenti. Non si hanno problemi se invece per ogni $n \in \mathbb{N}^+$ chiamiamo $P'_{1,n} = f'_{1,n} \cdot P_{0,n}$ la parte dominata da $P_{0,n}$ di $P_{1,n}$ e definiamo, sul supporto di $P_{0,n}$, la seguente funzione, a valori in $[-\infty, +\infty[$:

$$\Lambda_n \stackrel{\text{def}}{=} \log \left(\frac{dP'_{1,n}}{dP_{0,n}} \right) \tag{2.1}$$

D'ora in poi scriveremo informalmente $\Lambda_n = \log(dP_{1,n}/dP_{0,n})$, ma tale rapporto sarà sempre inteso nel senso dato dalla formula (2.1).

Prima di continuare, diamo un paio di definizioni.

Definizione 2.1.1. Si chiama **statistica** disponibile sull'esperimento \mathcal{E}_n una qualsiasi funzione misurabile T_n da (X_n, \mathcal{A}_n) in \mathbb{R} .

Definizione 2.1.2. Data una successione $(T_n)_{n \in \mathbb{N}^+}$, si dice che essa è **convergente** a 0 in $P_{0,n}$ -probabilità se:

$$\lim_{n \rightarrow +\infty} \int_{X_n} (1 \wedge |T_n|) dP_{0,n} = 0$$

Osservazione 17. Con le notazioni date sopra, indicheremo la legge di T_n sotto $P_{0,n}$ con $L(T_n|P_{0,n})$. Questa è una probabilità definita su $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

Ricordando il teorema di Prohorov, per misure definite su $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, diciamo che la successione di probabilità $(L(T_n|P_{0,n}))_{n \in \mathbb{N}^+}$ è relativamente compatta rispetto alla convergenza stretta se e solo se è tesa, ossia se per ogni $\varepsilon > 0$ esiste un compatto della forma $[-M_\varepsilon, M_\varepsilon]$ tale che per ogni $n \in \mathbb{N}^+$ vale:

$$L(T_n|P_{0,n})([-M_\varepsilon, M_\varepsilon]^c) = P_{0,n}(|T_n| > M_\varepsilon) < \varepsilon$$

In tali ipotesi, possiamo estrarre una sottosuccessione convergente ad un limite, che è una probabilità su \mathbb{R} : tale limite è detto **punto di accumulazione**.

Esponiamo ora il teorema più importante di questa sezione.

Teorema 2.1.1. Sia $(M_n)_{n \in \mathbb{N}^+}$ l'immagine su $\tilde{\mathbb{R}}$ della successione $(S_n)_{n \in \mathbb{N}^+}$ (ove per ogni $n \in \mathbb{N}^+$ vale $S_n = P_{0,n} + P_{1,n}$), ottenuta nel modo descritto precedentemente. Sia poi:

$$F_n(z) = p_0(z)M_n(z)$$

la legge del logaritmo del rapporto di verosimiglianza sotto $P_{0,n}$, e sia $(T_n)_{n \in \mathbb{N}^+}$ una successione di statistiche, ove per ogni $n \in \mathbb{N}^+$ T_n risulta una statistica disponibile sull'esperimento binario $\mathcal{E}_n = \{P_{0,n}, P_{1,n}\}$.

Le seguenti sei proprietà sono tutte equivalenti:

1. Se $(T_n)_{n \in \mathbb{N}^+} \rightarrow 0$ in $P_{0,n}$ -probabilità, allora $(T_n)_{n \in \mathbb{N}^+} \rightarrow 0$ in $P_{1,n}$ -probabilità;
2. Se $(T_n)_{n \in \mathbb{N}^+}$ è tale che $(L(T_n|P_{0,n}))_{n \in \mathbb{N}^+}$ è relativamente compatta (ossia è tesa), allora essa è tale che $(L(T_n|P_{1,n}))_{n \in \mathbb{N}^+}$ è relativamente compatta (ossia tesa);
3. Se M è un punto d'accumulazione di $(M_n)_{n \in \mathbb{N}^+}$, allora $M(\{+\infty\}) = 0$;
4. Se F è un punto di accumulazione di $(F_n)_{n \in \mathbb{N}^+}$, allora:

$$\int_{\tilde{\mathbb{R}}} e^z F(dz) = 1 ;$$

5. Se $(z_n)_{n \in \mathbb{N}^+} \subseteq [-\infty, +\infty[$ è tale che $(z_n)_{n \in \mathbb{N}^+} \nearrow +\infty$, allora:

$$(M_n([z_n, +\infty]))_{n \in \mathbb{N}^+} \rightarrow 0 ;$$

6. Per ogni $\varepsilon > 0$ esistono $N_\varepsilon \in \mathbb{N}^+$ e $b_\varepsilon \in \mathbb{R}$ tali che per ogni $n \geq N_\varepsilon$ vale:

$$M_n([b_\varepsilon, +\infty]) < \varepsilon .$$

Dimostrazione. Iniziamo dimostrando che (5.) è equivalente a (6.).

Supponiamo che non valga (6.). Per definizione, allora, esiste $\varepsilon_0 > 0$ tale che per ogni $b \in \mathbb{R}$ frequentemente vale:

$$M_n([b, +\infty]) \geq \varepsilon_0$$

Data una qualsiasi successione $(b_k)_{k \in \mathbb{N}^+} \subseteq [-\infty, +\infty[$, tale che $(b_k)_{k \in \mathbb{N}^+} \nearrow +\infty$, esiste una successione strettamente crescente $(n_k)_{k \in \mathbb{N}^+} \subseteq \mathbb{N}^+$ tale che, per ogni $k \in \mathbb{N}^+$, vale

$M_{n_k}([b_k, +\infty]) \geq \varepsilon_0$. Considerando allora la successione $(\beta_j)_{j \in \mathbb{N}^+}$ ottenuta in questo modo:

$$\begin{cases} \beta_j = b_1, & j \leq n_1 \\ \beta_j = b_k, & n_{k-1} < j \leq n_k, \quad k \geq 2 \end{cases}$$

si ha $(\beta_j)_{j \in \mathbb{N}^+} \nearrow +\infty$, ma $M_j([\beta_j, +\infty]) \geq \varepsilon_0$ frequentemente, da cui non vale (5.). Per contronominale, allora, (5.) \Rightarrow (6.). L'implicazione opposta è davvero banale, perciò la omettiamo.

Dimostriamo ora che (3.) è equivalente a (4.).

Iniziamo notando che $p_0(z) = 1/(1 + e^z)$ è una funzione invertibile, continua e limitata su $\tilde{\mathbb{R}}$ (entrambi i fatti sono ovvi). Se dunque M è punto di accumulazione di $(M_n)_{n \in \mathbb{N}^+}$, allora esiste una sottosuccessione $(M_{n_j})_{j \in \mathbb{N}^+}$ tale che $(M_{n_j})_{j \in \mathbb{N}^+} \rightarrow M$ strettamente.

Dato che $\tilde{\mathbb{R}}$ è compatto, ciò vuol dire che, per ogni funzione continua ϕ :

$$\lim_{n \rightarrow +\infty} \int_{\tilde{\mathbb{R}}} \phi(z) dM_{n_j}(z) = \int_{\tilde{\mathbb{R}}} \phi(z) dM(z)$$

Allora vale:

$$\begin{aligned} \lim_{n \rightarrow +\infty} \int_{\tilde{\mathbb{R}}} \phi(z) p_0(z) dM_{n_j}(z) &= \int_{\tilde{\mathbb{R}}} \phi(z) p_0(z) dM(z) \\ \lim_{n \rightarrow +\infty} \int_{\tilde{\mathbb{R}}} \phi(z) dF_{n_j}(z) &= \int_{\tilde{\mathbb{R}}} \phi(z) dF(z), \end{aligned}$$

da cui F , tale che $F = p_0 \cdot M$, è punto di accumulazione per $(F_n)_{n \in \mathbb{N}^+}$.

Sfruttando poi l'invertibilità di p_0 , se F è punto di accumulazione per $(F_n)_{n \in \mathbb{N}^+}$, allora esiste M tale che $F = p_0 \cdot M$, e risulta che M è punto di accumulazione per $(M_n)_{n \in \mathbb{N}^+}$.

Ricordando ora che $p_1(z) = e^z/(1 + e^z) = e^z \cdot p_0(z)$, si ha:

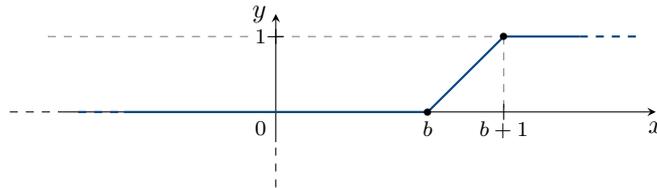
$$\int_{\tilde{\mathbb{R}}} e^z dF(z) = \int_{[-\infty, +\infty[} e^z dF(z) = \int_{[-\infty, +\infty[} p_1(z) dM(z) = 1 - M(\{+\infty\}),$$

dato che $+\infty$ ha massa nulla sotto F , visto che $p_0(\{+\infty\}) = 0$.

L'integrale vale dunque 1 se e solo se $M(\{+\infty\}) = 0$, da cui l'equivalenza delle due proprietà.

Dimostriamo ora l'equivalenza tra (3.) e (6.).

Premettiamo un'osservazione. Sia M un punto d'accumulazione per $(M_n)_{n \in \mathbb{N}^+}$, e sia $(M_{n_j})_{j \in \mathbb{N}^+} \rightarrow M$ una sottosuccessione convergente strettamente. Sia poi $b \in \mathbb{R}$. Si consideri allora la funzione ϕ_b , che vale 0 fino a b e 1 oltre $b + 1$, disegnata nel seguente grafico:



Si ha che ϕ_b è continua e limitata, da cui:

$$\lim_{j \rightarrow +\infty} \int_{\tilde{\mathbb{R}}} \phi_b(z) dM_{n_j}(z) = \int_{\tilde{\mathbb{R}}} \phi_b(z) dM(z)$$

Se (6.) vale, allora, fissato $\varepsilon > 0$, esiste b reale tale che:

$$\lim_{j \rightarrow +\infty} \int_{\tilde{\mathbb{R}}} \phi_b(z) dM_{n_j}(z) = \int_{\tilde{\mathbb{R}}} \phi_b(z) dM(z) \leq \varepsilon$$

Dunque:

$$\lim_{b \rightarrow +\infty} \int_{\tilde{\mathbb{R}}} \phi_b(z) dM(z) = M(\{+\infty\}) \leq \varepsilon,$$

da cui si ha $M(\{+\infty\}) = 0$ per l'arbitrarietà di ε , da cui si ottiene (3.).

Se invece (6.) non vale, allora esiste $\varepsilon_0 > 0$ tale che per ogni b vale l'opposto di quanto scritto sopra. Allora:

$$\int_{\tilde{\mathbb{R}}} \phi_b(z) dM(z) \geq \varepsilon_0$$

Dunque, contrariamente a prima, si ha:

$$\lim_{b \rightarrow +\infty} \int_{\tilde{\mathbb{R}}} \phi_b(z) dM(z) = M(\{+\infty\}) \geq \varepsilon_0,$$

da cui non sussiste (3.).

Dimostriamo ora che (6.) \Rightarrow (1.).

Supponiamo che $(T_n)_{n \in \mathbb{N}^+}$ tenda a 0 in $P_{0,n}$ -probabilità:

$$\lim_{n \rightarrow +\infty} \int_{\tilde{\mathbb{R}}} (1 \wedge |T_n|) dP_{0,n} = 0$$

Sia $\Lambda_n = \log(dP'_{1,n}/dP_{0,n})$ (ove $P'_{1,n}$ è la parte dominata da $P_{0,n}$ di $P_{1,n}$): scriviamo, per semplicità, $\Lambda_n = \log(dP_{1,n}/dP_{0,n})$.

Sia, per l'esattezza, $P_{1,n} = P'_{1,n} + P''_{0,n}$, ove $P'_{1,n}$ è dominata da $P_{0,n}$, e $P''_{0,n}$ è $P_{0,n}$ -singolare. Queste due probabilità sono portate rispettivamente da A_n e B_n , e vale $X_n = A_n \cup B_n$. Possiamo chiedere che $A_n \cap B_n$ sia vuoto, così da ottenere una partizione di X_n .

Sia allora $\Lambda_n^* \stackrel{\text{def}}{=} \Lambda_n$ su A_n , e $\Lambda_n^* \stackrel{\text{def}}{=} +\infty$ su B_n . Con questa notazione, per $b \in \mathbb{R}$:

- $M_n([b, +\infty]) = P_{0,n}(\Lambda_n^* \geq b) + P_{1,n}(\Lambda_n^* \geq b) = P_{0,n}(\Lambda_n \geq b) + P_{1,n}(\Lambda_n^* \geq b)$;
- $$\begin{aligned} \int_{\tilde{\mathbb{R}}} (1 \wedge |T_n|) dP_{1,n} &= \int_{\tilde{\mathbb{R}}} (1 \wedge |T_n|) \mathbb{I}(\Lambda_n^* < b) dP_{1,n} + \int_{\tilde{\mathbb{R}}} (1 \wedge |T_n|) \mathbb{I}(\Lambda_n^* \geq b) dP_{1,n} \leq \\ &\leq \int_{\tilde{\mathbb{R}}} (1 \wedge |T_n|) \mathbb{I}(\Lambda_n < b) dP_{1,n} + P_{1,n}(\Lambda_n^* \geq b) \leq \\ &\leq \int_{\tilde{\mathbb{R}}} (1 \wedge |T_n|) \mathbb{I}(\Lambda_n < b) e^{\Lambda_n} dP_{0,n} + M_n([b, +\infty]) \leq \\ &\leq e^b \int_{\tilde{\mathbb{R}}} (1 \wedge |T_n|) dP_{0,n} + M_n([b, +\infty]). \end{aligned}$$

Al tendere di n a $+\infty$, l'integrale al secondo membro tende a 0 per ipotesi, mentre $M_n([b, +\infty])$ tende a 0 perchè vale (6.), per b abbastanza elevato. Dunque $(T_n)_{n \in \mathbb{N}^+}$ tende a 0 in $P_{1,n}$ -probabilità (ossia vale (1.)).

Dimostriamo ora che (1.) \Rightarrow (5.).

Sia $(z_n)_{n \in \mathbb{N}^+} \subseteq [-\infty, +\infty[$ tale che $(z_n)_{n \in \mathbb{N}^+} \rightarrow +\infty$. Sia poi, per ogni $n \in \mathbb{N}^+$, $T_n = \mathbb{I}([z_n, +\infty])$. Allora:

$$\int_{\mathbb{R}} T_n(z) dP_{0,n}(z) = \int_{z_n}^{+\infty} \frac{1}{1+e^z} dM_n(z) \leq \frac{1}{1+e^{z_n}} \int_{\mathbb{R}} dM_n(z) = \frac{2}{1+e^{z_n}},$$

da cui:

$$\lim_{n \rightarrow +\infty} \int_{\mathbb{R}} T_n(z) dP_{0,n}(z) = 0$$

Dato che vale (1.):

$$\lim_{n \rightarrow +\infty} \int_{\mathbb{R}} T_n(z) dP_{1,n}(z) = \lim_{n \rightarrow +\infty} \int_{z_n}^{+\infty} \frac{e^z}{1+e^z} dM_n(z) = 0,$$

e dato che, per $z \geq 0$, vale $e^z/(1+e^z) \leq 2$, si ha:

$$\lim_{n \rightarrow +\infty} M_n([z_n, +\infty]) = 0,$$

ossia (5.).

Per concludere, dimostriamo che (2.) è equivalente, ad esempio, a (6.).

Supponiamo che valga (6.), e sia $(T_n)_{n \in \mathbb{N}^+}$ tale che $(L(T_n|P_{0,n}))_{n \in \mathbb{N}^+}$ risulti relativamente compatta, ossia tesa. Allora, fissato $\varepsilon > 0$, esiste M_ε tale che, per ogni $n \in \mathbb{N}^+$, vale $P_{0,n}(|T_n| > M_\varepsilon) < \varepsilon$. Fissato $b \in \mathbb{R}$ arbitrario, si ha:

$$P_{1,n}(|T_n| > M_\varepsilon) \leq e^b P_{0,n}(|T_n| > M_\varepsilon) + M_n([b, +\infty]),$$

e dato che:

- Per (6.) è possibile scegliere selezionare b tale che $M_n([b, +\infty]) < \varepsilon/2$ per ogni $n \in \mathbb{N}^+$;
- Per ipotesi, dato che $(L(T_n|P_{0,n}))_{n \in \mathbb{N}^+}$ è tesa, è possibile selezionare M_ε tale che per ogni $n \in \mathbb{N}^+$ valga $P_{0,n}(|T_n| > M_\varepsilon) < \varepsilon/2 \cdot e^{-b}$,

si ha che $(L(T_n|P_{1,n}))_{n \in \mathbb{N}^+}$ è tesa, ossia (2.).

Dimostriamo ora l'implicazione opposta. Definendo $T_n = dP'_{1,n}/dP_{0,n}$ su A_n , $T_n = n$ su B_n (si riveda l'implicazione (6.) \Rightarrow (1.) se questa definizione non è chiara), dato b reale si ha, per $n \in \mathbb{N}^+$:

$$M_n([b, +\infty]) = P_{0,n}(T_n \geq b) + P_{1,n}(T_n \geq b),$$

da cui, se vale (1.), per concludere è sufficiente dimostrare che $(L(T_n|P_{0,n}))_{n \in \mathbb{N}^+}$ è relativamente compatta, ossia tesa. Ciò è conseguenza immediata della disuguaglianza di Markov, in quanto:

$$P_{0,n}(T_n \geq b) \leq \frac{\mathbb{E}_n^0[T_n]}{b} = \frac{1}{b}$$

Si ottiene dunque la proprietà (6.), e ciò conclude la dimostrazione. \square

Osservazione 18. Un esperimento binario $\mathcal{E}_n = \{P_{0,n}, P_{1,n}\}$ è equivalente alla sua rappresentazione di Blackwell, equivalente all'esperimento binario $\mathcal{F}_n = \{F_{0,n}, F_{1,n}\}$, definito su $[-\infty, +\infty]$, con $F_{i,n} = p_i \cdot M_n$. La convergenza di $(\mathcal{E}_n)_{n \in \mathbb{N}^+}$ è dunque equivalente alla convergenza delle successioni di probabilità (definite su $[-\infty, +\infty]$) $(F_{0,n})_{n \in \mathbb{N}^+}$ e $(F_{1,n})_{n \in \mathbb{N}^+}$ rispettivamente a F_0 e F_1 (due misure sempre definite su $[-\infty, +\infty]$).

Teorema 2.1.2. Nelle ipotesi del teorema (2.1.1), le sei proprietà sono equivalenti alla seguente:

7. Se $\mathcal{E} = \{P_0, P_1\}$ è un punto d'accumulazione per una successione $(\mathcal{E}_n)_{n \in \mathbb{N}^+}$, allora $P_1 \ll P_0$.

Dimostrazione. Dimostriamo ad esempio che (7.) è equivalente a (3.).

Abbiamo visto, nel capitolo 1, che se $\mathcal{E} = \{P_0, P_1\}$ e $\mathcal{F} = \{F_0, F_1\}$ sono esperimenti equivalenti, allora $P_1 \ll P_0$ se e solo se $F_1 \ll F_0$. E la condizione $M(\{+\infty\}) = 0$, in effetti, è equivalente alla condizione $F_1 \ll F_0$, dato che la densità di F_0 rispetto a M si annulla solamente in $+\infty$. Ciò conclude la dimostrazione. \square

Definizione 2.1.3. Se la successione di esperimenti binari $(\mathcal{E}_n)_{n \in \mathbb{N}^+}$, ove $\mathcal{E}_n = \{P_{0,n}, P_{1,n}\}$ per ogni $n \in \mathbb{N}^+$, soddisfa una delle sette condizioni viste in questi due teoremi, si dice che la successione $(P_{1,n})_{n \in \mathbb{N}^+}$ è **contigua** alla successione $(P_{0,n})_{n \in \mathbb{N}^+}$.

Notiamo che la relazione di contiguità non è a priori riflessiva.

Definizione 2.1.4. Se accade che $(P_{1,n})_{n \in \mathbb{N}^+}$ e $(P_{0,n})_{n \in \mathbb{N}^+}$ sono ognuna contigua all'altra, in accordo con la definizione (2.1.3), esse si dicono **contigue**.

Osservazione 19. Di solito, nei casi pratici, si ha che la successione $(L(\Lambda_n | P_{0,n}))_{n \in \mathbb{N}^+}$ ha come limite su $\tilde{\mathbb{R}}$ una distribuzione portata da \mathbb{R} . Ciò vuol dire che il limite M della proprietà (3.) è tale che $M(\{-\infty\}) = 0$, oltre a $M(\{+\infty\}) = 0$.

Se dunque, se $(L(\Lambda_n | P_{0,n}))_{n \in \mathbb{N}^+}$ ha un limite portato da \mathbb{R} , allora $(P_{1,n})_{n \in \mathbb{N}^+}$ è contigua a $(P_{0,n})_{n \in \mathbb{N}^+}$ se e solo se $(P_{0,n})_{n \in \mathbb{N}^+}$ è contigua a $(P_{1,n})_{n \in \mathbb{N}^+}$.

Nella proposizione che segue, sarà $\Lambda_n = \log(dP'_{1,n}/dP_{0,n})$, secondo la formula (2.1).

Proposizione 2.1.3. Sia $(T_n)_{n \in \mathbb{N}^+}$ una successione di variabili aleatorie reali, ove per ogni $n \in \mathbb{N}^+$ T_n risulta anche una statistica disponibile sull'esperimento $\mathcal{E}_n = \{P_{0,n}, P_{1,n}\}$. Assumiamo che:

- $(P_{1,n})_{n \in \mathbb{N}^+}$ sia contigua a $(P_{0,n})_{n \in \mathbb{N}^+}$;
- La successione delle leggi congiunte $(L((T_n, \Lambda_n) | P_{0,n}))_{n \in \mathbb{N}^+}$ tenda a un limite F su $\mathbb{R} \times \tilde{\mathbb{R}}$.

Allora $(L((T_n, \Lambda_n) | P_{1,n}))_{n \in \mathbb{N}^+}$ tende a un limite G su $\mathbb{R} \times \tilde{\mathbb{R}}$, e inoltre:

$$G(dt, d\lambda) = e^\lambda \cdot F(dt, d\lambda)$$

Dimostrazione. Sia ϕ una funzione continua e limitata, definita in $\mathbb{R} \times \tilde{\mathbb{R}}$, nelle due variabili (t, λ) . Sia poi, per $i = 0, 1$, $F_{i,n} = L((T_n, \Lambda_n) | P_{i,n})$.

Allora, detto $\Omega = \mathbb{R} \times \tilde{\mathbb{R}}$, vale:

$$\int_{\Omega} \phi(t, \lambda) F_{1,n}(dt, d\lambda) = \int_{\Omega} \phi(t, \lambda) e^{\lambda} F_{0,n}(dt, d\lambda) + \eta_n ,$$

ove la successione $(\eta_n)_{n \in \mathbb{N}^+}$ dei contributi della parte singolare di $P_{1,n}$ rispetto a $P_{0,n}$, tende a 0 per l'ipotesi di contiguità.

L'integrale al secondo membro, inoltre, si può scrivere nel seguente modo:

$$\begin{aligned} & \int_{\Omega} \phi(t, \lambda) e^{\lambda} F_{0,n}(dt, d\lambda) = \\ & = \int_{\Omega} \phi(t, \lambda) (b \wedge e^{\lambda}) F_{0,n}(dt, d\lambda) + \int_{\Omega} \phi(t, \lambda) [e^{\lambda} - b]^+ F_{0,n}(dt, d\lambda) , \end{aligned}$$

dove $[e^{\lambda} - b]^+$ è uguale a $(e^{\lambda} - b)$ se tale valore è positivo, 0 altrimenti.

