



SAPIENZA
UNIVERSITÀ DI ROMA

Facoltà di Scienze Matematiche Fisiche e Naturali
Dipartimento di Matematica "G. Castelnuovo"
Corso di Laurea Magistrale in Matematica Applicata

Tesi di Laurea Magistrale

Metodo di Stein
e applicazioni al problema del change
point detection

Relatrice:
Ch.ma. Prof.ssa
Alessandra Faggionato

Candidato:
Simmaco Di Lillo
matr. 2016774

Anno Accademico 2022/2023

Indice

1	Introduzione	3
2	Metriche di probabilità	10
3	Il metodo di Stein	14
3.1	L'equazione di Stein	14
3.2	Leave one out	20
3.3	Exchangeable pairs	24
3.4	Zero-bias	29
3.5	Size-bias	33
4	Change-point detection con metodi grafici	39
4.1	Il change-point detection	39
4.2	Alcune definizioni sui grafi	40
4.3	Grafi associati a quantità osservate	43
4.4	Le variabili aleatorie $R_G(t)$ e $Z_G(t)$	44
4.5	Teorema di approssimazione del processo $Z_G(\cdot)$	46
4.6	Il test statistico	47
4.7	Dimostrazione del Lemma 4.4.1	50
5	Derivazione del Teorema 4.5.1	55
5.1	Le variabili $R_G^B(t)$ e $Z_G^B(t)$ definito sulle n -uple	55
5.2	Le variabili aleatorie $n^B(t)$ e $X^B(t)$	60
5.3	Un risultato di approssimazione	62
5.4	Legami tra le due distribuzioni: bootstrap e permutation null	70
5.5	Derivazione del Teorema 4.5.1	73
6	Discussione dell'Ansatz 1	75
6.1	$Z_G^B(t)$ e $Z_G^B(s)$	77
6.2	$X^B(t)$ e $X^B(s)$	85
6.3	$X^B(t)$ e $Z_G^B(t)$	87
6.4	$X^B(t)$ e $Z_G^B(s)$	92
A	Richiami di analisi funzionale	93
A.1	Assoluta continuità	93
A.2	Funzioni a tratti	97
A.3	Funzioni lipschitziane	97

B	Richiami di probabilità	98
B.1	Probabilità generale	98
B.2	Convergenza in legge	100
B.3	Valore atteso condizionato	100
B.4	Processi stocastici e vettori multivariati	102
C	Alcuni risultati sui grafi	103
C.1	Alcuni algoritmi sui grafi	105
C.1.1	Algoritmo di Kruskal	106
C.1.2	Accoppiamento di costo minimo	106
	Bibliografia	110

1 | Introduzione

La tesi è divisa in due parti: nella prima parte presenteremo nel dettaglio il metodo di Stein per l'approssimazione gaussiana analizzando alcune differenti tecniche (leave one out, exchangeable pairs, zero-bias e size-bias). Nella seconda parte, invece, riprendendo e ampliando il lavoro svolto in [8], mostreremo un'applicazione del metodo di Stein al problema del change-point detection.

Il teorema del limite centrale è risultato fondamentale nel campo della probabilità e statistica.

Teorema 1 (Teorema del limite centrale). *Sia $(X_n)_{n \in \mathbb{N}}$ una sequenza di variabili aleatorie indipendenti, identicamente distribuite e con varianza finita. Allora per ogni $x \in \mathbb{R}$ vale*

$$\lim_{n \rightarrow +\infty} \mathbb{P} \left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \leq x \right) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy$$

dove $S_n := \sum_{i=1}^n X_i$, $\mu := \mathbb{E}[X_1]$ e $\sigma^2 := \text{Var}(X_1)$.

Tale teorema spiega perché la distribuzione normale appare in aree molto diverse tra di loro come per esempio nel gioco d'azzardo, negli errori di misurazione, nel campionamento e nella meccanica statistica.

Tipicamente siamo interessati, soprattutto nelle applicazioni, ad avere informazioni circa l'andamento dell'errore che si commette utilizzando l'approssimazione normale. Risulta, dunque, fondamentale una formulazione quantitativa del teorema del limite centrale. Uno dei primi risultati in questa direzione è