Fissato $\varepsilon > 0$, dunque:

- Per b sufficientemente grande, il secondo integrale al secondo membro è minore di ε , in valore assoluto;
- Il primo integrale tende a:

$$\int_{\Omega} \phi(t, \lambda) (b \wedge e^{\lambda}) F(dt, d\lambda) ,$$

che tende per $b \rightarrow +\infty$ a:

$$\int_{\Omega} \phi(t, \lambda) e^{\lambda} F(dt, d\lambda) .$$

Si ottiene quindi:

$$\lim_{n \rightarrow +\infty} L((T_n, \Lambda_n) | P_{1,n}) = G ,$$

con $G(dt, d\lambda) = e^{\lambda} \cdot F(dt, d\lambda)$ (ossia la tesi). □

Facciamo ora un paio di esempi, che riprenderemo anche nel prossimo capitolo.

Esempio. Supponiamo che $L(\Lambda_n | P_{0,n})$ converga a una gaussiana:

$$F = N(m, \sigma^2)$$

Allora $L(\Lambda_n | P_{1,n})$ converge a G , tale che:

$$G(d\lambda) = e^{\lambda} F(d\lambda)$$

Dunque, dato che G dev'essere una misura di probabilità:

$$\int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}\sigma} e^{x - \frac{(x-m)^2}{2\sigma^2}} dx = \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}\sigma} e^{x - \frac{(x-m)^2}{2\sigma^2}} dx = 1 ,$$

e dato che:

$$x - \frac{(x-m)^2}{2\sigma^2} = \frac{-x^2 + 2(m+\sigma^2)x - m^2}{2\sigma^2} = -\frac{(x - (m + \sigma^2))^2}{2\sigma^2} + \frac{(m + \sigma^2)^2 - m^2}{2\sigma^2} ,$$

bisogna imporre:

$$\frac{(m + \sigma^2)^2 - m^2}{2\sigma^2} = 0 \Rightarrow m^2 + 2m\sigma^2 + \sigma^4 - m^2 = 2m\sigma^2 + \sigma^4 = 0,$$

e supponendo $\sigma > 0$:

$$m = -\frac{\sigma^2}{2}$$

Con questa condizione, $G = N(m + \sigma^2, \sigma^2)$.

Esempio. Supponiamo che $L((T_n, \Lambda_n)|P_{0,n})$ converga a un vettore gaussiano bidimensionale $F = N_2(\mathbf{m}, \Gamma)$, ove il vettore delle medie \mathbf{m} e la matrice delle covarianze Γ sono:

$$\mathbf{m} = \begin{bmatrix} a \\ b \end{bmatrix}, \quad \Gamma = \begin{bmatrix} A & C \\ C & B \end{bmatrix}$$

Dato un vettore gaussiano $\mathbf{X} \sim F$, e dato un vettore bidimensionale \mathbf{u} , allora, si ha che il logaritmo della funzione caratteristica di \mathbf{X} , valutato in \mathbf{u} , vale:

$$\begin{aligned} \psi_0(\mathbf{u}) &\stackrel{\text{def}}{=} \log \left(\mathbb{E} \left[e^{i(\mathbf{u} \cdot \mathbf{X})} \right] \right) = \log \left(\mathbb{E} \left[e^{i(u_1 T + u_2 \Lambda)} \right] \right) = \\ &= i(\mathbf{m} \cdot \mathbf{u}) - \frac{1}{2}(\Gamma \mathbf{u} \cdot \mathbf{u}) = iu_1 a + iu_2 b - \frac{1}{2} [Au_1^2 + 2Cu_1 u_2 + Bu_2^2] \end{aligned}$$

Se dunque $G = e^\Lambda F$, si ha sostanzialmente, dato $\mathbf{Y} \sim G$:

$$\begin{aligned} \psi_1(\mathbf{u}) &\stackrel{\text{def}}{=} \log \left(\mathbb{E} \left[e^{i(\mathbf{u} \cdot \mathbf{Y})} \right] \right) = \log \left(\mathbb{E} \left[e^{i(u_1 T + u_2 \Lambda) + \Lambda} \right] \right) = \log \left(\mathbb{E} \left[e^{i(u_1 T + (u_2 - i)\Lambda)} \right] \right) = \\ &= iu_1 a + i(u_2 - i)b - \frac{1}{2} [Au_1^2 + 2Cu_1(u_2 - i) + B(u_2 - i)^2] = \\ &= iu_1 a + iu_2 b + b + iu_1 C + iu_2 B + \frac{B}{2} - \frac{1}{2} [Au_1^2 + 2Cu_1 u_2 + Bu_2^2], \end{aligned}$$

da cui bisogna imporre $b = -B/2$. Sotto tale condizione $G = N_2(\mathbf{m}_1, \Gamma)$, ove il vettore delle medie, questa volta, è:

$$\mathbf{m}_1 = \begin{bmatrix} a + C \\ b + B \end{bmatrix} = \begin{bmatrix} a + C \\ B/2 \end{bmatrix}$$

Osservazione 20. Abbiamo visto, nel lemma (1.4.1), un criterio per la convergenza di una successione di esperimenti: tramite la proposizione (2.1.3), possiamo semplificare tale criterio.

Se infatti esiste un particolare $s \in \Theta$ tale che tutte le successioni $(P_{\theta,n})_{n \in \mathbb{N}^+}$, al variare di $\theta \in \Theta$ (ricordiamo che Θ , in questo caso, è finito), sono contigue alla successione $(P_{s,n})_{n \in \mathbb{N}^+}$, allora è sufficiente controllare che $L(\mathbf{v}_n^s | P_{s,n})$ converge (si riveda la sezione (1.3) per la definizione del vettore \mathbf{v}_n^s).

Basta infatti considerare le coppie (T_n, Λ_n) , al variare di $n \in \mathbb{N}^+$, ove, per $n \in \mathbb{N}^+$:

$$T_n = \mathbf{v}_n^s, \quad \Lambda_n = \log(\mathbf{v}_n^s)$$

e applicare la proposizione (2.1.3).

2.2 Distanza di Hellinger

Nell'ambito delle misure finite con segno, due norme che abbiamo usato prima d'ora sono:

- La norma 1:

$$\|\mu\|_1 = \sup_f \left| \int_{\Omega} f \, d\mu \right|,$$

con f misurabile e tale che $|f| \leq 1$ in Ω ;

- La norma dual-Lipschitziana:

$$\|\mu\|_D = \sup_f \left| \int_{\Omega} f \, d\mu \right|,$$

con f misurabile, tale che $|f| \leq 1$ in Ω , e tale che $\|f(x) - f(y)\| \leq d(x, y)$, ove d è una distanza definita su Ω .

Sappiamo che a priori vale $\|\mu\|_D \leq \|\mu\|_1$.

Nessuna di queste due distanze si comporta bene quando si effettuano prodotti diretti di spazi (operazione che si compie quando si studiano osservazioni indipendenti). In questi casi, c'è una distanza che si comporta molto meglio.

Definizione 2.2.1. Dato uno spazio misurabile (Ω, \mathcal{A}) , e due misure positive P, Q su di esso, si definisce **distanza di Hellinger** $h(P, Q)$ tra P e Q :

$$h(P, Q) \stackrel{\text{def}}{=} \sqrt{\frac{1}{2} \int_{\Omega} (\sqrt{dP} - \sqrt{dQ})^2}$$

Per comodità, consideriamo il suo quadrato:

$$h^2(P, Q) = \frac{1}{2} \int_{\Omega} (\sqrt{dP} - \sqrt{dQ})^2 \quad (2.2)$$

Definizione 2.2.2. Se P, Q sono probabilità, si definisce **affinità** $\rho(P, Q)$ tra P e Q :

$$\rho(P, Q) \stackrel{\text{def}}{=} \int_{\Omega} \sqrt{dP} \, dQ$$

Con tale definizione, si ha, direttamente dalla formula (2.2):

$$h^2(P, Q) = 1 - \rho(P, Q) \quad (2.3)$$

Se esiste una misura μ che domina sia P che Q , allora dette $f = dP/d\mu$ e $g = dQ/d\mu$, possiamo scrivere l'affinità in maniera più familiare:

$$\rho(P, Q) = \int_{\Omega} \sqrt{f} \, g \, d\mu$$

Chiaramente $\rho(P, Q)$ non varia al variare della misura dominante μ' considerata; stesso discorso per $h^2(P, Q)$.

Proposizione 2.2.1. Date due probabilità P, Q su (Ω, \mathcal{A}) , si ha la seguente catena di disuguaglianze:

$$h^2(P, Q) \leq \frac{1}{2} \|P - Q\|_1 \leq h(P, Q) \sqrt{2 - h^2(P, Q)} = \sqrt{1 - \rho^2(P, Q)}$$

Dimostrazione. Sia $\mu = P + Q$, e siano $f = dP/d\mu$, $g = dQ/d\mu$. Allora, dato che $f \geq 0$, $g \geq 0$, si ha:

$$\begin{cases} \sqrt{f} - \sqrt{g} \leq \sqrt{f} + \sqrt{g} \\ \sqrt{g} - \sqrt{f} \leq \sqrt{g} + \sqrt{f} \end{cases} \Rightarrow |\sqrt{f} - \sqrt{g}| \leq \sqrt{f} + \sqrt{g},$$

da cui:

$$\begin{aligned} h^2(P, Q) &= \frac{1}{2} \int_{\Omega} (\sqrt{f} - \sqrt{g})^2 d\mu \leq \\ &\leq \frac{1}{2} \int_{\Omega} |\sqrt{f} - \sqrt{g}| |\sqrt{f} + \sqrt{g}| d\mu \leq \frac{1}{2} \int_{\Omega} |f - g| d\mu = \frac{1}{2} \|P - Q\|_1 \end{aligned}$$

Si ha inoltre, sfruttando la disuguaglianza di Cauchy-Schwarz:

$$\begin{aligned} \int_{\Omega} |\sqrt{f} - \sqrt{g}| |\sqrt{f} + \sqrt{g}| d\mu &\leq \sqrt{\int_{\Omega} (\sqrt{f} - \sqrt{g})^2 d\mu} \sqrt{\int_{\Omega} (\sqrt{f} + \sqrt{g})^2 d\mu} = \\ &= \sqrt{(1 - \rho(P, Q))(1 + \rho(P, Q))} = h(P, Q) \sqrt{2 - h^2(P, Q)}, \end{aligned}$$

ossia la tesi. \square

Osservazione 21. Si ha che la somma degli errori per un test tra P e Q che minimizza tale somma è:

$$\|P \wedge Q\|_1 = 1 - \frac{1}{2} \|P - Q\|_1,$$

da cui la quantità $\frac{1}{2} \|P - Q\|_1$ ha un significato statistico non banale. Per la proposizione (2.2.1), allora, si può analizzare $h(P, Q)$ invece di $\frac{1}{2} \|P - Q\|_1$.

Vediamo ora come $h(P, Q)$ e $\rho(P, Q)$ si comportano bene con i prodotti diretti.

Sia J un insieme di indici. Sia $(\mathcal{E}_j)_{j \in J}$ un insieme di osservazioni indipendenti, indicizzato in J , ove per ogni $j \in J$ \mathcal{E}_j è un esperimento binario definito sullo spazio misurabile $(\Omega_j, \mathcal{A}_j)$, che consta di due probabilità $P_{0,j}, P_{1,j}$. Considerato allora il seguente spazio misurabile:

$$(\Omega, \mathcal{A}) = \left(\prod_{j \in J} \Omega_j, \bigotimes_{j \in J} \mathcal{A}_j \right),$$

consideriamo l'esperimento binario $\mathcal{E} = \{\mathbf{P}_0, \mathbf{P}_1\}$, ove:

$$\mathbf{P}_0 = \bigotimes_{j \in J} P_{0,j}, \quad \mathbf{P}_1 = \bigotimes_{j \in J} P_{1,j}$$

Direttamente dalla definizione (2.2.2), allora, si ha:

$$\begin{aligned} \rho(\mathbf{P}_0, \mathbf{P}_1) &= \int_{\Omega} \sqrt{d\mathbf{P}_0 d\mathbf{P}_1} = \int_{\Omega} \sqrt{d \bigotimes_{j \in J} P_{0,j} d \bigotimes_{j \in J} P_{1,j}} = \\ &= \prod_{j \in J} \int_{\Omega_j} \sqrt{dP_{0,j} dP_{1,j}} = \prod_{j \in J} \rho(P_{0,j}, P_{1,j}) \end{aligned}$$

Vale allora, se $\mathbf{h} = h(\mathbf{P}_0, \mathbf{P}_1)$, e per $j \in J$ $h_j = h(P_{0,j}, P_{1,j})$, usando la formula (2.3):

$$1 - \mathbf{h}^2 = \prod_{j \in J} (1 - h_j^2) \leq e^{-\sum_{j \in J} h_j^2},$$

ove l'ultima disuguaglianza è abbastanza nota.

Cosa ci dice tale catena di disuguaglianze? Se consideriamo una successione di insiemi di esperimenti indipendenti $(\mathcal{E}_n)_{n \in \mathbb{N}^+}$, ove per ogni $n \in \mathbb{N}^+$ vale:

$$\mathbf{P}_{0,n} = \bigotimes_{j \in J} P_{0,j,n}, \quad \mathbf{P}_{1,n} = \bigotimes_{j \in J} P_{1,j,n},$$

allora:

- Sostenere che $\prod_{j \in J} (1 - h_{j,n}^2)$ resta distante da 0, per $n \rightarrow +\infty$, è equivalente a dire che $\mathbf{P}_{0,n}$ e $\mathbf{P}_{1,n}$ non sono mai completamente separate, ossia la somma degli errori di un eventuale test di $\mathbf{P}_{0,n}$ contro $\mathbf{P}_{1,n}$ che minimizza tale somma, ossia $\|\mathbf{P}_{0,n} \wedge \mathbf{P}_{1,n}\|_1$, resta distante da 0;
- Sostenere che $\prod_{j \in J} (1 - h_{j,n}^2)$ resta distante da 1, per $n \rightarrow +\infty$, è equivalente a dire che gli esperimenti $\mathcal{E}_n = \{\mathbf{P}_{0,n}, \mathbf{P}_{1,n}\}$ sono distanti dall'esperimento banale $\mathcal{B} = \{\mathbf{Q}, \mathbf{Q}\}$ ove le due probabilità coincidono.

2.3 Trasformata di Hellinger

Abbiamo visto, nella sezione (2.2), che la distanza di Hellinger si comporta bene con i prodotti diretti di esperimenti, che tratteremo nei prossimi capitoli. Introduciamo ora, quindi, un nuovo operatore, anch'esso molto funzionale nello studio dei prodotti diretti di esperimenti.

Sia $\mathcal{E} = \{P_\theta \mid \theta \in \Theta\}$ un esperimento, e sia $\alpha = \{\alpha_\theta \mid \theta \in \Theta\}$ un vettore tale che:

- Per ogni $\theta \in \Theta$ vale $\alpha_\theta \geq 0$;
- Vale:

$$\sum_{\theta \in \Theta} \alpha_\theta = 1;$$

- Vale $\alpha_\theta > 0$ solo per un numero finito di $\theta \in \Theta$.

Definizione 2.3.1. Si definisce **trasformata di Hellinger** $\mathcal{H}(\mathcal{E})$ dell'esperimento \mathcal{E} , valutata in α :

$$\mathcal{H}(\mathcal{E})(\alpha) \stackrel{\text{def}}{=} \int_{\Omega} \prod_{\theta \in \Theta} (dP_\theta)^{\alpha_\theta}$$

Anche in questo caso, detta μ una misura dominante, si può scrivere:

$$\mathcal{H}(\mathcal{E})(\alpha) = \int_{\Omega} \prod_{\theta \in \Theta} f_\theta^{\alpha_\theta} d\mu,$$

con $f_\theta = dP_\theta/d\mu$, per ogni $\theta \in \Theta$ (è sufficiente richiedere che μ domini le probabilità relativi ai coefficienti strettamente maggiori di 0, che sono in numero finito: una tale misura, dunque, esiste sempre).

Osservazione 22. L'affinità vista nella scorsa sezione:

$$\rho(P, Q) = \int_{\Omega} \sqrt{dP dQ}$$

è nient'altro che la valutazione in $\mathbf{x} = \left[\frac{1}{2} \quad \frac{1}{2}\right]^t$ della trasformata di Hellinger dell'esperimento binario $\mathcal{E} = \{P, Q\}$.

Definizione 2.3.2. Dato un insieme di parametri Θ , si chiama **prodotto diretto** $\mathcal{E} \otimes \mathcal{F}$ di due esperimenti $\mathcal{E} = \{P_\theta \mid \theta \in \Theta\}$ e $\mathcal{F} = \{Q_\theta \mid \theta \in \Theta\}$ l'esperimento che consiste nel compiere prima \mathcal{E} e poi \mathcal{F} (indipendentemente dai risultati ottenuti compiendo \mathcal{E}):

$$\mathcal{E} \otimes \mathcal{F} = \{P_\theta \otimes Q_\theta \mid \theta \in \Theta\}$$

In analogia con la trasformata di Fourier, che ha la seguente proprietà (che enunciamo solamente):

$$\mathcal{F}(f * g)(\xi) = \mathcal{F}(f)(\xi) \cdot \mathcal{F}(g)(\xi) ,$$

(ove al primo membro è presente l'operatore di convoluzione), si ha che la trasformata di Hellinger di un prodotto diretto di esperimenti è il prodotto delle trasformate di Hellinger dei singoli esperimenti.

Per il teorema di Fubini si ha infatti:

$$\int_{\Omega \times \Omega} \prod_{\theta \in \Theta} (d[P_\theta \otimes Q_\theta])^{\alpha_\theta} = \left(\int_{\Omega} \prod_{\theta \in \Theta} (dP_\theta)^{\alpha_\theta} \right) \cdot \left(\int_{\Omega} \prod_{\theta \in \Theta} (dQ_\theta)^{\alpha_\theta} \right) ,$$

da cui si ottiene la proprietà voluta.

Questa è la proprietà più importante della trasformata di Hellinger: ve ne sono tuttavia altre. Per mostrarne alcune, sia Θ finito, e consideriamo il semplice $U(\Theta)$ visto nello scorso capitolo. Ricordiamo che $u \in U(\Theta)$ se e solo se $u_\theta \geq 0$ per ogni $\theta \in \Theta$, e:

$$\sum_{\theta \in \Theta} u_\theta = 1$$

Proposizione 2.3.1. La trasformata di Hellinger di $\mathcal{E} = \{P_\theta \mid \theta \in \Theta\}$:

$$\mathcal{H}(\mathcal{E})(\alpha) = \int_{\Omega} \prod_{\theta \in \Theta} (dP_\theta)^{\alpha_\theta}$$

caratterizza tale esperimento a meno di equivalenza (intesa secondo la definizione (1.2.3)). Inoltre, data una successione di esperimenti $(\mathcal{E}_n)_{n \in \mathbb{N}^+}$, ove si ha $\mathcal{E}_n = \{P_{\theta,n} \mid \theta \in \Theta\}$ per ogni $n \in \mathbb{N}^+$, essa converge a un limite $\mathcal{E} = \{P_\theta \mid \theta \in \Theta\}$, nel senso della distanza Δ , se e solo se si ha il seguente limite:

$$\lim_{n \rightarrow +\infty} \mathcal{H}(\mathcal{E}_n) = \mathcal{H}(\mathcal{E}) ,$$

puntualmente su $U(\Theta)$.

Dimostrazione. Sia, al solito:

$$S = \sum_{\theta \in \Theta} P_\theta ,$$

e sia, per ogni $\theta \in \Theta$, $f_\theta = dP_\theta/dS$ una versione della densità. Allora:

$$\mathcal{H}(\mathcal{E})(\alpha) = \int_{\Omega} \prod_{\theta \in \Theta} f_\theta^{\alpha_\theta} dS$$

dipende solamente dalla legge su $U(\Theta)$ del vettore $[f_{\theta_1} \dots f_{\theta_k}]^t$: dunque esperimenti equivalenti hanno la stessa trasformata di Hellinger, dato che $\mathcal{H}(\mathcal{E})$ dipende solo dal tipo di \mathcal{E} .

L'implicazione opposta è di tipo costruttivo, ed è decisamente più tecnica: non ne diamo i dettagli, e concludiamo qui. \square

Notiamo ora che la proprietà di contiguità può essere controllata facendo uso delle trasformate di Hellinger. Consideriamo allora una successione di esperimenti binari (useremo la notazione già vista in passato). Sia, per $n \in \mathbb{N}^+$ e $0 < \alpha < 1$:

$$\mathcal{H}_n(\alpha) \stackrel{\text{def}}{=} \mathcal{H}(\mathcal{E}_n)(\alpha) = \int_{\Omega} (dP_{0,n})^{1-\alpha} (dP_{1,n})^{\alpha} \quad (2.4)$$

Osservazione 23. Sia \mathcal{E} un esperimento binario, e siano:

$$f_0 = dP_0/(dP_0 + dP_1) \quad , \quad f_1 = dP_1/(dP_0 + dP_1)$$

Sia $D \subseteq X$ l'insieme dove entrambe le densità f_0, f_1 sono non nulle. Allora, se $0 < \alpha < 1$, vale:

$$\mathcal{H}(\mathcal{E})(\alpha) = \int_{\Omega} (dP_0)^{1-\alpha} (dP_1)^{\alpha} = \int_D (dP_0)^{1-\alpha} (dP_1)^{\alpha} \quad ,$$

da cui:

- Si ha:

$$\lim_{\alpha \rightarrow 0^+} \mathcal{H}(\mathcal{E})(\alpha) = P_0(D) \quad ,$$

ossia la massa della parte di P_0 dominata da P_1 ;

- Si ha:

$$\lim_{\alpha \rightarrow 1^-} \mathcal{H}(\mathcal{E})(\alpha) = P_1(D) \quad ,$$

ossia la massa della parte di P_1 dominata da P_0 .

Attraverso il teorema (2.1.1), possiamo enunciare (e non dimostrare, perchè la dimostrazione è ora ovvia) la seguente proposizione.

Proposizione 2.3.2. Con la notazione della formula (2.4), $(P_{1,n})_{n \in \mathbb{N}^+}$ è contigua a $(P_{0,n})_{n \in \mathbb{N}^+}$ se e solo se:

$$\lim_{\alpha \rightarrow 1^-} \liminf_{n \rightarrow +\infty} \mathcal{H}_n(\alpha) = 1$$

Osservazione 24. Lo statistico E.N.Torgersen ha dimostrato (noi non lo faremo) che se un esperimento \mathcal{E} è migliore di un esperimento \mathcal{F} , allora $\mathcal{H}(\mathcal{E}) \leq \mathcal{H}(\mathcal{F})$, e che il viceversa a priori non vale. Si veda, a tal riguardo, [13].

Capitolo 3

Shift gaussiani e esperimenti poissoniani

In questo capitolo si parlerà di due classi di esperimenti molto importanti, per via della loro versatilità dal punto di vista computazionale:

- Chiameremo **esperimenti gaussiani** i cosiddetti **shift gaussiani**, ove sarà supposta nota (e fissata a priori) la struttura della covarianza, e lo studio si restringerà alla stima del valore atteso. Analizzeremo il tutto nel caso infinito-dimensionale, quello più generale, e attraverso degli esempi vedremo che alcune proprietà statistiche possono essere dedotte da proprietà puramente geometriche dell'insieme Θ dei parametri, visto come sottoinsieme di un opportuno spazio di Hilbert \mathbb{H} ;
- Gli **esperimenti poissoniani** si presentano quando si studiano i cosiddetti **processi puntuali**, di cui daremo una breve presentazione. Non sono computazionalmente trattabili quanto gli esperimenti gaussiani, ma sono decisamente più importanti: un arbitrario esperimento \mathcal{E} può essere infatti approssimato mediante un esperimento poissoniano $\mathbf{P}(\mathcal{E})$, mediante la tecnica della **poissonizzazione**.

Entrambi i tipi di esperimento, gaussiano e poissoniano, assumono un'importanza davvero rilevante soprattutto in situazioni in cui lo statistico effettua più osservazioni, tra loro indipendenti.

3.1 Shift gaussiani

Gli esperimenti gaussiani, di cui parleremo in questa sezione, sono molto usati nell'ambito della statistica asintotica, sia per la loro trattabilità in termini computazionali, sia per l'esistenza del Teorema Limite Centrale.

Dopo aver dato la definizione, vedremo che ognuno di questi esperimenti definisce uno spazio di Hilbert \mathbb{H} , non necessariamente finitamente generato. Vedremo, infine, che le caratteristiche geometriche dell'insieme Θ , visto come sottoinsieme di \mathbb{H} , determineranno le proprietà statistiche dell'esperimento relativo.

Consideriamo un esperimento $\mathcal{G} = \{G_\theta \mid \theta \in \Theta\}$, con $\Theta \subseteq \mathbb{R}^k$, tale che G_θ ha, per ogni $\theta \in \Theta$, la seguente densità rispetto alla misura di Lebesgue:

$$f_\theta(\mathbf{x}) = \sqrt{\frac{\det M_\theta}{(2\pi)^k}} e^{-\frac{1}{2}\langle M_\theta(\mathbf{x}-\theta), (\mathbf{x}-\theta) \rangle} \quad (3.1)$$

Sappiamo che, per una tale distribuzione:

- θ è il **vettore delle medie**;
- $M_\theta \in GL(k, \mathbb{R})$, e la sua inversa è la **matrice delle covarianze**.

Useremo la seguente notazione, d'ora in poi (anche se quella definita non è a priori una norma):

$$\|\mathbf{x}\|_M^2 \stackrel{\text{def}}{=} \langle M\mathbf{x}, \mathbf{x} \rangle$$

Osservazione 25. Notiamo che $\theta \in \Theta$ ha un duplice ruolo: è sia l'indice della distribuzione, sia il suo vettore delle medie.

A priori M_θ potrebbe variare al variare di $\theta \in \Theta$ (si parlerebbe allora di **eteroschedasticità**), mentre supporremo per semplicità $M_\theta = M$ (e si parlerà allora di **omoschedasticità**). Infatti, seguendo la definizione data a suo tempo da C.F. Gauss, bisogna controllare che, in tali ipotesi, la stima di massima verosimiglianza per un campione (X_1, \dots, X_n) di taglia n e legge G_θ è ben definita, ed è la media empirica.

In effetti, considerato un campione (X_1, \dots, X_n) di taglia n e legge $P_\theta = f_\theta \cdot \lambda^k$, la stima di massima verosimiglianza $\hat{\theta}$ risolve la seguente equazione:

$$\left. \frac{d}{d\theta} L(\theta, \omega) \right|_{\theta=\hat{\theta}(\omega)} = 0,$$

da cui bisogna imporre (tralasciando le costanti moltiplicative, ininfluenti per i nostri scopi):

$$\sum_{j=1}^k \frac{d}{d\theta} \langle M(\mathbf{x}_j - \theta), (\mathbf{x}_j - \theta) \rangle = -2 \sum_{j=1}^k M(\mathbf{x}_j - \theta) = 0$$

Sfruttando l'invertibilità di M , si deve allora imporre:

$$\sum_{j=1}^k (\mathbf{x}_j - \theta) = \sum_{j=1}^k \mathbf{x}_j - k\theta = 0,$$

da cui, come previsto, $\hat{\theta}(\omega)$ risulta uguale alla media empirica.

Scelto ora $s \in \Theta$, si ha:

$$\begin{aligned} \log \left(\frac{dG_\theta}{dG_s}(\mathbf{x}) \right) &= \log \left(e^{-\frac{1}{2}[\|\mathbf{x}-\theta\|_M^2 - \|\mathbf{x}-s\|_M^2]} \right) = \\ &= -\frac{1}{2} [\langle M(\mathbf{x}-\theta), \mathbf{x}-\theta \rangle - \langle M(\mathbf{x}-s), \mathbf{x}-s \rangle] = (\theta-s)^t M\mathbf{x} - \frac{1}{2} \|\theta-s\|_M^2, \end{aligned}$$

da cui:

$$\log \left(\frac{dG_\theta}{dG_s} \right) = (\theta-s)^t M\mathbf{X} - \frac{1}{2} \|\theta-s\|_M^2 \quad (3.2)$$

Definizione 3.1.1. Un processo stocastico $(X_t)_{t \in \mathcal{T}}$ è detto **processo gaussiano** se per ogni insieme finito di indici t_1, \dots, t_h , $(X_{t_1}, \dots, X_{t_h})$ è un vettore gaussiano. La legge del processo è identificata dal vettore delle medie $\mathbf{m}(t) = \mathbb{E}[X_t]$ e dal nucleo di covarianza $K(t, u) = \text{Cov}(X_t, X_u)$.

Risulta in effetti che $\theta \rightarrow \log(dG_\theta/dG_s)$ è un processo gaussiano (non entriamo nei dettagli).

Definizione 3.1.2. Si definisce **esperimento gaussiano** (o **shift gaussiano**) indicizzato da Θ una famiglia \mathcal{G} di probabilità tale che:

- Per ogni $\theta_1, \theta_2 \in \Theta$: $G_{\theta_1} \sim G_{\theta_2}$;
- Esiste $s \in \Theta$, detto **origine** di Θ , tale che $\theta \rightarrow \log(dG_\theta/dG_s)$ è un processo gaussiano.

Scriviamo ora:

$$\log\left(\frac{dG_t}{dG_s}\right) = \mathbf{X}(t) + \mathbf{b}(t),$$

ove:

$$\mathbf{b}(t) \stackrel{\text{def}}{=} \mathbb{E}^s \left[\log\left(\frac{dG_t}{dG_s}\right) \right]$$

Allora si ha:

- Per ogni $t \in \Theta$, $\mathbb{E}^s[\mathbf{X}(t)] = 0$;
- Per ogni $t, u \in \Theta$, $K(t, u) = \text{Cov}_\circ(\mathbf{X}(t), \mathbf{X}(u)) = \mathbb{E}^\circ[\mathbf{X}(t)\mathbf{X}(u)]$ (il simbolo usato enfatizza che il valore atteso può essere calcolato rispetto a una qualsiasi probabilità G_θ , $\theta \in \Theta$: giustificheremo ciò più avanti).

Per effettuare qualche osservazione sugli esperimenti gaussiani, appena introdotti, notiamo che la trasformata di Hellinger di \mathcal{G} , per una funzione opportuna α , risulta essere:

$$\mathcal{H}(\mathcal{G})(\alpha) = e^{\frac{1}{2}[\sum_{t, u \in \Theta} \alpha(t)\alpha(u)K(t, u) - \sum_{t \in \Theta} \alpha(t)K(t, t)]} \quad (3.3)$$

Osservazione 26. Tale espressione non dipende da s . Concludiamo che, nella definizione di shift gaussiano data precedentemente, s non ricopre un ruolo particolare, ma il processo $t \rightarrow \mathbf{X}(t)$, se è gaussiano sotto G_s , è gaussiano sotto G_θ per ogni $\theta \in \Theta$.

In particolare l'affinità è:

$$\rho(t, u) = \int_{\Omega} \sqrt{dG_t dG_u} = e^{\frac{1}{8}[2K(t, u) - K(t, t) - K(t, u)]},$$

e si può dimostrare che il tipo di un esperimento \mathcal{G} , univocamente determinato dalla sua trasformata di Hellinger e quindi in questo caso dal nucleo di covarianza $K(t, u)$, è in realtà univocamente determinato dall'affinità $\rho(t, u)$, dato che vale:

$$K(t, u) = \frac{1}{2} [q^2(t, u) - q^2(s, t) - q^2(s, u)],$$

ove $q^2(t, u) \stackrel{\text{def}}{=} -8 \log(\rho(t, u))$ è una funzione di cui analizzeremo il significato tra poco.

Cambiamo ora l'origine di Θ , nel senso della definizione (3.1.2), considerando u e non s . Allora, riprendendo il secondo esempio dopo la proposizione (2.1.3), relativo a distribuzioni

normali multidimensionali, si può dimostrare che la funzione di covarianza del processo è indipendente dall'origine di Θ (dunque $K(t, u) = \mathbb{E}^\circ [\mathbf{X}(t)\mathbf{X}(u)]$ è ben definita), e che se $\mathbb{E}^s [\mathbf{X}(t)] = 0$, allora si ha $\mathbb{E}^u [\mathbf{X}(t)] = K(t, u)$.

I valori attesi subiscono dunque uno shift, quando si cambia l'origine dell'insieme dei parametri: ecco perchè questi esperimenti son detti shift gaussiani.

Osservazione 27. Finora abbiamo scritto la trasformata di Hellinger usando le sommatorie. Nella formula (3.3), ad esempio, siamo partiti dalla seguente espressione:

$$\mathcal{H}(\mathcal{G})(\alpha) = \int_{\Omega} e^{\sum_{t \in \Theta} \alpha(t) [\mathbf{X}(t) + \mathbf{b}(t)]} dG_s$$

Per i nostri scopi sarà più comodo sostituire le sommatorie con gli integrali. Nel caso precedente avremmo dunque scritto:

$$\mathcal{H}(\mathcal{G})(\alpha) = \int_{\Omega} e^{\int_{\Theta} [\mathbf{X}(t) + \mathbf{b}(t)] d\mu_\alpha(t)} dG_s,$$

ove $\mu_\alpha \in \mathcal{M}$ è un opportuno elemento dello spazio vettoriale delle misure finite con segno e con supporto finito su Θ , che abbiamo denotato con \mathcal{M} . A livello di notazione, inoltre, indichiamo con \mathcal{M}_0 il sottospazio vettoriale di \mathcal{M} delle misure μ tali che $\mu(\Theta) = 0$.

Definiamo infine:

$$\Lambda(t, s) \stackrel{\text{def}}{=} \log(dG_t/dG_s)$$

Data $\mu \in \mathcal{M}_0$, l'integrale di tale funzione rispetto a μ non varia al variare di $s \in \Theta$. Infatti, dati $s_1, s_2 \in \Theta$, si ha:

$$\int_{\Theta} \Lambda(t, s_1) d\mu(t) = \int_{\Theta} \Lambda(t, s_2) + \Lambda(s_2, s_1) d\mu(t) = \int_{\Theta} \Lambda(t, s_2) d\mu(t),$$

essendo $\mu(\Theta) = 0$. Definiamo ora, per $\mu \in \mathcal{M}_0$, la seguente seminorma:

$$\|\mu\|^2 \stackrel{\text{def}}{=} \text{Var}_s \left(\int_{\Theta} \Lambda(t, s) d\mu(t) \right)$$

Sfruttando il fatto che $\mu \in \mathcal{M}_0$, e usando i teoremi di Fubini-Tonelli, si ha:

$$\begin{aligned} \|\mu\|^2 &= \mathbb{E}^s [\mathbb{E}^\mu [\Lambda(t, s)] \mathbb{E}^\mu [\Lambda(u, s)]] - (\mathbb{E}^s [\mathbb{E}^\mu [\Lambda(t, s)]])^2 = \\ &= \mathbb{E}^s [\mathbb{E}^{\mu \times \mu} [\mathbf{X}(t)\mathbf{X}(u)]] = \mathbb{E}^{\mu \times \mu} [\mathbb{E}^s [\mathbf{X}(t)\mathbf{X}(u)]] = \mathbb{E}^{\mu \times \mu} [K(t, u)], \end{aligned}$$

dato che:

$$\mathbb{E}^s [\mathbb{E}^\mu [\Lambda(t, s)]] = \mathbb{E}^s [\mathbb{E}^\mu [\mathbf{X}(t)]] = \mathbb{E}^\mu [\mathbb{E}^s [\mathbf{X}(t)]] = 0$$

Dato che per quanto visto precedentemente $K(t, u) = \mathbb{E}^s [\mathbf{X}(t)\mathbf{X}(u)] = \mathbb{E}^\circ [\mathbf{X}(t)\mathbf{X}(u)]$, in effetti si ha:

$$\|\mu\|^2 = \text{Var}_\diamond \left(\int_{\Theta} \mathbf{X}(t) d\mu(t) \right)$$

Si può dimostrare, abbastanza facilmente, che valgono le seguenti espressioni alternative:

$$\|\mu\|^2 = \int_{\Theta \times \Theta} K(t, u) d\mu(t) d\mu(u) = -\frac{1}{2} \int_{\Theta \times \Theta} q^2(t, u) d\mu(t) d\mu(u) \quad (3.4)$$

Osservazione 28. Tale seminorma soddisfa l'identità del parallelogramma, dunque date $\mu, \nu \in \mathcal{M}_0$ si ha:

$$\left\| \frac{\mu + \nu}{2} \right\|^2 + \left\| \frac{\mu - \nu}{2} \right\|^2 = \frac{1}{2} (\|\mu\|^2 + \|\nu\|^2) \quad (3.5)$$

Ciò può essere utile per definire uno spazio di Hilbert \mathbb{H} associato a \mathcal{G} : si definisce $\mathbf{0}_{\mathbb{H}}$ la misura $\mu \in \mathcal{M}_0$ tale che $\|\mu\| = 0$ (supponiamo che vi sia unicità), e poi si procede con il completamento isometrico e denso dello spazio, che diventa quindi uno spazio di Hilbert \mathbb{H} incluso in \mathcal{M}_0 .

Noi vedremo però una costruzione diversa, che ci permetterà di estendere l'insieme dei parametri Θ a tutto \mathbb{H} .

Consideriamo $s \in \Theta$ dato dalla definizione (3.1.2) (\mathcal{G} è uno shift gaussiano). A $\theta \in \Theta$ associamo la misura $\mu_\theta = \delta_\theta - \delta_s \in \mathcal{M}_0$, somma algebrica di due misure del tipo Delta di Dirac. Se ora $\mathbf{X}(t) = \Lambda(t, s) - \mathbb{E}^s[\Lambda(t, s)]$, per $\mu \in \mathcal{M}_0$ definiamo il seguente funzionale lineare, da \mathcal{M}_0 nello spazio delle variabili gaussiane, munito della norma L^2 :

$$\langle \mathbf{X}, \mu \rangle \stackrel{\text{def}}{=} \int_{\Theta} \mathbf{X}(t) d\mu(t)$$

Tale funzionale è continuo, e si estende per continuità rispetto alla norma L^2 allo spazio \mathbb{H} . Per $\mu \in \mathbb{H}$ si definisce quindi $\langle \mathbf{X}, \mu \rangle$ operando un limite.

Si hanno le seguenti uguaglianze:

$$\mathbb{E}^s[\langle \mathbf{X}, \mu \rangle] = 0 \quad , \quad \mathbb{E}^s[\langle \mathbf{X}, \mu \rangle^2] = \|\mu\|^2$$

Con una notazione simile a quella usata per ottenere la formula (3.2), ma adattata al contesto, sia $G_{(\mu)}$ la probabilità tale che:

$$dG_{(\mu)} = e^{\langle \mathbf{X}, \mu \rangle - \frac{1}{2}\|\mu\|^2} dG_s \quad (3.6)$$

Per $\theta \in \Theta$, si ha $G_\theta = G_{(\delta_\theta - \delta_s)}$, per cui d'ora in poi identificheremo $\theta \in \Theta$ con $(\delta_\theta - \delta_s) \in \mathbb{H}$.

Date ora $\mu, \nu \in \mathbb{H}$, si ha:

$$\rho(G_{(\mu)}, G_{(\nu)}) = \int_{\Omega} \sqrt{dG_{(\mu)} dG_{(\nu)}} = \int_{\Omega} e^{\frac{1}{2}\langle x, \mu + \nu \rangle - \frac{1}{4}(\|\mu\|^2 + \|\nu\|^2)} dG_s$$

Sfruttando le relazioni (3.5) e (3.6) si ha:

$$\rho(G_{(\mu)}, G_{(\nu)}) = \int_{\Omega} e^{\langle x, \frac{\mu + \nu}{2} \rangle - \frac{1}{2}(\|\frac{\mu + \nu}{2}\|^2 + \|\frac{\mu - \nu}{2}\|^2)} dG_s = e^{\frac{1}{8}\|\mu - \nu\|^2}$$

Dall'espressione (3.4), poi, si ottiene il significato intrinseco della funzione q :

$$q(t, u) = \|\delta_t - \delta_u\|$$

Questa formula si ottiene valutando esplicitamente $\|\delta_t - \delta_u\|$: non riportiamo i dettagli.

Osservazione 29. Potrebbero esistere punti distinti $t, u \in \Theta$ tali che $q(t, u) = 0$: noi assumeremo però che q sia una metrica su Θ .

Consideriamo ora due esperimenti gaussiani $\mathcal{G}_1, \mathcal{G}_2$, con insiemi dei parametri (muniti di una distanza) $(\Theta_1, q_1), (\Theta_2, q_2)$, e spazi di Hilbert associati $\mathbb{H}_1, \mathbb{H}_2$.

Teorema 3.1.1. Supponiamo che esista un'isometria ξ tra Θ_1 e Θ_2 , ossia un'applicazione iniettiva tale che per ogni $t, u \in \Theta_1$:

$$q_2(\xi(t), \xi(u)) = q_1(t, u)$$

Allora essa si estende ad un'isometria affine $\hat{\xi} : \mathbb{H}_1 \rightarrow \mathbb{H}_2$, che diventa lineare se l'origine s di Θ_1 e l'origine σ di Θ_2 (secondo la definizione (3.1.2)) sono tali che $\xi(s) = \sigma$ (e tale condizione si può ottenere, se gli spazi Θ_1 e Θ_2 lo permettono, definendo a partire da ξ l'isometria associata $\xi^*(t) = \xi(t) + (\sigma - \xi(s))$).

Tale teorema, in realtà, deriva direttamente da come sono stati definiti, in via costruttiva, gli spazi di Hilbert \mathbb{H}_1 e \mathbb{H}_2 .

Osservazione 30. Discende da tutto quanto visto in questa sezione che gli esperimenti \mathcal{G}_1 e $\mathcal{G}_{\xi(1)}$, ove $\mathcal{G}_{\xi(1)} = \{G_{\xi(t)} \mid t \in \Theta_1\}$, sono equivalenti.

Le proprietà statistiche di un esperimento gaussiano \mathcal{G} , dunque, dipendono solamente dalle proprietà geometriche dello spazio dei parametri Θ , visto come sottoinsieme di \mathbb{H} .

Concludiamo con un paio di applicazioni di quanto visto in questa sezione.

Esempio. Come primo esempio, consideriamo uno shift gaussiano k -dimensionale. Consideriamo un vettore aleatorio gaussiano $\mathbf{X} \sim N_k(\mathbf{0}, \Gamma)$, ove $\Gamma \in M(k, \mathbb{R})$ è la matrice delle covarianze, tale che $\Gamma_{i,j} = \mathbb{E}[X_i X_j]$, per $i, j = 1, \dots, n$.

Considereremo allora la famiglia dei vettori gaussiani del tipo $\mathbf{Y}(t) = t + \mathbf{X}$, al variare di $t \in \mathbb{R}^k$. Detta λ la misura di Lebesgue su \mathbb{R}^k , sia poi $P_t = \mathbf{Y}(t) \cdot \lambda$. Si ha:

$$\Lambda(t, 0) = \log \left(\frac{dP_t}{dP_0} \right) = t' M \mathbf{X} - \frac{1}{2} t' M t,$$

ove $M = \Gamma^{-1}$ (stiamo supponendo, in base alla formula (3.1) e alla definizione (3.1.2), che Γ sia invertibile).

Se ora $\mathbf{X}(t) = t' M \mathbf{X}$, per $\mu \in \mathcal{M}_0$ si ha:

$$\int_{\mathbb{R}^k} \mathbf{X}(t) d\mu(t) = \int_{\mathbb{R}^k} t' M \mathbf{X} d\mu(t) = \left[\int_{\mathbb{R}^k} t d\mu(t) \right]' M \mathbf{X}$$

La mappa che manda $\mu \in \mathcal{M}_0$ in $\int_{\mathbb{R}^k} t d\mu(t) \in \mathbb{R}^k$ è lineare, e la seminorma è data in questo caso dalla formula seguente:

$$\|\mu\|^2 = \left[\int_{\mathbb{R}^k} t d\mu(t) \right]' M \left[\int_{\mathbb{R}^k} t d\mu(t) \right],$$

da cui si ottiene che lo spazio di Hilbert \mathbb{H} associato all'esperimento è isomorfo allo spazio $(\mathbb{R}^k, \|\cdot\|_M)$, ove $\|\theta\|_M = \theta' M \theta$.

Esempio. Analizzeremo ora un esempio di **segnale con disturbo**. Consideriamo l'intervallo chiuso $\mathcal{T} = [0, 1]$, che chiameremo **intervallo temporale**. Consideriamo ora il processo stocastico $(Y(\tau))_{\tau \in \mathcal{T}}$, tale che, per ogni $\tau \in \mathcal{T}$ vale:

$$dY(\tau) = g(\tau) d\tau + \sigma dW(\tau), \quad (3.7)$$

ove σ è una costante nota, e $(W(\tau))_{\tau \in \mathcal{T}}$ è un processo di Wiener: tale processo costituisce il cosiddetto **rumore bianco gaussiano**. Lo spazio dei parametri, in questo caso, è uno spazio di funzioni: nella formula (3.7), ad esempio, il parametro è la funzione g .

Con un'integrazione si ottiene allora:

$$Y(\tau) = \int_0^\tau g(s) ds + \sigma W(\tau) \quad (3.8)$$

Definizione 3.1.3. Dato uno spazio di Hilbert $(\mathcal{H}, \|\cdot\|)$, si definisce **processo gaussiano canonico** di \mathcal{H} è una mappa lineare $\mathbf{Z} : x \rightarrow Z(x)$, tale che per ogni $x \in \mathcal{H}$ si ha:

- $\mathbb{E}[Z(x)] = 0$;
- $\mathbb{E}[|Z(x)|^2] = \|x\|^2$.

Osservazione 31. Notiamo che:

$$\mathbb{E}[|Z(x-y)|^2] = \|x-y\|^2 = \|x\|^2 + \|y\|^2 - 2\langle x, y \rangle,$$

ma per linearità vale anche:

$$\mathbb{E}[|Z(x-y)|^2] = \mathbb{E}[|Z(x) - Z(y)|^2] = \mathbb{E}[|Z(x)|^2] + \mathbb{E}[|Z(y)|^2] - 2\mathbb{E}[Z(x)Z(y)]$$

Direttamente dalla definizione (3.1.3), si ha allora che per ogni $x, y \in \mathcal{H}$ vale:

$$\text{Cov}(Z(x), Z(y)) = \langle x, y \rangle$$

Enunciamo ora un teorema che useremo tra poco: per una trattazione completa, si veda [12].

Teorema 3.1.2 (di Bochner). Una funzione continua $f : \mathcal{H} \rightarrow \mathbb{C}$, definita su uno spazio di Hilbert \mathcal{H} , è la funzione caratteristica di un'opportuna probabilità μ_f su \mathcal{H} se e solo se $f(0) = 1$ e f risulta **semidefinita positiva**, nel senso che scelti comunque $t_1, \dots, t_n \in \mathcal{H}$ e $z_1, \dots, z_n \in \mathbb{C}$, vale:

$$\sum_{h=1}^n \sum_{k=1}^n f(t_h - t_k) z_h \bar{z}_k \geq 0$$

Tornando alla nostra definizione di processo gaussiano canonico, la sua funzione caratteristica sarebbe:

$$\phi_{\mathbf{Z}}(x) = \mathbb{E}\left[e^{iZ(x)}\right] = e^{-\frac{\|x\|^2}{2}},$$

da cui esiste una probabilità su \mathcal{H} che genera la mappa desiderata, per il teorema (3.1.2). Si ha infatti $\phi_{\mathbf{Z}}(0) = 1$, e non è difficile dimostrare che $\phi_{\mathbf{Z}}$ è semidefinita positiva.

Procedendo in maniera costruttiva, invece, data una base ortonormale $\{x_\alpha\}_{\alpha \in A}$ e considerato un insieme di variabili gaussiane indipendenti $\{\gamma_\alpha\}_{\alpha \in A}$, tutte con media nulla e varianza unitaria, la mappa può essere definita in questo modo:

$$x = \sum_{\alpha \in A} \hat{c}_\alpha x_\alpha \rightarrow Z(x) = \sum_{\alpha \in A} \hat{c}_\alpha \gamma_\alpha$$

Passiamo ora alla costruzione del rumore bianco: sia $\mathcal{H} = L^2(\mathbb{R})$, lo spazio di Hilbert delle funzioni reali a quadrato integrabile rispetto alla misura di Lebesgue 1-dimensionale λ ,

munito della norma $\|\cdot\|_{L^2}$ (che d'ora in poi indicheremo con $\|\cdot\|$). Sia poi \mathbf{Z} il processo gaussiano canonico di \mathcal{H} .

Per ottenere W , allora, si può definire la mappa $\tau \rightarrow W(\tau) = Z(\mathbb{I}_{<\tau})$, con $\mathbb{I}_{<\tau}$ funzione indicatrice dell'intervallo semiaperto $[0, \tau[$.

Le formule (3.7) e (3.8), ora, sono chiare. Per ottenere lo shift gaussiano $\mathcal{G} = \{G_\theta \mid \theta \in \Theta\}$, si considera allora un sottoinsieme Θ di $L^2(\mathbb{R})$ (supponiamo per semplicità che la funzione identicamente nulla appartenga a Θ). Allora si ha $\mathcal{H} = \text{Span}(\Theta)$, ove il generato e la chiusura sono intese in $L^2(\mathbb{R})$.

Osservazione 3.2. Dato che un esperimento gaussiano è definito (a meno di equivalenza) dalle proprietà geometriche di Θ , visto come sottoinsieme di \mathcal{H} , ogni esperimento gaussiano può essere rappresentato usando la formula (3.7), se Θ , munito della distanza associata q , risulta uno spazio metrico separabile.

Concludiamo con un risultato molto recente sui rumori gaussiani, formulato nel 1996. Per una trattazione completa, si veda [11].

Teorema 3.1.3 (di Nussbaum). Sia Θ un insieme di densità definite in $[0, 1]$, con $\Theta \subseteq \Sigma_{\varepsilon, M, \alpha}$, per qualche $\varepsilon > 0, M > 0, \alpha > \frac{1}{2}$, ove:

$$\Sigma_{\varepsilon, M, \alpha} = \left\{ f : [0, 1] \rightarrow \mathbb{R} \mid \int_0^1 f(x) dx = 1, f \geq \varepsilon, f \in C^{0, \alpha}, \|f\|_{C^{0, \alpha}} = M \right\}$$

Per ogni $n \in \mathbb{N}^+$ si considerino gli esperimenti:

- $\mathcal{E}_n = \{P_f^n \mid f \in \Theta\}$ ove $P_f^n = (f \cdot \lambda)^{\otimes n}$;
- $\mathcal{G}_n = \{G_f^n \mid f \in \Theta\}$, ove G_f^n è la probabilità indotta dal processo stocastico seguente:

$$dY_{f, n}(\tau) = \sqrt{f(\tau)} d\tau + \frac{1}{2\sqrt{n}} dW(\tau), \quad \tau \in [0, 1]$$

Allora si ha $\lim_{n \rightarrow +\infty} \Delta(\mathcal{E}_n, \mathcal{G}_n) = 0$.

Se dunque l'insieme dei parametri soddisfa le condizioni espone nel teorema (3.1.3), l'esperimento di partenza può essere ben approssimato considerando un opportuno shift gaussiano.

3.2 Esperimenti poissoniani

Andremo ora a esaminare i cosiddetti esperimenti poissoniani, molto utili quando si analizzano fenomeni naturali (faremo alcuni esempi più avanti). In particolare introdurremo la tecnica della **poissonizzazione** di un dato esperimento \mathcal{E} , ed esporremo alcuni risultati, in questo e nei prossimi capitoli, che mostreranno l'utilità di questa tipologia di esperimenti.

Sia Ω un insieme, e sia $\mathcal{A} \subseteq \mathcal{P}(\Omega)$ un'algebra di parti, ossia un sottoinsieme di $\mathcal{P}(\Omega)$ contenente Ω e chiuso per unione e complementazione.

Definizione 3.2.1. Si definisce **funzione di massa** (in inglese **content**) una funzione $\lambda : \mathcal{A} \rightarrow [0, +\infty[$ finitamente additiva: per due insiemi $A, B \in \mathcal{A}$ disgiunti, allora, vale $\lambda(A \cup B) = \lambda(A) + \lambda(B)$ (useremo d'ora in avanti il termine inglese, per enfatizzare il fatto che non si tratta a priori di una misura, ossia di una funzione che sia anche σ -additiva).

Definizione 3.2.2. Dato un content λ , si definisce **processo poissoniano non omogeneo** con funzione di intensità λ una funzione Z a valori nella classe delle variabili aleatorie di legge poissoniana, tale che:

- Se $A_1, \dots, A_n \in \mathcal{A}$ sono disgiunti a coppie, allora:

$$Z(A_1 \cup \dots \cup A_n) = Z(A_1) + \dots + Z(A_n);$$

- Se $A_1, \dots, A_n \in \mathcal{A}$ sono disgiunti a coppie, allora le variabili $Z(A_1), \dots, Z(A_n)$ sono indipendenti;
- Per $A \in \mathcal{A}$, $Z(A)$ è una variabile di legge poissoniana con valore atteso $\lambda(A)$.

Osservazione 33. Anche in questo caso dovremmo dimostrare che dato un content λ esiste uno e un solo processo poissoniano con funzione di intensità λ : non lo faremo nel dettaglio. Ci limitiamo a dire che anche in questo caso si usa una generalizzazione del teorema (3.1.2); l'unicità deriva invece da alcune proprietà delle funzioni caratteristiche.

Definizione 3.2.3. Nelle notazioni delle definizioni (3.2.1) e (3.2.2), dato un insieme Θ di parametri, per ogni $\theta \in \Theta$ sia μ_θ un content definito su \mathcal{A} . Sia poi, per ogni $\theta \in \Theta$, P_θ la legge del processo poissoniano Z_θ con funzione di intensità μ_θ . Allora $\mathcal{P} = \{P_\theta \mid \theta \in \Theta\}$ è detto **esperimento poissoniano**.

Gli esperimenti poissoniani sono molto frequenti quando si modellizzano fenomeni naturali, soprattutto in ambito fisico o biologico. Alcuni esempi sono i seguenti:

- I processi poissoniani sono usati per modellizzare la quantità di reclami pervenuti ad un'agenzia assicurativa, ognuno corredato della cifra in denaro richiesta. Sia $\Omega = [0, +\infty[\times [0, +\infty[$ il prodotto dell'asse temporale per quello finanziario: allora $Z(A \times B)$ descrive il numero di reclami pervenuti nell'arco temporale A che sono corredati da richieste in denaro appartenenti al range B ;
- Spesso la durata di vita di un nucleo di una sostanza radioattiva è governata da una legge esponenziale, nel senso che dato un tempo $t > 0$, la probabilità che il nucleo abbia una durata di vita maggiore di t è $\mathbb{P}(T > t) = e^{-\lambda t}$, ove λ è nota come **costante di decadimento**. Considerato allora una mole di tale sostanza, e fissato $h \in \mathbb{N}$, si può usare un processo poissoniano per calcolare la probabilità che il numero di decadimenti in un certo intervallo temporale $[\tau_1, \tau_2]$, con $0 < \tau_1 < \tau_2$, sia h ;
- Gli incidenti stradali, in particolare la loro posizione spazio-temporale, possono essere descritti usando dei processi poissoniani: ciò può servire per prevenirne altri in futuro, attraverso un miglioramento dell'illuminazione, o della segnaletica, in punti opportuni della città.

Potremmo fare altri esempi, ma già questi dovrebbero essere sufficienti per comprendere l'importanza degli esperimenti poissoniani. Piuttosto, cominciamo con una prima proposizione.

Proposizione 3.2.1. Sia \mathcal{P} un esperimento poissoniano, nel senso della definizione (3.2.3). Allora la trasformata di Hellinger di \mathcal{P} , data una funzione α opportuna (si riveda la definizione (2.3.1)) è:

$$\mathcal{H}(\mathcal{P})(\alpha) = \int_{\Omega} \prod_{\theta \in \Theta} (dP_\theta)^{\alpha_\theta} = e^{\int_{\Omega} [\prod_{\theta \in \Theta} (d\mu_\theta)^{\alpha_\theta} - \sum_{\theta \in \Theta} \alpha_\theta d\mu_\theta]} \quad (3.9)$$

In particolare, l'affinità tra due probabilità P_s, P_t è:

$$\rho(s, t) = e^{-\frac{1}{2} \int_{\Omega} (\sqrt{d\mu_s} - \sqrt{d\mu_t})^2}$$

Dimostrazione. Sia $A \subseteq \Omega$: allora $Z_{\theta}(A)$ è una variabile poissoniana di parametro λ_{θ} , con $\lambda_{\theta} = \mu_{\theta}(A)$, da cui per $k \in \mathbb{N}$ si ha:

$$p_{\theta}(k) = e^{-\lambda_{\theta}} \frac{\lambda_{\theta}^k}{k!}$$

Si ottiene allora, per tale insieme, la trasformata di Hellinger seguente:

$$\phi_A(\alpha) = \sum_{n \in \mathbb{N}} \left[\frac{1}{k!} e^{-\sum_{\theta \in \Theta} \lambda_{\theta} \alpha_{\theta}} \left(\prod_{\theta \in \Theta} \lambda_{\theta}^{\alpha_{\theta}} \right)^k \right] = e^{-\sum_{\theta \in \Theta} \alpha_{\theta} \lambda_{\theta} + \prod_{\theta \in \Theta} \lambda_{\theta}^{\alpha_{\theta}}}$$

Se ora consideriamo l'unione disgiunta di una famiglia finita di insiemi A_1, \dots, A_k (sia essa B), per indipendenza delle variabili poissoniane si ha $\phi_B(\alpha) = \prod_{j=1}^k \phi_{A_j}(\alpha)$, e per ottenere la formula (3.9) si può raffinare sempre più la partizione di X , e poi procedere operando un limite. \square

Ricordiamo che, dato un esperimento $\mathcal{E} = \{P_{\theta} \mid \theta \in \Theta\}$, due probabilità P_s, P_t sono singolari (ossia hanno supporti essenziali disgiunti) se e solo se $\rho(s, t) = 0$. Dato allora un esperimento poissoniano \mathcal{P} , come nella definizione (3.2.3), possiamo dunque scrivere $\Theta = \coprod_{j \in J} \Theta_j$, ove:

- Se θ_1, θ_2 appartengono ad un insieme Θ_j comune, allora:

$$\int_{\Omega} \left(\sqrt{d\mu_{\theta_1}} - \sqrt{d\mu_{\theta_2}} \right)^2 < +\infty \Rightarrow \rho(\theta_1, \theta_2) > 0,$$

da cui le misure $P_{\theta_1}, P_{\theta_2}$ non sono singolari (non è detto che siano equivalenti, a priori);

- Se $\theta_1 \in \Theta_{j_1}, \theta_2 \in \Theta_{j_2}$, con $j_1 \neq j_2$, allora:

$$\int_{\Omega} \left(\sqrt{d\mu_{\theta_1}} - \sqrt{d\mu_{\theta_2}} \right)^2 = +\infty \Rightarrow \rho(\theta_1, \theta_2) = 0,$$

da cui le misure $P_{\theta_1}, P_{\theta_2}$ sono singolari.

Enfatizziamo il fatto che non è detto che due misure non singolari siano equivalenti. Non si ha infatti un risultato analogo a quello che si ha nel caso di variabili gaussiane, che ora enunciamo (per una dimostrazione di questo teorema, si veda [1]).

Teorema 3.2.2 (di Feldman-Hájek). Siano $\mu_1 = N_k(\mathbf{m}_1, \Gamma), \mu_2 = N_k(\mathbf{m}_2, \Gamma)$. Allora si ha che $\mu_1 \sim \mu_2$ oppure μ_1 e μ_2 sono singolari, e si ha la prima alternativa se e solo se $(\mathbf{m}_1 - \mathbf{m}_2) \in \text{Im } Q$, ove $Q = \sqrt{\Gamma} \in S(k, \mathbb{R})$ è la radice quadrata di Γ . In particolare, se Γ è invertibile, le misure sono tra loro equivalenti.

Un altro risultato che non si conserva quando si passa dagli esperimenti gaussiani a quelli poissoniani è il seguente (che è stato già dimostrato precedentemente).

Proposizione 3.2.3. Il tipo di un esperimento gaussiano \mathcal{G} è univocamente determinato dall'affinità $\rho(s, t)$.

Nonostante ciò, l'importanza che gli esperimenti poissoniani hanno in molti casi pratici è indubbia: cerchiamo di apprezzarla ora, effettuando qualche considerazione.

Definizione 3.2.4. Dato un esperimento \mathcal{E} , sia $\mathcal{E}^k = \mathcal{E} \otimes \dots \otimes \mathcal{E}$ l'esperimento ottenuto ripetendo k volte \mathcal{E} (con \mathcal{E}^0 uguale all'esperimento banale, dove tutte le probabilità sono uguali, tale che $\mathcal{H}(\mathcal{E}^0) \equiv 1$). Si definisce **esperimento poissonizzato** di \mathcal{E} l'esperimento $\mathbf{P}(\mathcal{E}) = \mathcal{E}^N$ ottenuto considerando una variabile aleatoria discreta N di legge poissoniana di parametro 1, e considerando l'esperimento \mathcal{E}^k , con $k \in \mathbb{N}$, se N assume il valore k .

Definizione 3.2.5. Sia $(\mathcal{E}_j)_{j \in J}$ una famiglia di esperimenti, con J insieme finito di indici di cardinalità ξ , e sia:

$$\mathcal{F} = \bigotimes_{j \in J} \mathcal{E}_j$$

Si definisce allora **esperimento infinitamente divisibile associato** (in breve, **esperimento associato**):

$$\mathbf{Aid}(\mathcal{F}) \stackrel{\text{def}}{=} \bigotimes_{j \in J} \mathbf{P}(\mathcal{E}_j)$$

Osservazione 34. Notiamo subito che $\mathbf{Aid}(\mathcal{F})$ è un esperimento poissoniano. Se infatti, per $j \in J$, $\phi_j(\alpha)$ è la trasformata di Hellinger di \mathcal{E}_j in α , allora la trasformata di Hellinger $\psi_j(\alpha)$ di $\mathbf{P}(\mathcal{E}_j)$ in α è:

$$\psi_j(\alpha) = \sum_{k=0}^{+\infty} \frac{e^{-1}}{k!} \phi_j(\alpha)^k = e^{-1} \sum_{k=0}^{+\infty} \frac{\phi_j(\alpha)^k}{k!} = e^{\phi_j(\alpha)-1},$$

da cui:

$$\mathcal{H}(\mathbf{Aid}(\mathcal{F}))(\alpha) = \prod_{j \in J} \psi_j(\alpha) = e^{\sum_{j \in J} \phi_j(\alpha) - \xi},$$

da cui $\mathbf{Aid}(\mathcal{F})$ ha una trasformata di Hellinger della forma vista nella formula (3.9): concludiamo che esso è un esperimento poissoniano (a meno di equivalenza, nel senso della definizione (1.2.3)).

Prima di continuare, fissiamo la notazione. Per α opportuno, siano:

- $\gamma_1(\alpha) \stackrel{\text{def}}{=} \frac{1}{2} \sup_{j \in J} [1 - \phi_j(\alpha)];$
- $\gamma_2(\alpha) \stackrel{\text{def}}{=} \frac{1}{\xi} [1 - \bar{\phi}(\alpha)]^{-2} \sum_{j \in J} [\phi_j(\alpha) - \bar{\phi}(\alpha)]^2,$

con:

$$\bar{\phi}(\alpha) = \frac{1}{\xi} \sum_{j \in J} \phi_j(\alpha),$$

(ove $\gamma_2(\alpha) = 0$, per convenzione, se $\bar{\phi}(\alpha) = 1$), e sia:

$$\gamma(\alpha) \stackrel{\text{def}}{=} \left(\gamma_1(\alpha) \wedge \frac{2}{e} \frac{1 + \gamma_2(\alpha)}{\xi} \right)$$

Proposizione 3.2.4. Con la notazione fin qui usata, per α opportuno si ha la seguente catena di disuguaglianze:

$$0 \leq \mathcal{H}(\mathbf{Aid}(\mathcal{F}))(\alpha) - \mathcal{H}(\mathcal{F})(\alpha) \leq \gamma(\alpha)$$

Dimostrazione. Sia $J = \{1, 2, \dots, \xi\}$. Notiamo che, per ogni $j \in J$, vale $0 \leq \phi_j(\alpha) \leq 1$, da cui $e^{-1} \leq \psi_j(\alpha) \leq 1$. In particolare, dato che la funzione $h : \mathbb{R} \rightarrow \mathbb{R}$ definita da $h(x) = [e^{x-1} - x]$ è sempre non negativa, concludiamo che per ogni $j \in J$ e per ogni α opportuno vale:

$$0 \leq \phi_j(\alpha) \leq \psi_j(\alpha) \leq 1$$

Volendo essere più precisi, con uno sviluppo di Taylor al second'ordine si ottiene:

$$\begin{aligned} 0 &\leq \psi_j(\alpha) - \phi_j(\alpha) = e^{\phi_j(\alpha)-1} - \phi_j(\alpha) \leq \\ &\leq 1 + \phi_j(\alpha) - 1 + \frac{1}{2}(\phi_j(\alpha) - 1)^2 - \phi_j(\alpha) = \frac{1}{2}(1 - \phi_j(\alpha))^2 \end{aligned}$$

Definiamo ora i seguenti prodotti parziali, per ogni $j \in J$ (il prodotto vuoto è da porre uguale a 1):

$$\Pi_j(\alpha) \stackrel{\text{def}}{=} \left(\prod_{k < j} \psi_k(\alpha) \right) \left(\prod_{k > j} \phi_k(\alpha) \right)$$

Notiamo che $\Pi_1(\alpha)\phi_1(\alpha) = \phi(\alpha)$ e $\Pi_\xi(\alpha)\psi_\xi(\alpha) = \psi(\alpha)$. Inoltre si ha, per $j \in J$ con $j \neq 1$, $\Pi_{j-1}(\alpha)\psi_j(\alpha) = \Pi_j(\alpha)\phi_j(\alpha)$. Allora vale la seguente uguaglianza:

$$\psi(\alpha) - \phi(\alpha) = \sum_{j \in J} \Pi_j(\alpha) [\psi_j(\alpha) - \phi_j(\alpha)]$$

Dato che $\Pi_j(\alpha) \leq \psi(\alpha)\psi_j(\alpha)^{-1}$, e che $\psi_j(\alpha) \geq e^{-1}$, si ha dunque:

$$0 \leq \psi(\alpha) - \phi(\alpha) \leq e \psi(\alpha) \sum_{j \in J} (\psi_j(\alpha) - \phi_j(\alpha)) \leq \frac{e \psi(\alpha)}{2} \sum_{j \in J} (1 - \phi_j(\alpha))^2$$

Definiamo ora:

$$\omega(\alpha) \stackrel{\text{def}}{=} \sum_{j \in J} (1 - \phi_j(\alpha))$$

Allora $\psi(\alpha) = e^{-\omega(\alpha)}$, e inoltre vale $\omega(\alpha)\psi(\alpha) = \omega(\alpha) e^{-\omega(\alpha)} \leq e^{-1}$. Dunque:

$$\sum_{j \in J} (1 - \phi_j(\alpha))^2 \leq \omega(\alpha) \sup_{j \in J} [1 - \phi_j(\alpha)] = \omega(\alpha) [2 \gamma_1(\alpha)] ,$$

da cui si ottiene la prima disuguaglianza:

$$0 \leq \psi(\alpha) - \phi(\alpha) \leq \frac{e \psi(\alpha)}{2} \sum_{j \in J} (1 - \phi_j(\alpha))^2 \leq \frac{e \psi(\alpha)}{2} \omega(\alpha) [2 \gamma_1(\alpha)] \leq \gamma_1(\alpha)$$

Per la seconda disuguaglianza, notiamo che:

$$\sum_{j \in J} (1 - \phi_j(\alpha))^2 = \sum_{j \in J} (1 - \bar{\phi}(\alpha) + \bar{\phi}(\alpha) - \phi_j(\alpha))^2 =$$

$$\begin{aligned}
&= \sum_{j \in J} (\bar{\phi}(\alpha) - \phi_j(\alpha))^2 + \sum_{j \in J} (1 - \bar{\phi}(\alpha))^2 = \xi (1 + \gamma_2(\alpha)) (1 - \bar{\phi}(\alpha))^2 = \\
&= \xi (1 + \gamma_2(\alpha)) \frac{1}{\xi^2} \omega(\alpha)^2 = \xi^{-1} (1 + \gamma_2(\alpha)) \omega(\alpha)^2,
\end{aligned}$$

e dato che $\omega(\alpha)^2 \psi(\alpha) = \omega(\alpha)^2 e^{-\omega(\alpha)} \leq 4 e^{-2}$, si ottiene:

$$\begin{aligned}
0 \leq \psi(\alpha) - \phi(\alpha) &\leq \frac{e \psi(\alpha)}{2} \sum_{j \in J} (1 - \phi_j(\alpha))^2 = \\
&= \frac{e \psi(\alpha)}{2} \frac{1}{\xi} (1 + \gamma_2(\alpha)) \omega(\alpha)^2 \leq \frac{2}{e} \frac{1 + \gamma_2(\alpha)}{\xi},
\end{aligned}$$

da cui la tesi. \square

Osservazione 35. La disuguaglianza riguardante $\gamma_1(\alpha)$ acquista maggior senso quando ciascuno degli esperimenti di cui è costituito \mathcal{F} , visto separatamente, contiene informazioni limitate, pressochè irrilevanti; la disuguaglianza riguardante $\gamma_2(\alpha)$, invece, acquista maggior senso quando \mathcal{F} consta di più ripetizioni di uno stesso esperimento, o è simile ad un esperimento di tale forma.

Sfortunatamente, dati due esperimenti $\mathcal{E}_1, \mathcal{E}_2$, non è facile trovare relazioni tra la distanza $\Delta(\mathcal{E}_1, \mathcal{E}_2)$, introdotta nella definizione (1.2.2), e la distanza uniforme delle relative trasformate di Hellinger $\mathcal{H}(\mathcal{E}_1), \mathcal{H}(\mathcal{E}_2)$. A titolo d'esempio, vediamo ora una proposizione che lega queste due grandezze, nel caso in cui Θ ha cardinalità finita.

Proposizione 3.2.5. Esiste una funzione $\omega_k : [0, +\infty[\rightarrow [0, +\infty[$, crescente, continua e infinitesima in 0 (ossia un **modulo di continuità**) tale che, dati due esperimenti $\mathcal{E}_1, \mathcal{E}_2$ aventi come comune spazio dei parametri Θ (insieme finito di cardinalità k), allora:

$$\|\mathcal{H}(\mathcal{E}_1) - \mathcal{H}(\mathcal{E}_2)\|_\infty \leq \varepsilon \Rightarrow \Delta(\mathcal{E}_1, \mathcal{E}_2) \leq \omega_k(\varepsilon)$$

Dimostrazione. La dimostrazione sarà molto simile a quella già vista per il teorema (1.3.5), e farà uso del corollario (1.3.6) e della proposizione (2.3.1).

Supponendo per assurdo che la proposizione sia falsa, esisterebbe allora una coppia di successioni di esperimenti $(\mathcal{E}_{1,n})_{n \in \mathbb{N}^+}, (\mathcal{E}_{2,n})_{n \in \mathbb{N}^+}$ e un numero reale $b > 0$ tali che:

- $\|\mathcal{H}(\mathcal{E}_{1,n}) - \mathcal{H}(\mathcal{E}_{2,n})\|_\infty \leq 2^{-n}$;
- $\Delta(\mathcal{E}_{1,n}, \mathcal{E}_{2,n}) > b$.

Per il corollario (1.3.6), a meno di estrarre sottosuccessioni possiamo supporre che le due successioni convergano rispettivamente a \mathcal{F}_1 e \mathcal{F}_2 . Le loro trasformate di Hellinger, per la proposizione (2.3.1) e per quanto detto sopra, devono coincidere: allora $\Delta(\mathcal{F}_1, \mathcal{F}_2) = 0$, e ciò è assurdo perchè dovrebbe valere $\Delta(\mathcal{F}_1, \mathcal{F}_2) \geq b$. \square

Esiste un teorema che prova all'incirca l'implicazione opposta di quest'ultima proposizione, ma la convergenza delle trasformate è solo puntuale.

Osservazione 36. Analizzando quanto visto finora, è chiaro che l'obiettivo è quello di formulare un teorema simile al teorema (3.1.3), usando però gli esperimenti poissoniani al posto degli shift gaussiani. Sfortunatamente, non è così immediato trovare condizioni sufficienti nel caso degli esperimenti poissonizzati. I primi risultati sono stati formulati da

L. Le Cam nel 1974 (si veda [5]), mentre i risultati che esporremo ora sono molto recenti (2007). Per una dimostrazione dei risultati che verranno enunciati di seguito, si veda [10].

Dato un esperimento \mathcal{E} indicizzato in Θ , sia $\mathcal{E}^n = \mathcal{E} \otimes \dots \otimes \mathcal{E}$ l'esperimento ottenuto ripetendo n volte \mathcal{E} . Presa ora una variabile aleatoria discreta N_n , di legge poissoniana di parametro n , sia $\mathcal{F}_n = \mathcal{E}^{N_n}$. Il nostro obiettivo è quello di cercare condizioni necessarie e sufficienti per cui valga:

$$\lim_{n \rightarrow +\infty} \Delta(\mathcal{E}^n, \mathcal{F}_n) = 0$$

Innanzitutto supponiamo che Θ sia un insieme di densità definite in $[0, 1]$, e che sia quindi $\mathcal{E}^n = \{P_f^n \mid f \in \Theta\}$, con $P_f^n = (f \cdot \lambda)^{\otimes n}$, con λ misura di Lebesgue su \mathbb{R} .

Data una densità f definita in $[0, 1]$, e dati due naturali positivi $j \leq k$, definiamo il seguente operatore approssimante:

$$J_{j,k}(f) \stackrel{\text{def}}{=} k \int_{\frac{j-1}{k}}^{\frac{j}{k}} f(x) dx,$$

e definiamo dunque l'approssimante $\bar{f}_{(k)}$ in questo modo:

$$\bar{f}_{(k)} \stackrel{\text{def}}{=} \sum_{j=1}^k J_{j,k}(f) \mathbb{I}_{\left[\frac{j-1}{k}, \frac{j}{k}\right[} \quad (3.10)$$

Definizione 3.2.6. Con le notazioni usate nella formula (3.10), data una densità f in $[0, 1]$, e dati α, p, q tali che $1 \leq p \leq +\infty$, $1 \leq q \leq +\infty$ e $0 < \alpha \leq 1$, si definisce **norma di Besov** di f :

$$\|f\|_{\alpha,p,q} \stackrel{\text{def}}{=} \left[\left| \int_0^1 f(x) dx \right|^q + \sum_{k=0}^{+\infty} \left(2^{k\alpha} \|\bar{f}_{(2^{k+1})} - \bar{f}_{(2^k)}\|_p \right)^q \right]^{\frac{1}{q}}$$

Definizione 3.2.7. Nelle notazioni usate finora, dato $M \geq 0$ si definisce **palla di Besov** di raggio M :

$$B_{p,q}^\alpha \stackrel{\text{def}}{=} \{f \in L^1(\mathbb{R}) \mid \|f\|_{\alpha,p,q} \leq M\}$$

Teorema 3.2.6. Sia \mathcal{E} un esperimento indicizzato in Θ , con $\Theta \subseteq B_{p,q}^\alpha(M)$, per qualche quaterna α, p, q, M con $1 \leq p, q \leq +\infty$, $M \geq 0$, $0 < \alpha \leq 1$. Supponiamo che $\alpha p < 1/2$. Allora:

$$\lim_{n \rightarrow +\infty} \Delta(\mathcal{E}^n, \mathcal{F}_n) = 0$$

Osservazione 37. Sempre in [10] viene mostrata la costruzione di un controesempio: nella fattispecie, viene mostrato un esperimento \mathcal{E} indicizzato in Θ , con $\Theta \subseteq B_{p,q}^\alpha(M)$ (ove però $\alpha p > 1/2$), tale che $\Delta(\mathcal{E}^n, \mathcal{F}_n) \not\rightarrow 0$.

Capitolo 4

Alcuni risultati preliminari

In questo capitolo, riassumeremo alcuni risultati ausiliari, che serviranno poi per comprendere il capitolo 5, dove analizzeremo una generalizzazione del Teorema Limite Centrale.

Salvo avvisi contrari, in questo capitolo parleremo di somme di variabili aleatorie indipendenti: esporremo alcune dimostrazioni elementari ma originali che saranno utili nel prossimo capitolo, utilizzando alcune idee di J.W.Lindeberg e P.Lévy. Concluderemo infine con la nozione di **differenziabilità secondo Peano**.

Finora abbiamo analizzato varie proprietà riguardanti gli esperimenti statistici, mentre ora la nostra analisi riguarderà soprattutto le variabili aleatorie reali. Ciò sarà utile nel prossimo capitolo, quando torneremo ad analizzare la particolare classe degli esperimenti statistici binari, già presentati nella sezione (2.1).

4.1 La procedura di Lindeberg

Esporremo ora la procedura usata da J.W.Lindeberg per dare una dimostrazione elementare del Teorema Limite Centrale: si veda [8] per l'articolo originale. L'idea è molto semplice, e userà sostanzialmente la formula di Taylor con resto in forma di Lagrange. Nel prossimo capitolo, poi, verrà mostrato esplicitamente che il teorema di Lindeberg, che dimostreremo in questa sezione, costituisce una generalizzazione del classico Teorema Limite Centrale.

Sia $(X_k)_{k=1}^n$ una n -upla di variabili aleatorie reali tra loro indipendenti, e sia $(Y_k)_{k=1}^n$ una n -upla dello stesso tipo, ove per ogni k da 1 a n Y_k risulta indipendente da X_1, \dots, X_n . L'idea di fondo sarà quella di considerare un'opportuna funzione $f(X_1 + \dots + X_n)$ della somma, e di sostituire le variabili X_j con le variabili Y_j : una procedura molto semplice, che sarà più chiara tra poco.

Definiamo, per $k = 1, \dots, n$ (le somme vuote sono da intendersi uguali a 0):

$$R_k \stackrel{\text{def}}{=} \sum_{j < k} X_j + \sum_{j > k} Y_j$$

Allora, data una funzione f sufficientemente regolare, si ha:

$$f\left(\sum_{j=1}^n X_j\right) - f\left(\sum_{j=1}^n Y_j\right) = \sum_{k=1}^n [f(R_k + X_k) - f(R_k + Y_k)] , \quad (4.1)$$

dato che per ogni $k = 2, \dots, n$ vale $R_{k-1} + X_{k-1} = R_k + Y_k$.

Dimostriamo ora una proposizione, la cui utilità sarà chiara tra non molto.

Proposizione 4.1.1. Supponiamo che $X_1, \dots, X_n, Y_1, \dots, Y_n$ siano variabili tra loro indipendenti, e supponiamo che per ogni $j = 1, \dots, n$ valgano le seguenti condizioni:

- $\mathbb{E}[X_j] = \mathbb{E}[Y_j] = 0$;
- $\mathbb{E}[X_j^2] = \mathbb{E}[Y_j^2] = \sigma_j^2 > 0$.

Sia poi, per ogni $j = 1, \dots, n$, $\beta_j = \mathbb{E}[|X_j|^3] + \mathbb{E}[|Y_j|^3]$.

Considerata allora una funzione $f \in C^2(\mathbb{R})$ a valori reali, tale che f'' risulti λ -Lipschitziana per qualche $\lambda \geq 0$, si ha:

$$\left| \mathbb{E}\left[f\left(\sum_{j=1}^n X_j\right)\right] - \mathbb{E}\left[f\left(\sum_{j=1}^n Y_j\right)\right] \right| \leq \frac{\lambda}{2} \sum_{j=1}^n \beta_j$$

Dimostrazione. Usando la formula di Taylor con resto di Lagrange, con centro in R_k :

- $f(R_k + X_k) = f(R_k) + X_k f'(R_k) + \frac{1}{2} X_k^2 f''(R_k) + W_{k,1}$;
- $f(R_k + Y_k) = f(R_k) + Y_k f'(R_k) + \frac{1}{2} Y_k^2 f''(R_k) + W_{k,2}$,

ove:

- $W_{k,1} = \frac{1}{2} [f''(\xi_k) - f''(R_k)] X_k^2$, con ξ_k compreso tra R_k e $R_k + X_k$;
- $W_{k,2} = \frac{1}{2} [f''(\eta_k) - f''(R_k)] Y_k^2$, con η_k compreso tra R_k e $R_k + Y_k$.

Sfruttando l'indipendenza di tutte le variabili, il valore atteso della differenza è:

$$\begin{aligned} |\mathbb{E}[f(R_k + X_k) - f(R_k + Y_k)]| &= |\mathbb{E}[W_{k,1} - W_{k,2}]| \leq \mathbb{E}[|W_{k,1}| + |W_{k,2}|] \leq \\ &\leq \frac{\lambda}{2} \mathbb{E}[X_k^2 |\xi_k - R_k| + Y_k^2 |\eta_k - R_k|] \leq \frac{\lambda}{2} \mathbb{E}[|X_k|^3 + |Y_k|^3] = \frac{\lambda}{2} \beta_j , \end{aligned}$$

da cui si ottiene la tesi sommando su k e sfruttando la formula (4.1), quindi la disuguaglianza triangolare. \square

Osservazione 38. Chiaramente la proposizione (4.1.1) ammette una naturale generalizzazione nel caso di due famiglie numerabili di variabili aleatorie tra loro indipendenti, che acquista senso quando $\sum_{j=1}^{+\infty} \beta_j < +\infty$.

Vediamo ora come sfruttare la proposizione (4.1.1): prima, però introduciamo una nuova distanza tra probabilità.

Definizione 4.1.1. Siano P, Q probabilità su \mathbb{R} , aventi funzioni di ripartizione F_P, F_Q . Si definisce **norma di Kolmogorov** della differenza $P - Q$:

$$\|P - Q\|_K \stackrel{\text{def}}{=} \|F_P - F_Q\|_\infty$$

Teorema 4.1.2 (di Lindeberg). Supponiamo che le variabili X_1, \dots, X_n siano indipendenti, centrate, dotate di momento secondo, ed equilimitate in modulo da un'opportuna costante a . Sia, per ogni $j = 1, \dots, n$, $\mathbb{E}[X_j^2] = \sigma_j^2$. Sia poi:

$$s^2 = \sum_{j=1}^n \sigma_j^2$$

Siano infine P la legge della variabile $S = X_1 + \dots + X_n$, e sia $G \sim N(0, s^2)$. Allora esiste una costante C , non dipendente da n , tale che:

$$\|P - G\|_K \leq C \left(\frac{a}{s}\right)^{\frac{1}{4}}$$

Dimostrazione. In questa dimostrazione, la proposizione (4.1.1) sarà essenziale. Sia:

$$g(x) = \int_{-\infty}^x g'(t) dt,$$

ove (si noti la presenza della funzione β di Eulero, che rende g' una densità):

$$g'(t) \stackrel{\text{def}}{=} \frac{1}{\beta(m, m)} x^{m-1} (1-x)^{m-1} \mathbb{I}_{[0,1]}$$

Allora $g(x) = 0$ per $x < 0$, e $g(x) = 1$ per $x \geq 1$. Scegliamo $m \geq 6$, così da avere le prime quattro derivate di g continue e limitate.

Operiamo ora una riscalatura, definendo $f(x) = g(\tau x)$, con $\tau > 0$: risulta evidente che $f(x) = 0$ per $x < 0$, e $f(x) = 1$ per $x \geq \tau^{-1}$.

Definiamo ora, per $j = 1, \dots, n$, $Y_j \sim N(0, \sigma_j^2)$. Se dunque $S = X_1 + \dots + X_n$, e $T = Y_1 + \dots + Y_n \sim N(0, s^2)$, si ha, per la proposizione (4.1.1):

$$\mathbb{P}(S - x > 0) \leq \mathbb{E} \left[f \left(S - x + \frac{1}{\tau} \right) \right] \leq \mathbb{E} \left[f \left(T - x + \frac{1}{\tau} \right) \right] + C_g \tau^3 a s^2,$$

ove C_g è una costante che dipende solo da g . Non riportiamo i dettagli, dicendo solo che si è sfruttato la seguente disuguaglianza (fornita implicitamente dalle ipotesi):

$$a \geq \max_{1 \leq j \leq n} \sigma_j$$

Sfruttando ora le proprietà di f , e il fatto che $T \sim N(0, s^2)$, si ha:

$$\mathbb{E} \left[f \left(T - x + \frac{1}{\tau} \right) \right] \leq \mathbb{P}(T - x > 0) + \mathbb{P} \left(-\frac{1}{\tau} \leq T - x < 0 \right) < \mathbb{P}(T - x > 0) + \frac{1}{\tau \sqrt{2\pi} s}$$

Ripetendo il ragionamento in maniera pressochè identica, si ottiene:

$$\|P - G\|_K \leq \frac{1}{\tau \sqrt{2\pi} s} + C_g \tau^3 a s^2,$$

e minimizzando il secondo membro in τ , con calcoli semplici, si ottiene la tesi (la costante finale non sarà C_g , ma sarà una costante D_g proporzionale a $(C_g)^{1/4}$). \square

Osservazione 39. Notiamo che l'ipotesi di equilimitatezza non è strettamente necessaria, in quanto si può ottenere la tesi ponendo $a = \max_{1 \leq j \leq n} \sigma_j$.

Osservazione 40. Anche il teorema (4.1.2) ammette una naturale generalizzazione nel caso numerabile, che acquista senso quando $\sum_{j=1}^{+\infty} \sigma_j^2 < +\infty$.

Osservazione 41. Il termine $(a/s)^{1/4}$ può essere migliorato di molto: C.G.Esseen, con argomenti decisamente più impegnativi, ha dimostrato che, nelle ipotesi del teorema (4.1.2), il secondo membro potrebbe essere della forma $\Gamma_g(a/s)$, con Γ_g costante universale. Per l'articolo relativo, si veda [3]. Detto ciò, sottolineiamo il fatto che la procedura di Lindeberg, oltre a essere facilmente comprensibile, si applica in molti casi dove le altre procedure falliscono, o non si applicano facilmente: ad esempio essa si applica, senza sostanziali modifiche, per variabili aleatorie a valori in uno spazio di Hilbert generico.

4.2 Lo split di Lévy

Analizziamo ora un metodo introdotto da P.Lévy, che insieme alla procedura di Lindeberg, analizzata nella scorsa sezione, consentirà di dimostrare la versione generalizzata del Teorema Limite Centrale che presenteremo nel capitolo 5.

Sia μ una probabilità su \mathbb{R} . Essa si può scrivere, in modo chiaramente non unico, come somma di due misure non banali: sarà allora $\mu = \mu_1 + \mu_2$. Se $\mu_2(\mathbb{R}) = \alpha$ (dunque $\mu_1(\mathbb{R}) = (1 - \alpha)$), possiamo definire due probabilità $P_1 = (1 - \alpha)^{-1} \cdot \mu_1$, $P_2 = \alpha^{-1} \cdot \mu_2$, e scrivere allora μ usando una combinazione convessa:

$$\mu = (1 - \alpha)P_1 + \alpha P_2$$

Il discorso è analogo nel caso di particolari variabili aleatorie. Sia ξ una variabile aleatoria discreta, portata da $A = \{0, 1\}$, con $\mathbb{P}(\xi = 1) = 1 - \mathbb{P}(\xi = 0) = \alpha$.

Siano poi $U' \sim P_1$, $V' \sim P_2$ due variabili aleatorie reali, e supponiamo che ξ, U', V' siano indipendenti. Allora è facile convincersi che:

$$X \stackrel{\text{def}}{=} (1 - \xi)U' + \xi V' \sim \mu$$

Considerata allora una successione di variabili aleatorie reali indipendenti $(X_k)_{k \in \mathbb{N}^+}$, ove per $k \in \mathbb{N}^+$ $X_k \sim \mu_k$, possiamo applicare lo split a ciascuna variabile, e scrivere:

$$X_k = (1 - \xi_k)U'_k + \xi_k V'_k,$$

ove $(\xi_k)_{k \in \mathbb{N}^+}$, $(U'_k)_{k \in \mathbb{N}^+}$, $(V'_k)_{k \in \mathbb{N}^+}$ sono successioni di variabili aleatorie reali, tra loro indipendenti.

Supponiamo che, nello split di $(X_k)_{k \in \mathbb{N}^+}$, sia ben definito ben ogni $k \in \mathbb{N}^+$ il valore atteso $m_k = \mathbb{E}[U'_k]$. Definiamo allora:

$$U_k \stackrel{\text{def}}{=} U'_k - m_k, \quad V_k \stackrel{\text{def}}{=} V'_k - m_k$$

In questo modo, si ha, per ogni $k \in \mathbb{N}^+$:

- $\mathbb{E}[U_k] = 0$;
- $X_k = m_k + (1 - \xi_k)U_k + \xi_k V_k$.

Esempio. Nelle notazioni usate fin qui, un esempio di split spesso utilizzato è il seguente. Dato $a \geq 0$:

$$X_k = (1 - \gamma_k)U'_k + \gamma_k V'_k,$$

con $\gamma_k = \mathbb{P}(|X_k| > a) \in]0, 1[$, per non avere decomposizioni banali, e:

- $U'_k \sim \mu_{k,1} = [(1 - \gamma_k)^{-1} \mathbb{I}_{[-a,a]}] \cdot \mu_k$;
- $V'_k \sim \mu_{k,2} = [\gamma_k^{-1} \mathbb{I}_{[-a,a]^c}] \cdot \mu_k$.

Notiamo che in tal caso $|U'_k| \leq a$, dunque è ben definito il valore atteso $m_k = \mathbb{E}[U'_k]$.

Osservazione 42. D'ora in poi, data una legge μ su \mathbb{R} , una variabile aleatoria $X \sim \mu$ e un insieme misurabile $A \in \mathcal{B}(\mathbb{R})$, diremo che Y ha la legge di X condizionata all'evento $X \in A$ se:

$$Y \sim \frac{1}{\mathbb{P}(X \in A)} \mathbb{I}_A \cdot \mu,$$

ove ovviamente $\mathbb{P}(X \in A) = \mu(A) > 0$. Dunque le variabili descritte precedentemente hanno, per ogni $k \in \mathbb{N}^+$, la legge di X_k condizionata rispettivamente agli eventi $|X_k| \leq a$ e $|X_k| > a$.

Introduciamo ora una nuova successione $(\eta_k)_{k \in \mathbb{N}^+}$ di variabili aleatorie reali, indipendenti tra loro e indipendenti da tutte le variabili aleatorie presentate fino ad ora, tale che per ogni $k \in \mathbb{N}^+$ vale $\eta_k \sim \xi_k$. Allora si ha:

$$\sum_{k=1}^{+\infty} (X_k - m_k) = \sum_{k=1}^{+\infty} (1 - \eta_k)U_k + \sum_{k=1}^{+\infty} \xi_k V_k + \sum_{k=1}^{+\infty} (\eta_k - \xi_k)U_k,$$

ma adesso è chiaro che abbiamo guadagnato una proprietà aggiuntiva. Al secondo membro, infatti, le prime due sommatorie sono indipendenti. In più, se per ogni $k \in \mathbb{N}^+$ definiamo $\alpha_k = \mathbb{P}(\xi_k = 1)$ e poi definiamo $\alpha = \max_{k \in \mathbb{N}^+} \alpha_k$, allora la terza sommatoria al secondo membro diventa tanto più trascurabile, rispetto alla prima sommatoria, quanto più α è piccolo in modulo. Ad essere precisi, enunciamo una proposizione: per la dimostrazione, si veda [6].

Proposizione 4.2.1. Nelle notazioni usate finora, siano:

$$S = \sum_{k=1}^{+\infty} (X_k - m_k), \quad T = \sum_{k=1}^{+\infty} [(1 - \eta_k)U_k + \xi_k V_k]$$

Siano poi P_S, P_T le probabilità immagine (su \mathbb{R}) di S, T rispettivamente. Allora:

$$\|P_S - P_T\|_K \leq 13 \alpha^{\frac{1}{3}}$$

Osservazione 43. Notiamo che la varianza della terza sommatoria è:

$$\begin{aligned} \mathbb{E} \left[\left(\sum_{k=1}^{+\infty} (\eta_k - \xi_k)U_k \right)^2 \right] &= \mathbb{E} \left[\sum_{k=1}^{+\infty} (\eta_k - \xi_k)^2 U_k^2 \right] = \mathbb{E} \left[\sum_{k=1}^{+\infty} (\alpha_k - 2\alpha_k^2 + \alpha_k) U_k^2 \right] \leq \\ &\leq 2\alpha \mathbb{E} \left[\sum_{k=1}^{+\infty} (1 - \alpha_k) U_k^2 \right] = \mathbb{E} \left[\sum_{k=1}^{+\infty} (1 - \eta_k)^2 U_k^2 \right] = \mathbb{E} \left[\left(\sum_{k=1}^{+\infty} (1 - \eta_k) U_k \right)^2 \right], \end{aligned}$$

dunque è minore di 2α volte la varianza della prima sommatoria.

Osservazione 44. È dunque chiaro, dalla proposizione (4.2.1) e dall'osservazione (43), che è conveniente cercare uno split della forma $X_k = m_k + (1 - \alpha_k) U_k + \alpha_k V_k$, con α_k piccolo in modulo.

4.3 La disuguaglianza di simmetrizzazione di Lévy

Esponiamo ora un risultato dovuto ancora a P.Lévy, dalla dimostrazione abbastanza semplice, ma dall'importanza notevole (tanto da meritare una propria sezione).

Definizione 4.3.1. Data una probabilità μ su \mathbb{R} , si definisce **mediana** di μ un numero reale m tale che $\mu([-\infty, m]) \geq \frac{1}{2}$ e $\mu([m, +\infty]) \geq \frac{1}{2}$, oppure tale che $F_\mu(m^-) \leq \frac{1}{2} \leq F_\mu(m)$, dove F_μ è la funzione di ripartizione di μ .

Con questa definizione, ogni misura sulla retta ammette almeno una mediana: se ne esiste una, questa non è a priori unica.

Teorema 4.3.1 (Disuguaglianza di simmetrizzazione, di Lévy). Sia $(X_k)_{k=1}^n$ un insieme finito di variabili aleatorie reali indipendenti, definite in uno spazio probabilizzato $(\Omega, \mathcal{A}, \mathbb{P})$. Sia, per $k = 1, \dots, n$, $S_k = X_1 + \dots + X_k$, e sia poi, per $k = 1, \dots, (n-1)$, $T_k = X_{k+1} + \dots + X_n$: supponiamo che T_1, \dots, T_{n-1} ammettano 0 come mediana.

Allora, per $a \geq 0$:

$$\mathbb{P}(\sup_{1 \leq k \leq n} |S_k| \geq a) \leq 2 \mathbb{P}(|S_n| \geq a)$$

Dimostrazione. Sia $\omega \in \Omega$ tale che $\sup_{1 \leq k \leq n} |S_k(\omega)| \geq a$, e sia j il minimo indice tale che $|S_j(\omega)| \geq a$. Supponiamo che sia $S_j(\omega) \geq a$, il caso $S_j(\omega) \leq -a$ è identico. Si ha $S_n(\omega) = S_j(\omega) + T_k(\omega)$, e dato che T_k ha 0 come mediana, la probabilità che sia $|S_n(\omega)| \geq a$ è maggiore o uguale a $\frac{1}{2}$. Di qui la tesi. \square

Osservazione 45. Il teorema (4.3.1) si applica in particolare a sequenze finite del tipo $(Z_k)_{k=1}^n = (X_k - Y_k)_{k=1}^n$, con $X_1, \dots, X_n, Y_1, \dots, Y_n$ indipendenti, e $X_k \sim Y_k$ per ogni $k = 1, \dots, n$. Per ogni $k = 1, \dots, n$, Z_k è detta **simmetrizzata** di X_k , ed in particolare ammette 0 sia come valore atteso (ovvio) che come mediana. Per $a \geq 0$, infatti, vale:

$$\mathbb{P}(Z_k \geq a) = \mathbb{P}(Z_k \leq a),$$

da cui dev'essere necessariamente $\mathbb{P}(Z_k \geq 0) = \mathbb{P}(Z_k \leq 0) = \frac{1}{2}$.

Osservazione 46. Anche in questo caso si ha un'utile generalizzazione, nel caso in cui si ha una successione $(X_k)_{k \in \mathbb{N}^+}$ di variabili aleatorie reali indipendenti. Se per ogni $k \in \mathbb{N}^+$ si ha infatti $\mathbb{P}(|X_k| \geq \delta) < \gamma$, allora:

$$\mathbb{P}\left(\sup_{k \in \mathbb{N}^+} |X_k| \geq \delta\right) < 2\gamma$$

4.4 Condizioni per la shift-compattezza

Innanzitutto, per parlare di compattezza di una successione di probabilità $(P_n)_{n \in \mathbb{N}^+}$ o di convergenza ad una data probabilità P , dobbiamo introdurre una topologia: useremo allora la topologia indotta dalla classica **convergenza stretta**. Ricordiamo che si dice che $(P_n)_{n \in \mathbb{N}^+} \rightarrow P$ strettamente se per ogni funzione f continua e limitata si ha:

$$\lim_{n \rightarrow +\infty} \int_{\Omega} f dP_n = \int_{\Omega} f dP$$

Notiamo subito che la stessa topologia si può indurre usando la norma dual-Lipschitziana (si veda la definizione (1.3.4)), o la distanza di Lévy-Prohorov: abbiamo dimostrato l'equivalenza di tali norme, senza dimostrarla, nella sezione (1.4).

Sia ora, per ogni $n \in \mathbb{N}^+$, $(X_{n,j})_{j \in \mathbb{N}^+}$ una successione di variabili aleatorie reali, con la condizione che, per ogni $n \in \mathbb{N}^+$, la successione $(X_{n,j})_{j \in \mathbb{N}^+}$ è formata da variabili aleatorie indipendenti. Sia quindi $(S_n)_{n \in \mathbb{N}^+}$ la successione definita da:

$$S_n \stackrel{\text{def}}{=} \sum_{j=1}^{+\infty} X_{n,j}$$

Definizione 4.4.1. Sia $(X_n)_{n \in \mathbb{N}^+}$ una successione di variabili aleatorie. Si dice che la successione delle probabilità immagine di tali variabili è **shift-compatta** (o con abuso di notazione, che la successione delle variabili aleatorie lo è) se esiste una successione $(b_n)_{n \in \mathbb{N}^+} \subseteq \mathbb{R}$ tale che la successione delle probabilità immagine delle variabili aleatorie $Y_n = X_n - b_n$, con $n \in \mathbb{N}^+$, è relativamente compatta rispetto alla convergenza stretta, ossia è tesa (per il noto teorema di Prohorov).

Proposizione 4.4.1. Nelle notazioni precedenti, $(S_n)_{n \in \mathbb{N}^+}$ è shift-compatta se e solo se $(S_n - S'_n)_{n \in \mathbb{N}^+}$ (ove S'_n è una variabile aleatoria avente stessa legge di S_n e indipendente da essa) è relativamente compatta.

Dimostrazione. Dimostriamo la prima implicazione: sia dunque $(S_n)_{n \in \mathbb{N}^+}$ shift-compatta, e sia $(b_n)_{n \in \mathbb{N}^+}$ una successione di costanti tale che $(S_n - b_n)_{n \in \mathbb{N}^+}$ risulti una famiglia tesa. Allora per ogni $\varepsilon > 0$ esiste una costante $K_\varepsilon \geq 0$ tale che per ogni $n \in \mathbb{N}^+$:

$$\mathbb{P}(|S_n - b_n| > K_\varepsilon) \leq \varepsilon$$

Dato che:

$$\mathbb{P}(|S_n - S'_n| > M) = 2 \mathbb{P}(S_n - S'_n > M),$$

se per ogni $\varepsilon > 0$ scegliamo $M_\varepsilon = 2 K_\varepsilon$, si ha, per ogni $n \in \mathbb{N}^+$:

$$\mathbb{P}(S_n - S'_n > M_\varepsilon) \leq \mathbb{P}(S'_n < b_n - K_\varepsilon) + \mathbb{P}(S_n > b_n + K_\varepsilon) \leq \varepsilon + \varepsilon = 2\varepsilon,$$

da cui $\mathbb{P}(|S_n - S'_n| > M_\varepsilon) \leq 4\varepsilon$. Dunque $(S_n - S'_n)_{n \in \mathbb{N}^+}$ è tesa.

Dimostriamo ora la seconda implicazione. Sia $(S_n - S'_n)_{n \in \mathbb{N}^+}$ tesa, e sia $(\mu_n)_{n \in \mathbb{N}^+}$ tale che, per ogni $n \in \mathbb{N}^+$, μ_n è una mediana di S_n . Per ipotesi si ha che per ogni $\varepsilon > 0$ esiste una costante $D_\varepsilon \geq 0$ tale che per ogni $n \in \mathbb{N}^+$:

$$\mathbb{P}(|S_n - S'_n| > D_\varepsilon) \leq \varepsilon$$

Sfruttando le idee offerte dal teorema (4.3.1), si ha:

$$\mathbb{P}(|S_n - \mu_n| > D_\varepsilon) = \mathbb{P}(S_n - \mu_n > D_\varepsilon) + \mathbb{P}(S_n - \mu_n < -D_\varepsilon) ,$$

ove:

- $\mathbb{P}(S_n > D_\varepsilon + \mu_n) \leq 2 \mathbb{P}(S_n - S'_n > D_\varepsilon) \leq 2\varepsilon;$
- $\mathbb{P}(S_n < -D_\varepsilon + \mu_n) \leq 2 \mathbb{P}(S_n - S'_n < D_\varepsilon) \leq 2\varepsilon.$

Dunque $\mathbb{P}(|S_n - \mu_n| > D_\varepsilon) \leq 4\varepsilon$: $(S_n - \mu_n)_{n \in \mathbb{N}^+}$ è quindi tesa, da cui $(S_n)_{n \in \mathbb{N}^+}$ è shift-compatta. \square

Osservazione 47. Se $(S_n)_{n \in \mathbb{N}^+}$ è shift-compatta, sfruttando la proposizione (4.4.1) possiamo considerare la successione di costanti $(a_n)_{n \in \mathbb{N}^+}$, tale che per ogni $n \in \mathbb{N}^+$ a_n sia una mediana di S_n .

Sfruttando la proposizione (4.4.1), daremo ora condizioni necessarie e sufficienti per la relativa compattezza, nel caso di variabili $(X_{n,j})_{n,j \in \mathbb{N}^+}$ simmetriche.

Proposizione 4.4.2. Sia $(X_{n,j})_{n,j \in \mathbb{N}^+}$ un insieme numerabile di variabili aleatorie reali simmetriche e indipendenti. Sia poi, per $n \in \mathbb{N}^+$:

$$S_n = \sum_{j=1}^{+\infty} X_{n,j}$$

Allora $(S_n)_{n \in \mathbb{N}^+}$ forma una successione relativamente compatta se e solo se valgono le seguenti due condizioni:

- Per $\lambda > 0$, definiamo $X_{n,j,\lambda} = X_{n,j} \mathbb{I}_{(|X_{n,j}| < \lambda)}$. Allora, se $\lambda < +\infty$, esiste $b(\lambda) < +\infty$ tale che per ogni $n \in \mathbb{N}^+$:

$$\sum_{j=1}^{+\infty} \text{Var}(X_{n,j,\lambda}) \leq b(\lambda) ;$$

- Per ogni $\varepsilon > 0$ esiste $\Xi_\varepsilon < +\infty$ tale che, per ogni $n \in \mathbb{N}^+$:

$$\sum_{j=1}^{+\infty} \mathbb{P}(|X_{n,j}| \leq \Xi_\varepsilon) \leq \varepsilon .$$

Dimostrazione. Dimostriamo innanzitutto che la seconda condizione è necessaria. Dati $n, k \in \mathbb{N}^+$, definiamo:

$$S_{n,k} \stackrel{\text{def}}{=} \sum_{j=1}^k X_{n,j}$$

Per la relativa compattezza (data dall'ipotesi), per ogni $\varepsilon > 0$ esiste $A > 0$ tale che per ogni $n \in \mathbb{N}^+$ vale $\mathbb{P}(|S_n| \geq \frac{A}{2}) < \frac{\varepsilon}{2}$.

Per la proposizione (4.3.1) e l'osservazione (46), si ha allora:

$$\mathbb{P}\left(\sup_{k \in \mathbb{N}^+} |S_{n,k}| \geq \frac{A}{2}\right) < \varepsilon ,$$

da cui, essendo $X_{n,k} = S_{n,k} - S_{n,k-1}$:

$$\mathbb{P} \left(\sup_{k \in \mathbb{N}^+} |X_{n,k}| \geq A \right) < \varepsilon$$

Per l'indipendenza delle variabili, si ha ora:

$$\begin{aligned} 1 - \varepsilon &< 1 - \mathbb{P} \left(\sup_{k \in \mathbb{N}^+} |X_{n,k}| \geq A \right) = \mathbb{P} \left(\sup_{k \in \mathbb{N}^+} |X_{n,k}| < A \right) = \\ &= \prod_{k=1}^{+\infty} \mathbb{P}(|X_{n,k}| < A) = \prod_{k=1}^{+\infty} (1 - \mathbb{P}(|X_{n,k}| \geq A)) \leq e^{-\sum_{k=1}^{+\infty} \mathbb{P}(|X_{n,k}| \geq A)}, \end{aligned}$$

da cui, se $\varepsilon > 0$ è sufficientemente piccolo in modulo:

$$\sum_{k=1}^{+\infty} \mathbb{P}(|X_{n,k}| \geq A) < -\ln(1 - \varepsilon) = \varepsilon + o(\varepsilon) \leq 2\varepsilon$$

Per dimostrare che anche la prima condizione è necessaria, sia $\varepsilon_0 \leq 10^{-10}$ molto piccolo, e sia C la costante data dalla seconda condizione, vera per quanto appena dimostrato. Supponiamo per assurdo che esista $b > 0$ (senza perdere di generalità si può supporre che sia $b > C$) tale che:

$$\sup_{n \in \mathbb{N}^+} \left(\sum_{j=1}^{+\infty} \text{Var}(X_{n,j,b}) \right) = \sup_{n \in \mathbb{N}^+} \sigma_n^2 = +\infty,$$

ove per $n \in \mathbb{N}^+$ σ_n^2 è la varianza di $\sum_{j=1}^{+\infty} X_{n,j,b}$.

Eventualmente estraendo una sottosuccessione, si può supporre che la successione delle varianze sia monotona crescente, e tendente a $+\infty$. Se per ogni $n \in \mathbb{N}^+$ chiamiamo P_n la probabilità immagine di $\sum_{j=1}^{+\infty} X_{n,j,b}$, e definiamo $G_n = N(0, \sigma_n^2)$, per la proposizione (4.1.2) si ha, per un'opportuna costante Γ non dipendente da n :

$$0 \leq \lim_{n \rightarrow +\infty} \|P_n - G_n\|_K \leq \lim_{n \rightarrow +\infty} \Gamma \left(\frac{b}{\sigma_n} \right)^{\frac{1}{4}} = 0,$$

quindi:

$$\lim_{n \rightarrow +\infty} \|P_n - G_n\|_K = 0,$$

da cui si può concludere. Infatti, dato che $(G_n)_{n \in \mathbb{N}^+}$ non è tesa, non è tesa nemmeno $(P_n)_{n \in \mathbb{N}^+}$, e tale successione differisce da quella dipartenza solamente per termini che valgono 0 con probabilità maggiore di $(1 - 10^{-10})$.

Supponiamo ora che valgano le due condizioni. Innanzitutto, per ogni $\alpha, \beta > 0$ e per ogni $n \in \mathbb{N}^+$:

$$\mathbb{P}(|S_n| \geq 2\alpha) \leq \mathbb{P} \left(\left| \sum_{j=1}^{+\infty} X_{n,j,\beta} \right| \geq \alpha \right) + \mathbb{P} \left(\left| \sum_{j=1}^{+\infty} X_{n,j} \mathbb{I}_{(|X_{n,j}| > \beta)} \right| \geq \alpha \right)$$

Consideriamo il secondo addendo al secondo membro: certamente quella sommatoria vale zero, se ogni funzione indicatrice vale 0.

Dalla seconda condizione, se $M_1 > 0$ è tale che $\sum_{j=1}^{+\infty} \mathbb{P}(|X_{n,j}| \geq M_1) \leq \varepsilon/2$, allora:

$$\begin{aligned} \mathbb{P}\left(\sup_{j \in \mathbb{N}^+} \mathbb{I}_{(|X_{n,j}| \geq M_1)} = 0\right) &= 1 - \mathbb{P}\left(\sup_{j \in \mathbb{N}^+} \mathbb{I}_{(|X_{n,j}| \geq M_1)} > 0\right) \geq \\ &\geq 1 - \sum_{j=1}^{+\infty} \mathbb{P}(|X_{n,j}| > M_1) \geq 1 - \frac{\varepsilon}{2}, \end{aligned}$$

da cui, per ogni $n \in \mathbb{N}^+$ e per ogni $\alpha > 0$:

$$\mathbb{P}\left(\left|\sum_{j=1}^{+\infty} X_{n,j} \mathbb{I}_{(|X_{n,j}| > M_1)}\right| \geq \alpha\right) \leq \frac{\varepsilon}{2}$$

Consideriamo ora il secondo addendo al secondo membro. Usando la disuguaglianza di Chebychev e la prima condizione, dato $M_2 > 0$ tale che:

$$\frac{b^2(M_1)}{M_2} \leq \frac{\varepsilon}{2},$$

si ha per ogni $n \in \mathbb{N}^+$, ponendo $\alpha = M_2$:

$$\mathbb{P}\left(\left|\sum_{j=1}^{+\infty} X_{n,j, M_1}\right| \geq M_2\right) \leq \frac{\varepsilon}{2},$$

da cui, per ogni $n \in \mathbb{N}^+$:

$$\mathbb{P}(|S_n| \geq 2M_2) \leq \varepsilon,$$

da cui $(S_n)_{n \in \mathbb{N}^+}$ è tesa, cioè relativamente compatta. \square

Osservazione 48. Osserviamo che se $(S_n)_{n \in \mathbb{N}^+}$ è relativamente compatta, per ogni $\tau > 0$ si ha che la seguente quantità:

$$\sum_{j=1}^{+\infty} \mathbb{P}(|X_{n,j}| \geq \tau)$$

è limitata da una costante $d(\tau)$, indipendente da n (quest'osservazione tornerà utile nella dimostrazione del Teorema Limite Centrale).

Per dimostrare ciò, consideriamo $\varepsilon > 0$ e Ξ_ε come nella proposizione (4.4.2), e notiamo che se $\tau \geq \Xi_\varepsilon$ la tesi è banale. In caso contrario, si ha che:

$$\mathbb{P}(|X_{n,j}| \geq \tau) = \mathbb{P}(|X_{n,j}| \geq \Xi_\varepsilon) + \mathbb{P}(\tau \leq |X_{n,j}| < \Xi_\varepsilon),$$

e che, per la disuguaglianza di Chebichev:

$$\mathbb{P}(\tau \leq |X_{n,j}| < \Xi_\varepsilon) = \mathbb{P}(|X_{n,j, \Xi_\varepsilon}| \geq \tau) \leq \frac{\mathbb{E}[X_{n,j, \Xi_\varepsilon}^2]}{\tau^2} = \frac{\text{Var}(X_{n,j, \Xi_\varepsilon})}{\tau^2},$$

essendo le variabili simmetriche. Sommando si ottiene allora:

$$\sum_{j=1}^{+\infty} \mathbb{P}(|X_{n,j}| \geq \tau) \leq \varepsilon + \frac{b(\tau)}{\tau^2}$$

Osservazione 49. Osserviamo che nella proposizione (4.4.2), a partire dall'ipotesi di relativa compattezza, per dimostrare la seconda condizione abbiamo in effetti usato il fatto che le variabili $(X_{n,j})_{n,j \in \mathbb{N}^+}$ ammettono 0 come mediana, mentre per dimostrare la prima condizione abbiamo usato il fatto che le variabili $(X_{n,j,a})_{n,j \in \mathbb{N}^+}$, per a fissato, sono centrate. Queste due ipotesi sono soddisfatte nel caso di variabili simmetriche.

4.5 Differenziabilità secondo Peano e secondo Fréchet

Sia $f : S \subseteq \mathbb{R} \rightarrow \tilde{\mathbb{R}}$ una funzione (supponiamo che $0 \in \overset{\circ}{S}$).

Definizione 4.5.1. La funzione f si dice **differenziabile una volta secondo Peano** in 0 se:

$$f(x) = f(0) + ax + xR_1(x) ,$$

con $\lim_{x \rightarrow 0} R_1(x) = 0$.

Definizione 4.5.2. La funzione f si dice **differenziabile due volte secondo Peano** in 0 se:

$$f(x) = f(0) + ax + \frac{1}{2}bx^2 + x^2R_2(x) ,$$

con $\lim_{x \rightarrow 0} R_2(x) = 0$.

Siano ora $(\mathbb{E}, \|\cdot\|_{\mathbb{E}})$, $(\mathbb{F}, \|\cdot\|_{\mathbb{F}})$ spazi di Banach, per semplicità; sia $f : \mathbb{E} \rightarrow \mathbb{F}$.

Definizione 4.5.3. La funzione f si dice **differenziabile secondo Fréchet** in 0 se esiste un'applicazione lineare e continua $A : \mathbb{E} \rightarrow \mathbb{F}$ tale che:

$$f(x) = f(0) + Ax + \|x\|R(x) ,$$

con $\lim_{\|x\| \rightarrow 0} R(x) = 0$.

Con queste definizioni, risulta chiaro che la definizione di differenziabilità secondo Peano è l'adattamento su \mathbb{R} della versione di differenziabilità secondo Fréchet.

D'ora in poi indicheremo con $P[0, 2]$ l'insieme delle funzioni f differenziabili secondo Peano due volte, con $f(0) = 0$. Salvo avvisi contrari, indicheremo con a il differenziale primo e con b il differenziale secondo, intesi secondo Peano, di f .

Proposizione 4.5.1. Sia $(X_{n,j})_{n,j \in \mathbb{N}^+}$ una famiglia di variabili aleatorie reali, e sia, per ogni $n \in \mathbb{N}^+$, $S_n = \sum_{j=1}^{+\infty} X_{n,j}^2$. Supponiamo che la famiglia $(S_n)_{n \in \mathbb{N}^+}$ sia tesa, e che:

$$\left(\sup_{j \in \mathbb{N}^+} |X_{n,j}| \right) \xrightarrow{\mathbb{P}} 0$$

Per $f \in P[0, 2]$, allora, si ha che:

$$\left(\sum_{j=1}^{+\infty} f(X_{n,j}) - \sum_{j=1}^{+\infty} \left[aX_{n,j} + \frac{1}{2}bX_{n,j}^2 \right] \right) \xrightarrow{\mathbb{P}} 0$$

Dimostrazione. La differenza è uguale a $\sum_{j=1}^{+\infty} X_{n,j}^2 R_2(X_{n,j})$, in modulo minore di:

$$\left[\sup_{j \in \mathbb{N}^+} |R_2(X_{n,j})| \right] \sum_{j=1}^{+\infty} X_{n,j}^2$$

Dato che $R(x) \rightarrow 0$ quando $x \rightarrow 0$, è facile dimostrare che tale resto tende a 0 in probabilità, sfruttando le varie ipotesi. \square

Enunciamo ora una sorta di converso della proposizione (4.5.1), sotto ipotesi aggiuntive.

Proposizione 4.5.2. Sia $g \in P[0, 2]$ una funzione non singolare, ossia con differenziale primo $a \neq 0$ e tale che per ogni $\varepsilon > 0$ vale:

$$\inf_{|y| \geq \varepsilon} |g(y)| > 0$$

Allora se $(X_{n,j})_{n,j \in \mathbb{N}^+} = (g(Y_{n,j}))_{n,j \in \mathbb{N}^+}$ è una famiglia che rispetta le due condizioni viste nella proposizione (4.5.1), ossia:

- $\left(\sum_{j \in \mathbb{N}^+} X_{n,j}^2 \right)_{n \in \mathbb{N}^+}$ è una famiglia tesa;
- $\left(\sup_{j \in \mathbb{N}^+} |X_{n,j}| \right)_{n \in \mathbb{N}^+} \xrightarrow{\mathbb{P}} 0$,

allora:

- $\left(\sum_{j \in \mathbb{N}^+} Y_{n,j}^2 \right)_{n \in \mathbb{N}^+}$ è una famiglia tesa;
- $\left(\sup_{j \in \mathbb{N}^+} |Y_{n,j}| \right)_{n \in \mathbb{N}^+} \xrightarrow{\mathbb{P}} 0$.

Capitolo 5

Un Teorema Limite Centrale per “array” infinitesimali

In questo capitolo, l'ultimo, useremo tutto quanto visto precedentemente per presentare una generalizzazione del Teorema Limite Centrale. Tale risultato verrà ancora ampliato nella sezione conclusiva di questa tesi, non prima di aver presentato tre lemmi tecnici, ma dimostrabili abbastanza facilmente, noti come **i tre lemmi di Le Cam**.

Vedremo poi, molto brevemente, come tutto quanto visto negli ultimi due capitoli si applica nel caso di successioni di esperimenti binari, ossia tali che l'insieme dei parametri Θ ha cardinalità 2.

5.1 Teorema Limite Centrale: una prima generalizzazione

Siamo ora pronti per enunciare e dimostrare una versione del Teorema Limite Centrale decisamente più generale rispetto a quelle più note. In particolare, ad eccezione della versione che fa uso della cosiddetta **condizione di Lindeberg** (su cui non ci soffermiamo; si veda [9] se si è interessati), tutte le versioni più note richiedono che le variabili aleatorie siano tra loro indipendenti ma soprattutto identicamente distribuite. Quest'ipotesi, molto stringente, verrà decisamente indebolita nella formulazione che verrà presentata tra poco.

Come nella sezione (4.4), sia $(X_{n,j})_{n,j \in \mathbb{N}^+}$ una famiglia di variabili aleatorie reali, tali che per ogni $n \in \mathbb{N}^+$ $(X_{n,j})_{j \in \mathbb{N}^+}$ è formata da variabili indipendenti. In generale, chiameremo “array doppio”, o più semplicemente “array”, una successione di successioni di variabili aleatorie, come ad esempio $(X_{n,j})_{n,j \in \mathbb{N}^+}$.

Definizione 5.1.1. Nelle notazioni appena introdotte, $(X_{n,j})_{n,j \in \mathbb{N}^+}$ si dice **uniformemente asintoticamente trascurabile** se per ogni $\varepsilon > 0$:

$$\lim_{n \rightarrow +\infty} \sup_{j \in \mathbb{N}^+} \mathbb{P}(|X_{n,j}| > \varepsilon) = 0$$

Chiameremo **array infinitesimale** una famiglia di variabili aleatorie reali $(X_{n,j})_{n,j \in \mathbb{N}^+}$ che sia uniformemente asintoticamente trascurabile.

Cominciamo enunciando una proposizione, che discende dalla proposizione (4.4.2) e che tornerà utile in futuro: per la dimostrazione, si veda [6].

Proposizione 5.1.1. Sia $(X_{n,j})_{n,j \in \mathbb{N}^+}$ un array infinitesimale, e supponiamo che la successione $(S_n)_{n \in \mathbb{N}^+}$ sia shift-compatta; sia poi $\tau > 0$. Per ogni $n, j \in \mathbb{N}^+$, sia:

$$X_{n,j} = (1 - \xi_{n,j})U_{n,j} + \xi_{n,j}V_{n,j} ,$$

ove $U_{n,j}$ è la legge di $X_{n,j}$ condizionata all'evento $|X_{n,j}| < \tau$, mentre $V_{n,j}$ è la legge di $X_{n,j}$ condizionata all'evento $|X_{n,j}| \geq \tau$, e $\xi_{n,j} \sim \mathbb{I}_{(|X_{n,j}| \geq \tau)}$. Allora:

- $\lim_{n \rightarrow +\infty} \sup_{j \in \mathbb{N}^+} \mathbb{E}[\xi_{n,j}] = \lim_{n \rightarrow +\infty} \sup_{j \in \mathbb{N}^+} \mathbb{P}(|X_{n,j}| \geq \tau) = 0$;
- La quantità $\sum_{j=1}^{+\infty} \mathbb{E}[\xi_{n,j}] = \sum_{j=1}^{+\infty} \mathbb{P}(|X_{n,j}| \geq \tau)$ è limitata da una costante $h(\tau)$ indipendente da n .

Consideriamo dunque un array infinitesimale, e sia $\tau > 0$. Definiamo allora, per $n, j \in \mathbb{N}^+$:

$$Y_{n,j} \stackrel{\text{def}}{=} X_{n,j} - m_{n,j} , \quad (5.1)$$

ove $m_{n,j} = \mathbb{E}[U'_{n,j}]$, con $U'_{n,j}$ avente la legge di $X_{n,j}$ condizionata all'evento $|X_{n,j}| < \tau$.

Sia poi $G_{n,j}$ la misura di probabilità su \mathbb{R} tale che $Y_{n,j} \sim G_{n,j}$. Sia infine:

$$L_n \stackrel{\text{def}}{=} \sum_{j=1}^{+\infty} G_{n,j} ,$$

dove però la massa del punto 0 è posta uguale a 0.

Definizione 5.1.2. Data una probabilità μ sulla retta reale, si definisce **versione poissonizzata** di μ la legge $\pi(\mu)$ della variabile aleatoria $\sum_{j=1}^N \xi_j$, dove le variabili della successione $(\xi_j)_{j \in \mathbb{N}^+}$ sono indipendenti e identicamente distribuite, con legge μ , e N è una variabile aleatoria discreta, indipendente da tutte le altre, con legge poissoniana di parametro 1 (la somma vale 0 se N assume il valore 0).

Convienne però esprimere in maniera diversa tale definizione: introduciamo allora un nuovo operatore.

Definizione 5.1.3. Siano μ, ν due probabilità definite su $\mathcal{B}(\mathbb{R}^k)$, e siano $X \sim \mu, Y \sim \nu$ indipendenti. Si definisce **convoluzione** di μ e ν la legge di probabilità di $X + Y$: essa si indica con l'usuale notazione $\mu * \nu$. Per ogni $A \in \mathcal{B}(\mathbb{R}^k)$, allora:

$$(\mu * \nu)(A) = \int_{\mathbb{R}^k} \int_{\mathbb{R}^k} \mathbb{I}_A(x + y) d\mu(x) d\nu(y)$$

La definizione è analoga nel caso di misure finite con segno. Indicheremo con μ^2 la convoluzione $\mu * \mu$, e con μ^k la k -esima convoluzione. Infine, sarà $\mu^0 = \mathcal{I}$, dove \mathcal{I} è la misura discreta che conferisce massa unitaria a 0 (elemento neutro per la convoluzione appena definita).

Definizione 5.1.4. Con la notazione usata finora, data una probabilità μ sulla retta (ma la definizione è analoga, in generale, per misure finite con segno), si definisce **convoluzione esponenziale** di μ la seguente misura:

$$e^\mu \stackrel{\text{def}}{=} \sum_{k=0}^{+\infty} \frac{\mu^k}{k!}$$

Osservazione 50. Notiamo subito che, considerate due probabilità μ, ν sulla retta, si ha:

$$e^{\mu+\nu} = \sum_{k=0}^{+\infty} \frac{(\mu+\nu)^k}{k!} = \sum_{k=0}^{+\infty} \sum_{j=0}^k \binom{k}{j} \frac{\mu^j * \nu^{k-j}}{k!} = \sum_{j=0}^{+\infty} \frac{\mu^j}{j!} * \left(\sum_{k=j}^{+\infty} \frac{\nu^{k-j}}{(k-j)!} \right) = e^\mu * e^\nu$$

Con questa notazione, per ogni $n, j \in \mathbb{N}^+$ la versione poissonizzata di $G_{n,j}$ è:

$$\pi(G_{n,j}) = \sum_{k=0}^{+\infty} e^{-1} \frac{G_{n,j}^k}{k!} = e^{-1} e^{G_{n,j}} = e^{G_{n,j} - \mathcal{I}},$$

essendo, per una probabilità μ qualsiasi, $e^{\mu - a\mathcal{I}} = e^\mu * e^{-a\mathcal{I}} = e^{-a} e^\mu$. D'ora in poi, quindi, scriveremo con abuso di notazione $e^{\mu - a}$ in favore di $e^{\mu - a\mathcal{I}}$, quando ciò non creerà confusione.

Se $G_{n,j}$ ha massa $\gamma > 0$ in 0, definiamo $G'_{n,j} = G_{n,j} - \gamma\mathcal{I}$. Allora:

$$G_{n,j} - \mathcal{I} = G'_{n,j} + \gamma\mathcal{I} - \mathcal{I} = G'_{n,j} - \|G'_{n,j}\|\mathcal{I},$$

ove per una misura μ su \mathbb{R} vale $\|\mu\| = \mu(\mathbb{R})$: si può allora supporre che sia $G_{n,j}(\{0\}) = 0$.

Con la notazione finora introdotta, (adattando la definizione (3.2.5)), la legge infinitamente divisibile associata alla successione $(Y_{n,j})_{j \in \mathbb{N}^+}$ è la legge della variabile aleatoria:

$$\mathbf{W}_n = \sum_{j=1}^{+\infty} W_{n,j},$$

ove $W_{n,j} \sim \pi(G_{n,j})$ per ogni $n, j \in \mathbb{N}^+$. Tale legge, che indicheremo con $\mathbf{Aid}(n)$, è pari a:

$$\mathbf{Aid}(n) = \bigstar_{j=1}^{+\infty} \pi(G_{n,j}) = e^{\sum_{j=1}^{+\infty} [G_{n,j} - \mathcal{I}]} = e^{L_n - \|L_n\|}, \quad (5.2)$$

dove L_n è detta **misura di Lévy di $\mathbf{Aid}(n)$** .

Prima di esporre il principale risultato di questa sezione, enunciamo un paio di risultati preliminari: per le relative dimostrazioni, e per ulteriori dettagli a riguardo, si vedano [6] e [7].

Proposizione 5.1.2. Sia $(X_j)_{j=1}^n$ un insieme finito di variabili aleatorie reali indipendenti centrate, e siano per ogni $j = 1, \dots, n$ F_j la legge di X_j , $\sigma_j^2 = \mathbb{E}[X_j^2]$, Y_j la versione poissonizzata di X_j , avente legge $e^{F_j - 1}$. Siano poi:

$$S = \sum_{j=1}^n X_j, \quad T = \sum_{j=1}^n Y_j$$

Si supponga l'esistenza di una costante τ tale che $|X_j| \leq \tau$ per ogni $j = 1, \dots, n$.

Allora, per una funzione $f \in C^4(\mathbb{R})$, avente la derivata quarta limitata da una costante B , si ha:

$$|\mathbb{E}[f(S)] - \mathbb{E}[f(T)]| \leq B\tau^2 s^2,$$

ove $s^2 = \sigma_1^2 + \dots + \sigma_n^2$.

Osservazione 51. Usando la proposizione (5.1.2) si può ottenere una limitazione in norma di Kolmogorov, con un procedimento simile a quello visto nel teorema (4.1.2): non diamo i dettagli.

Proposizione 5.1.3. Siano X_1, \dots, X_n delle variabili aleatorie reali indipendenti, di leggi P_1, \dots, P_n , e sia S la legge della somma, ottenuta mediante convoluzione delle singole leggi. Siano poi $\pi(P_1), \dots, \pi(P_n)$ le versioni poissonizzate delle singole leggi, e sia T la legge della somma, ottenuta mediante convoluzione delle singole versioni poissonizzate (nelle notazioni usate prima, la legge infinitamente divisibile associata a $(X_j)_{j=1}^n$). Se per ogni $j = 1, \dots, n$ definiamo $\alpha_j = \mathbb{P}(X_j \neq 0)$, allora:

$$\|S - T\|_{L^1} \leq 2 \sum_{j=1}^n \alpha_j^2$$

Siamo ora pronti per esporre questa generalizzazione del Teorema Limite Centrale.

Teorema 5.1.4 (Limite Centrale, generalizzazione). Sia $(X_{n,j})_{n,j \in \mathbb{N}^+}$ un array infinitesimale, e supponiamo che $(S_n)_{n \in \mathbb{N}^+}$ sia shift-compatta (ove $S_n = X_{n,1} + X_{n,2} + \dots$). Sia $\tau > 0$. Definite per ogni $n \in \mathbb{N}^+$ le seguenti distribuzioni:

- $\mathbf{G}_n = \ast_{j=1}^{+\infty} G_{n,j}$, legge di $\mathbf{Y}_n = \sum_{j=1}^{+\infty} Y_{n,j}$, ove ogni variabile $Y_{n,j}$ è definita mediante la formula (5.1);
- $\mathbf{Aid}(n) = e^{L_n - \|L_n\|}$, come nella formula (5.2),

allora:

$$\lim_{n \rightarrow +\infty} \|\mathbf{G}_n - \mathbf{Aid}(n)\|_D = 0$$

Dimostrazione. Per ogni $n, j \in \mathbb{N}^+$ sia:

$$X_{n,j} = (1 - \xi_{n,j})U'_{n,j} + \xi_{n,j}V'_{n,j},$$

ove (le variabili aleatorie sono tutte tra loro indipendenti):

- $U'_{n,j}$ ha la legge di $X_{n,j}$ condizionata all'evento $|X_{n,j}| < \tau$;
- $V'_{n,j}$ ha la legge di $X_{n,j}$ condizionata all'evento $|X_{n,j}| \geq \tau$;
- $\xi_{n,j} \sim \mathbb{I}_{(|X_{n,j}| \geq \tau)}$.

Sia poi $m_{n,j} = \mathbb{E}[U'_{n,j}]$. Allora si ha:

$$\begin{aligned} Y_{n,j} &= X_{n,j} - m_{n,j} = (1 - \xi_{n,j})(U'_{n,j} - m_{n,j}) + \xi_{n,j}(V'_{n,j} - m_{n,j}) = \\ &= (1 - \xi_{n,j})U_{n,j} + \xi_{n,j}V_{n,j} = (1 - \eta_{n,j})U_{n,j} + \xi_{n,j}V_{n,j} + (\eta_{n,j} - \xi_{n,j})U_{n,j}, \end{aligned}$$

con $U_{n,j} = U'_{n,j} - m_{n,j}$, $V_{n,j} = V'_{n,j} - m_{n,j}$, $\eta_{n,j} \sim \xi_{n,j}$ indipendente da tutte le altre variabili presenti nella formula. In particolare, notiamo che, per costruzione, $\mathbb{E}[U_{n,j}] = 0$.

Dato che l'array è infinitesimale, eventualmente eliminando le variabili $X_{n,j}$, con $n \leq p$ (per un p opportuno), possiamo supporre che sia, per ogni $n, j \in \mathbb{N}^+$, $\mathbb{P}(|X_{n,j}| < \tau) \geq 1/2$. Allora, dato che $(S_n)_{n \in \mathbb{N}^+}$ è shift-compatta, adattando opportunamente la proposizione (4.4.2) si ha (per la prima condizione) che esiste una costante $b(\tau)$ tale che, per ogni $n \in \mathbb{N}^+$:

$$\sum_{j=1}^{+\infty} \text{Var}(U_{n,j}) \leq b(\tau)$$

Allora il termine $\sum_{j=1}^{+\infty}(\eta_{n,j} - \xi_{n,j})U_{n,j}$ tende a 0 in norma L^2 , pertanto può essere trascurato.

Restano da analizzare le prime due sommatorie del secondo membro, tra loro indipendenti: le analizzeremo separatamente, mostrando che la distribuzione di ognuna può essere approssimata con quella della distribuzione infinitamente divisibile associata.

Per la prima delle due somme, si può usare la proposizione (5.1.1), quindi la procedura di Lindeberg, e poi sfruttare il fatto che la convergenza in norma di Kolmogorov implica la convergenza rispetto alla distanza di Lévy, equivalente a quella in norma dual-Lipschitziana: date infatti due funzioni di ripartizione F e G , la loro distanza di Kolmogorov è maggiore o uguale di quella di Lévy (per una dimostrazione della disuguaglianza, si veda [4]; per una dimostrazione dell'equivalenza tra la convergenza in norma dual-Lipschitziana e quella rispetto alla distanza di Lévy, già accennata nella sezione (1.4), si veda [2]).

Per la seconda delle due somme, sia, per ogni $n, j \in \mathbb{N}^+$, $\beta_{n,j} = \mathbb{P}(\xi_{n,j}V_{n,j} \neq 0)$. Allora:

- Per la shift-compattatezza, usando la proposizione (5.1.1), vale:

$$\sup_{n \in \mathbb{N}^+} \left(\sum_{j=1}^{+\infty} \beta_{n,j} \right) < +\infty ;$$

- Essendo un array infinitesimale, si ha $\lim_{n \rightarrow +\infty} \sup_{j \in \mathbb{N}^+} \beta_{n,j} = 0$.

Allora, si ha $\lim_{n \rightarrow +\infty} \sum_{j=1}^{+\infty} \beta_{n,j}^2 = 0$, dato che:

$$\sum_{j=1}^{+\infty} \beta_{n,j}^2 \leq \left(\sup_{j \in \mathbb{N}^+} \beta_{n,j} \right) \sum_{j=1}^{+\infty} \beta_{n,j}$$

Da qui, usando la proposizione (5.1.2), si ottiene la convergenza in norma L^1 , che ovviamente implica la convergenza in norma dual-Lipschitziana. \square

Osservazione 52. Avendo introdotto la giusta notazione, possiamo ora osservare come già il teorema (4.1.2) sia una generalizzazione (peraltro dimostrata con tecniche del tutto elementari) del Teorema Limite Centrale, nella versione più conosciuta. Sia $Y \sim \mu$ una variabile aleatoria reale tale che $\mathbb{E}[Y] = 0$ e $\mathbb{E}[Y^2] = 1$, e consideriamo un array doppio $(X_{n,j})_{n,j \in \mathbb{N}^+}$ di variabili aleatorie reali indipendenti da Y tale che, fissato $n \in \mathbb{N}^+$:

- Le variabili $X_{n,1}, X_{n,2}, \dots$ risultano tra loro indipendenti;
- Per ogni $j \in \mathbb{N}^+$ vale $X_{n,j} \sim \frac{1}{\sqrt{n}}Y$ se $j \leq n$, $X_{n,j} \equiv 0$ se $j > n$ (sia μ_n tale che $X_{n,j} \sim \mu_n$, se $j \leq n$).

Utilizzando il teorema (4.1.2) e l'osservazione (39), per ogni $n \in \mathbb{N}^+$ si ottiene, definendo la legge normale standard $G = N(0, 1)$:

$$\left\| \bigstar_{k=1}^n \mu_n - G \right\|_K \leq C \left(\frac{1}{n} \right)^{\frac{1}{8}},$$

con C costante universale, da cui si ottiene la tesi del Teorema Limite Centrale, nella sua formulazione più comune.

5.2 I tre lemmi di Le Cam e una seconda generalizzazione

Nella sezione conclusiva di questa tesi enunceremo e dimostreremo tre lemmi riguardanti la trascurabilità di ben precise speranze condizionali, abbastanza tecnici, che sono conosciuti come **i tre lemmi di Le Cam**. Attraverso questi lemmi enunceremo una generalizzazione del teorema (5.1.4), e concluderemo sfruttando il lemma (1.4.1), che (lo avevamo già notato) permette di stabilire la convergenza di una successione di esperimenti (ad esempio binari) ad un dato esperimento limite mediante l'analisi dei rapporti di verosimiglianza, che sono variabili aleatorie (definite in spazi probabilizzati opportuni).

Ricordiamo che un esperimento binario è un esperimento $\mathcal{E} = \{P_0, P_1\}$ in cui lo spazio dei parametri è $\Theta = \{0, 1\}$. Per un tale esperimento, si definiscono:

- L'affinità $\rho(P_0, P_1) = \int_{\Omega} \sqrt{dP_0 dP_1}$;
- La distanza di Hellinger $h^2(P_0, P_1) = \frac{1}{2} \int_{\Omega} (\sqrt{dP_0} - \sqrt{dP_1})^2 = 1 - \rho(P_0, P_1)$;
- La massa $\beta(P_0, P_1)$ di P_1 singolare rispetto a P_0 . Sarà allora $P_1 = P'_1 + P''_1$, con $P'_1 \ll P_0$ e P''_1 singolare rispetto a P_0 , con:

$$\int_{\Omega} \frac{dP'_1}{dP_0} dP_0 = 1 - \beta(P_0, P_1)$$

Consideriamo ora, su $(\Omega, \mathcal{A}, P_0)$, la variabile aleatoria:

$$Y = \sqrt{\frac{dP_1}{dP_0}} - 1$$

Per ora l'introduzione di Y può destare qualche perplessità, ma più avanti vedremo come questa scelta non sarà stata del tutto casuale.

Enfatizziamo il fatto che non è detto a priori che sia $P_1 \ll P_0$, da cui dP_1/dP_0 non è una vera e propria derivata di Radon-Nikodym. Si ha in particolare, per quanto visto poc'anzi:

$$\int_{\Omega} \frac{dP_1}{dP_0} dP_0 = \int_{\Omega} \frac{dP'_1}{dP_0} dP_0 = 1 - \beta(P_0, P_1)$$

Si ha allora:

- $\mathbb{E}^0 [Y] = \int_{\Omega} Y dP_0 = \int_{\Omega} \sqrt{\frac{dP_1}{dP_0}} dP_0 - 1 = \int_{\Omega} \sqrt{dP_1 dP_0} - 1 = \rho(P_0, P_1) - 1 = -h^2(P_0, P_1)$;
- $\mathbb{E}^0 [Y^2] = \int_{\Omega} Y^2 dP_0 = \int_{\Omega} \frac{dP_1}{dP_0} dP_0 - 2 \int_{\Omega} \sqrt{\frac{dP_1}{dP_0}} dP_0 + 1 = 1 - \beta(P_0, P_1) + 2\rho(P_0, P_1) - 1 = 2h^2(P_0, P_1) - \beta(P_0, P_1) \leq 2h^2(P_0, P_1)$;
- $\text{Var}_0(Y) = \mathbb{E}^0 [Y^2] - (\mathbb{E}^0 [Y])^2 = (1 - \rho^2(P_0, P_1)) - \beta(P_0, P_1)$.

Ricordiamo poi una disuguaglianza già dimostrata:

$$h^2(P_0, P_1) \leq h(P_0, P_1)\sqrt{2 - h^2(P_0, P_1)} \leq h(P_0, P_1)\sqrt{2}$$

Consideriamo ora un array doppio di esperimenti binari $(\mathcal{E}_{n,j})_{n,j \in \mathbb{N}^+}$, con:

$$\mathcal{E}_{n,j} = \{P_{0,n,j}, P_{1,n,j}\}$$

Siano $\rho_{n,j}, h_{n,j}, \beta_{n,j}, Y_{n,j}$ le quantità e la variabile aleatoria (definite precedentemente) relative a $\mathcal{E}_{n,j}$. Definiamo poi i seguenti prodotti diretti di esperimenti:

$$\mathcal{E}_n \stackrel{\text{def}}{=} \bigotimes_{j=1}^{+\infty} \mathcal{E}_{n,j}$$

Supporremo d'ora in poi l'esistenza di due condizioni, una di limitatezza, l'altra di trascurabilità (d'ora in poi chiameremo queste condizioni \mathfrak{L} e \mathfrak{T}):

\mathfrak{L}) Per ogni $n \in \mathbb{N}^+$ la quantità $\sum_{j=1}^{+\infty} h_{n,j}^2$ è limitata da una certa costante $b \in [0, +\infty[$;

\mathfrak{T}) Si ha $\lim_{n \rightarrow +\infty} \sup_{j \in \mathbb{N}^+} h_{n,j}^2 = 0$.

Se ora $\mathcal{E}_n = \{\mathbf{P}_{0,n}, \mathbf{P}_{1,n}\}$, con:

$$\mathbf{P}_0 = \bigotimes_{j=1}^{+\infty} P_{0,n,j}, \quad \mathbf{P}_1 = \bigotimes_{j=1}^{+\infty} P_{1,n,j},$$

allora l'affinità ρ_n tra $\mathbf{P}_{0,n}$ e $\mathbf{P}_{1,n}$ è:

$$\rho_n = 1 - \mathbf{h}_n^2 = \prod_{j=1}^{+\infty} (1 - h_{n,j}^2) = \prod_{j=1}^{+\infty} \rho_{n,j}$$

Se allora valgono le due condizioni espresse sopra, sostanzialmente le probabilità non diventano singolari, al limite. Se infatti $\delta = \frac{1}{2}$, esiste $N \in \mathbb{N}^+$ tale che per $n \geq N$ vale $\sup_{j \in \mathbb{N}^+} h_{n,j}^2 \in [0, \delta]$, allora:

$$e^{-2h_{n,j}^2} \leq 1 - h_{n,j}^2 \leq e^{-h_{n,j}^2},$$

da cui:

$$e^{-2\sum_{j=1}^{+\infty} h_{n,j}^2} \leq 1 - \mathbf{h}_n^2 \leq e^{-\sum_{j=1}^{+\infty} h_{n,j}^2}$$

Notiamo poi che, per ogni $n \in \mathbb{N}^+$, per la disuguaglianza di Markov (fissato $\varepsilon > 0$):

$$\sup_{j \in \mathbb{N}^+} \mathbb{P}(|Y_{n,j}| > \varepsilon) = \sup_{j \in \mathbb{N}^+} \mathbb{P}(Y_{n,j}^2 > \varepsilon^2) \leq \sup_{j \in \mathbb{N}^+} \frac{\mathbb{E}[Y_{n,j}^2]}{\varepsilon^2} \leq \frac{2}{\varepsilon^2} \sup_{j \in \mathbb{N}^+} h_{n,j}^2,$$

da cui l'array doppio $(Y_{n,j})_{n,j \in \mathbb{N}^+}$ è un array infinitesimale (si noti che vi è anche indipendenza tra le variabili, per n fissato).

Lemma 5.2.1 (Primo Lemma di Le Cam). Siano soddisfatte le condizioni \mathfrak{L} e \mathfrak{T} , e sia $\mathbb{E}_n^0[X]$ il valore atteso di X sotto $\mathbf{P}_{0,n}$.

Allora:

- $-b \leq \sum_{j=1}^{+\infty} \mathbb{E}_n^0 [Y_{n,j}] \leq 0$;
- $\sum_{j=1}^{+\infty} \mathbb{E}_n^0 [Y_{n,j}^2] \leq 2b$;
- $\lim_{n \rightarrow +\infty} \sup_{j \in \mathbb{N}^+} \mathbb{E}_n^0 [Y_{n,j}^2] = 0$;
- Se sotto $\mathbf{P}_{0,n}$ si ha $\sum_{j=1}^{+\infty} Y_{n,j} \sim \mu_n$, allora $(\mu_n)_{n \in \mathbb{N}^+}$ è una famiglia tesa di probabilità su $\hat{\mathbb{R}}$.

Dimostrazione. Direttamente dalle ipotesi, si ha:

$$\mathbb{E}_n^0 [Y_{n,j}] = \int_{\Omega} Y_{n,j} dP_{0,n,j} = -h_{n,j}^2,$$

da cui si ha la prima delle quattro affermazioni della tesi:

$$-b \leq \sum_{j=1}^{+\infty} \mathbb{E}_n^0 [Y_{n,j}] = \mathbb{E}_n^0 \left[\sum_{j=1}^{+\infty} Y_{n,j} \right] \leq 0$$

Con un ragionamento analogo si ha la seconda affermazione:

$$\mathbb{E}_n^0 \left[\left(\sum_{j=1}^{+\infty} Y_{n,j} \right)^2 \right] = \sum_{j=1}^{+\infty} \mathbb{E}_n^0 [Y_{n,j}^2] \leq \sum_{j=1}^{+\infty} 2h_{n,j}^2 \leq 2b,$$

e la terza affermazione discende immediatamente dalla condizione \mathfrak{T} :

$$\begin{aligned} 0 \leq \lim_{n \rightarrow +\infty} \left(\sup_{j \in \mathbb{N}^+} \mathbb{E}_n^0 [Y_{n,j}^2] \right) &\leq \lim_{n \rightarrow +\infty} \left(\sup_{j \in \mathbb{N}^+} 2h_{n,j}^2 \right) = 0 \Rightarrow \\ &\Rightarrow \lim_{n \rightarrow +\infty} \left(\sup_{j \in \mathbb{N}^+} \mathbb{E}_n^0 [Y_{n,j}^2] \right) = 0 \end{aligned}$$

Le variabili $Z_n = \sum_{j=1}^{+\infty} Y_{n,j}$, per $n \in \mathbb{N}^+$, hanno dunque un valore atteso $-b \leq m_n \leq 0$ e una varianza $0 < \sigma_n^2 \leq 2b$. Allora, fissato $\varepsilon > 0$, per la disuguaglianza di Chebichev:

$$\mathbf{P}_{0,n} \left(|Z_n| \geq b + \sqrt{\frac{2b}{\varepsilon}} \right) \leq \mathbf{P}_{0,n} \left(|Z_n - \mathbb{E}_n^0 [Z_n]| \geq \sqrt{\frac{2b}{\varepsilon}} \right) \leq \frac{2b}{2b} \varepsilon = \varepsilon,$$

da cui $(\mu_n)_{n \in \mathbb{N}^+}$ è tesa. □

Lemma 5.2.2 (Secondo Lemma di Le Cam). Siano soddisfatte le condizioni \mathfrak{L} e \mathfrak{T} , e sia $A \subseteq \mathbb{R}$ con $0 \in \overset{\circ}{A}$: sia \mathbb{I}_A l'indicatrice dell'insieme A . Sia poi, per ogni coppia $n, j \in \mathbb{N}^+$:

$$c_{n,j} \stackrel{\text{def}}{=} \frac{\mathbb{E}_n^0 [Y_{n,j} \mathbb{I}_A(Y_{n,j})]}{\mathbb{E}_n^0 [\mathbb{I}_A(Y_{n,j})]}$$

Allora

$$\lim_{n \rightarrow +\infty} \left(\sum_{j=1}^{+\infty} c_{n,j}^2 \right) = 0$$

Dimostrazione. Dal lemma (5.2.1), e in particolare dalla terza delle quattro affermazioni della tesi, si ottiene che il denominatore di $c_{n,j}$ si avvicina uniformemente in j a 1 al tendere di n all'infinito. Infatti, dalla disuguaglianza di Markov:

$$\mathbf{P}_{0,n}(Y_{n,j} \notin A) \leq \mathbf{P}_{0,n}(|Y_{n,j}| \geq \varepsilon) \leq \frac{\mathbb{E}_n^0 [Y_{n,j}^2]}{\varepsilon^2},$$

ove $\varepsilon > 0$ è tale che $] -\varepsilon, \varepsilon [\subseteq A$. A meno di considerare un sottoarray $(c_{n,j})_{n \geq k, j \in \mathbb{N}^+}$, con k opportuno, possiamo ignorare il denominatore di $c_{n,j}$, supponendolo ad esempio compreso tra $\frac{1}{2}$ e 1.

Notiamo ora che:

$$\begin{aligned} \mathbb{E}_n^0 [Y_{n,j} \mathbb{I}_{A^c}(Y_{n,j})]^2 &\leq \mathbb{E}_n^0 [|Y_{n,j}| \mathbb{I}_{]-\varepsilon, \varepsilon[^c}(Y_{n,j})]^2 \leq \\ &\leq \frac{1}{\varepsilon} \mathbb{E}_n^0 [Y_{n,j}^2 \mathbb{I}_{]-\varepsilon, \varepsilon[^c}(Y_{n,j})]^2 \leq \frac{1}{\varepsilon} \mathbb{E}_n^0 [Y_{n,j}^2]^2 \end{aligned}$$

Allora, dato che:

$$\sum_{j=1}^{+\infty} \mathbb{E}_n^0 [Y_{n,j} \mathbb{I}_{A^c}(Y_{n,j})]^2 \leq \frac{1}{\varepsilon} \sum_{j=1}^{+\infty} \mathbb{E}_n^0 [Y_{n,j}^2]^2 \leq \frac{1}{\varepsilon} \left(\sup_{j \in \mathbb{N}^+} \mathbb{E}_n^0 [Y_{n,j}^2] \right) \sum_{j=1}^{+\infty} \mathbb{E}_n^0 [Y_{n,j}^2],$$

si ha:

$$\lim_{n \rightarrow +\infty} \sum_{j=1}^{+\infty} \mathbb{E}_n^0 [Y_{n,j} \mathbb{I}_{A^c}(Y_{n,j})]^2 = 0,$$

dato che:

$$\sup_{j \in \mathbb{N}^+} \mathbb{E}_n^0 [Y_{n,j}^2] \leq 2 \sup_{j \in \mathbb{N}^+} h_{n,j}^2, \quad \sum_{j=1}^{+\infty} \mathbb{E}_n^0 [Y_{n,j}^2] \leq 2 \sum_{j=1}^{+\infty} h_{n,j}^2 \leq 2b,$$

da cui:

$$\lim_{n \rightarrow +\infty} \sum_{j=1}^{+\infty} \mathbb{E}_n^0 [Y_{n,j}^2]^2 = \lim_{n \rightarrow +\infty} \left(\sup_{j \in \mathbb{N}^+} \mathbb{E}_n^0 [Y_{n,j}^2] \right) \sum_{j=1}^{+\infty} \mathbb{E}_n^0 [Y_{n,j}^2] = 0 \quad (5.3)$$

Inoltre:

$$\sum_{j=1}^{+\infty} \mathbb{E}_n^0 [Y_{n,j}]^2 = \sum_{j=1}^{+\infty} h_{n,j}^4 \leq \left(\sup_{j \in \mathbb{N}^+} h_{n,j}^2 \right) \sum_{j \in \mathbb{N}^+} h_{n,j}^2,$$

da cui:

$$\lim_{n \rightarrow +\infty} \sum_{j=1}^{+\infty} \mathbb{E}_n^0 [Y_{n,j}]^2 = 0$$

Per concludere, notiamo che:

$$\mathbb{E}_n^0 [Y_{n,j} \mathbb{I}_A(Y_{n,j})] = \mathbb{E}_n^0 [Y_{n,j}] - \mathbb{E}_n^0 [Y_{n,j} \mathbb{I}_{A^c}(Y_{n,j})],$$

da cui, usando la disuguaglianza $(a \pm b)^2 \leq 2a^2 + 2b^2$ e sommando su j :

$$\sum_{j=1}^{+\infty} \mathbb{E}_n^0 [Y_{n,j} \mathbb{I}_A(Y_{n,j})]^2 \leq 2 \left(\sum_{j=1}^{+\infty} \mathbb{E}_n^0 [Y_{n,j}]^2 + \sum_{j=1}^{+\infty} \mathbb{E}_n^0 [Y_{n,j} \mathbb{I}_{A^c}(Y_{n,j})]^2 \right),$$

Allora, per quanto visto prima:

$$\lim_{n \rightarrow +\infty} \sum_{j=1}^{+\infty} \mathbb{E}_n^0 [Y_{n,j} \mathbb{I}_A(Y_{n,j})]^2 = 0, \quad (5.4)$$

da cui la tesi. \square

Osservazione 53. Considerata $f \in P[0, 2]$, della forma:

$$f(x) = ax + \frac{b}{2}x^2 + x^2 R_2(x),$$

se $b \neq 0$ esiste $\tau > 0$ tale che, per $|x| \leq \tau$, $|R_2(x)| \leq \frac{|b|}{2}$.

Lemma 5.2.3 (Terzo Lemma di Le Cam). Siano soddisfatte le condizioni \mathfrak{L} e \mathfrak{T} . Sia $A \subseteq \mathbb{R}$, con $0 \in A$, e sia $f \in P[0, 2]$ limitata in A . Sia poi, per $n, j \in \mathbb{N}^+$:

$$c_{n,j} = \frac{\mathbb{E}_n^0 [f(Y_{n,j}) \mathbb{I}_A(Y_{n,j})]}{\mathbb{E}_n^0 [\mathbb{I}_A(Y_{n,j})]}$$

Allora:

$$\lim_{n \rightarrow +\infty} \left(\sum_{j=1}^{+\infty} c_{n,j}^2 \right) = 0$$

Dimostrazione. Come prima, possiamo trascurare il denominatore di $c_{n,j}$, analizzando solamente il numeratore. Sia $\tau > 0$ tale che $[-\tau, \tau] \subseteq A$ e:

- Se $b \neq 0$, allora $|R_2(x)| \leq \frac{b}{2}$ se $|x| \leq \tau$;
- Se $b = 0$, allora $|R_2(x)| \leq 1$ se $|x| \leq \tau$.

Allora:

$$\mathbb{E}_n^0 [f(Y_{n,j}) \mathbb{I}_A(Y_{n,j})] = \mathbb{E}_n^0 [f(Y_{n,j}) \mathbb{I}_{A_\tau}(Y_{n,j})] + \mathbb{E}_n^0 [f(Y_{n,j}) \mathbb{I}_{B_\tau}(Y_{n,j})], \quad (5.5)$$

con $A_\tau = [-\tau, \tau]$ e $B_\tau = A - A_\tau$. Per quanto riguarda il primo addendo al secondo membro nella formula (5.5):

$$\begin{aligned} \mathbb{E}_n^0 [f(Y_{n,j}) \mathbb{I}_{A_\tau}(Y_{n,j})] &= a \mathbb{E}_n^0 [Y_{n,j} \mathbb{I}_{A_\tau}(Y_{n,j})] + \frac{b}{2} \mathbb{E}_n^0 [Y_{n,j}^2 \mathbb{I}_{A_\tau}(Y_{n,j})] + \\ &+ \mathbb{E}_n^0 [R_2(Y_{n,j}) Y_{n,j}^2 \mathbb{I}_{A_\tau}(Y_{n,j})], \end{aligned}$$

ove:

$$\left| \frac{b}{2} \mathbb{E}_n^0 [Y_{n,j}^2 \mathbb{I}_{A_\tau}(Y_{n,j})] + \mathbb{E}_n^0 [R_2(Y_{n,j}) Y_{n,j}^2 \mathbb{I}_{A_\tau}(Y_{n,j})] \right| \leq (1 + |b|) \mathbb{E}_n^0 [Y_{n,j}^2 \mathbb{I}_{A_\tau}(Y_{n,j})]$$

Usando sempre la disuguaglianza $(a \pm b)^2 \leq 2a^2 + 2b^2$ e sommando in j si ha:

$$\begin{aligned} & \sum_{j=1}^{+\infty} \mathbb{E}_n^0 [f(Y_{n,j}) \mathbb{I}_{A_\tau}(Y_{n,j})]^2 \leq \\ & \leq 2 \left(a^2 \sum_{j=1}^{+\infty} \mathbb{E}_n^0 [Y_{n,j} \mathbb{I}_{A_\tau}(Y_{n,j})]^2 + (1 + |b|)^2 \sum_{j=1}^{+\infty} \mathbb{E}_n^0 [Y_{n,j}^2 \mathbb{I}_{A_\tau}(Y_{n,j})]^2 \right), \end{aligned}$$

e riguardando la dimostrazione del lemma (5.2.2), e in particolare la formula (5.4), si ha:

$$\lim_{n \rightarrow +\infty} \sum_{j=1}^{+\infty} \mathbb{E}_n^0 [f(Y_{n,j}) \mathbb{I}_{A_\tau}(Y_{n,j})]^2 = 0$$

Per quanto riguarda il secondo addendo al secondo membro nella formula (5.5), si ha:

$$| \mathbb{E}_n^0 [f(Y_{n,j}) \mathbb{I}_{B_\tau}(Y_{n,j})] | \leq C \mathbf{P}_{0,n}(Y_{n,j} \in B_\tau),$$

essendo f limitata in A . Per la disuguaglianza di Chebichev:

$$[\mathbf{P}_{0,n}(Y_{n,j} \in B_\tau)]^2 \leq [\mathbf{P}_{0,n}(|Y_{n,j}| \geq \tau)]^2 \leq \left(\frac{\mathbb{E}_n^0 [Y_{n,j}^2]}{\tau^2} \right)^2 = \frac{\mathbb{E}_n^0 [Y_{n,j}^2]^2}{\tau^4},$$

da cui, riguardando ancora una volta la dimostrazione del lemma (5.2.2), e in particolare la formula (5.3):

$$\lim_{n \rightarrow +\infty} \sum_{j=1}^{+\infty} \mathbb{E}_n^0 [Y_{n,j}^2]^2 = 0 \Rightarrow \lim_{n \rightarrow +\infty} \sum_{j=1}^{+\infty} \mathbb{E}_n^0 [f(Y_{n,j}) \mathbb{I}_{B_\tau}(Y_{n,j})]^2 = 0,$$

da cui la tesi. □

Osservazione 54. Data una funzione $f : (\Omega, \mathcal{A}, m) \rightarrow \mathbb{R}$ misurabile, indicheremo con $f(m)$ la misura immagine di m tramite f . Per $A \in \mathcal{B}(\mathbb{R})$, allora:

$$f(m)(A) = m(f^{-1}(A))$$

Notiamo che, se $X \sim m$, allora $f(X) \sim f(m)$. Notiamo che l'operatore \mathbf{f} che manda μ in $f(\mu)$ è lineare. Infatti, se μ, ν sono misure su (Ω, \mathcal{A}) , allora per $A \in \mathcal{B}(\mathbb{R})$:

$$f(\mu + \nu)(A) = (\mu + \nu)(f^{-1}(A)) = \mu(f^{-1}(A)) + \nu(f^{-1}(A)) = f(\mu)(A) + f(\nu)(A)$$

Questi tre lemmi hanno diverse applicazioni, nell'ambito dell'approssimazione in legge di somme di variabili aleatorie. Ad esempio enunciamo un teorema, di cui non riportiamo la dimostrazione, che rappresenta una generalizzazione del teorema (5.1.4).

Teorema 5.2.4. Siano soddisfatte le condizioni \mathfrak{L} e \mathfrak{F} , e sia $f \in P[0, 2]$ a valori reali e misurabile. Indichiamo con $G_{n,j}$ la legge di $Y_{n,j}$, e sia per ogni $n \in \mathbb{N}^+$ $L_n = \sum_{j=1}^{+\infty} G_{n,j}$ la misura di Lévy, con la massa a 0 posta uguale a 0.

Allora la distanza di Lévy tra la legge di $\sum_{j=1}^{+\infty} f(Y_{n,j})$ e la convoluzione esponenziale $e^{f(L_n) - \|f(L_n)\|}$ tende a 0, al tendere di n a $+\infty$.

Osservazione 55. Il teorema (5.2.4) ci dice, in sostanza, che si possono approssimare le distribuzioni delle somme del tipo:

$$\sum_{j=1}^{+\infty} f(Y_{n,j})$$

conoscendo le distribuzioni delle somme del tipo:

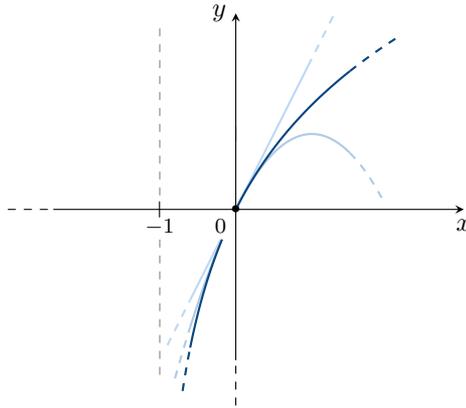
$$\sum_{j=1}^{+\infty} Y_{n,j}$$

È abbastanza chiaro che, se $f \in P[0, 2]$ è ad esempio iniettiva e non singolare in 0, allora, conoscendo le distribuzioni delle somme del primo tipo, si possono approssimare le distribuzioni delle somme del secondo tipo.

Concludiamo ora apprezzando l'utilità delle variabili $(Y_{n,j})_{n,j \in \mathbb{N}^+}$ introdotte all'inizio del capitolo. Se indichiamo con $\Lambda_{n,j}$ il logaritmo del rapporto di verosimiglianza dell'esperimento $\mathcal{E}_{n,j}$, allora possiamo scrivere $\Lambda_{n,j} = f(Y_{n,j})$, con:

$$f(y) = 2 \ln(1 + y)$$

Il grafico di tale funzione è il seguente (la funzione f è riportata in blu):



In particolare, $f \in P[0, 2]$, in quanto:

$$f(y) = 2y - y^2 + o(y^2),$$

per $y \rightarrow 0$ (i grafici della retta e della parabola approssimanti sono riportati in celeste, nel grafico precedente). Dunque, in sostanza, è applicabile il teorema (5.2.4), appena enunciato.

Osservazione 56. La funzione f definita prima è definita in $[-1, +\infty[$ ed è a valori in $[-\infty, +\infty[$, quindi non a valori reali: è dunque a priori errato applicare i teoremi visti finora per dedurre i relativi risultati. In realtà, si può definire la convoluzione per misure finite con segno definite sul semigruppato additivo $[-\infty, +\infty[$, ridefinire opportunamente una distanza su tale semigruppato (ad esempio $\delta(x_1, x_2) = |e^{x_1} - e^{x_2}|$, con $x_1, x_2 \in [-\infty, +\infty[$, con $e^{-\infty} = 0$), quindi ridefinire una distanza tra misure mediante la relativa norma dual-Lipschitziana. Una volta fatto ciò, i teoremi rimangono validi, e le dimostrazioni sono praticamente le stesse.

Applicando il teorema (5.2.4), pertanto, si ha che la legge di Λ_n , logaritmo del rapporto di verosimiglianza di \mathcal{E}_n , è approssimabile dalla convoluzione esponenziale $e^{f(L_n) - \|f(L_n)\|}$. Ricordando poi la linearità dell'operatore f che manda μ in $f(\mu)$, si ha che $e^{f(L_n) - \|f(L_n)\|}$ è esattamente la legge del logaritmo del rapporto di verosimiglianza di $\mathbf{Aid}(\mathcal{E}_n)$ (si veda la definizione (3.2.5)). Questi risultati possono poi essere usati per valutare la convergenza di $(\mathcal{E}_n)_{n \in \mathbb{N}^+}$ a un dato esperimento limite \mathcal{F} , usando il lemma (1.4.1): nel caso di esperimenti binari, infatti, il lemma è sicuramente applicabile, ed è anche funzionale. Se si è interessati, vi sono molti risultati a riguardo in [6] e [7]: noi invece concludiamo qui la trattazione.

Bibliografia

- [1] Da Prato, G. - Zabczyk, Z. *Stochastic equations in infinite dimensions*. Cambridge University Press, 2014.
- [2] Dudley, R.M. *Real Analysis and Probability*. Cambridge University Press, 2004.
- [3] Esseen, C.G. On the Lyapounov limit of error in the theory of probability. *Arkiv för matematik, astronomi och fysik*, 28A(9):1–19, 1942.
- [4] Huber, P.J. - Ronchetti, E.M. *Robust Statistics*. John Wiley & Sons, 2009.
- [5] Le Cam, L. On the information contained in additional observations. *The Annals of Statistics*, 2(4):630–649, 1974.
- [6] Le Cam, L. *Asymptotic Methods in Statistical Decision Theory*. Springer-Verlag, 1986.
- [7] Le Cam, L. - Yang, G.L. *Asymptotic in Statistics: some basic concepts*. Springer-Verlag, 2000.
- [8] Lindeberg, J.W. Über das Exponentialgesetz in der Wahrscheinlichkeitsrechnung. *Annales Academiae Scientiarum Fennicae*, 16:1–23, 1920.
- [9] Lindeberg, J.W. Eine neue Herleitung des Exponentialgesetzes in der Wahrscheinlichkeitsrechnung. *Mathematische Zeitschrift*, 15(1):211–225, 1922.
- [10] Low, M.G. - Zhou, H.H. A complement to Le Cam's theorem. *The Annals of Statistics*, 35(3):1146–1165, 2007.
- [11] Nussbaum, M. Asymptotic equivalence of density estimation and gaussian white noise. *The Annals of Statistics*, 24(6):2399–2430, 1996.
- [12] Rudin, W. *Fourier analysis on groups*. Wiley-Interscience, 1990.
- [13] Torgersen, E.N. *Comparison of statistical experiments*. Cambridge University Press, 1991.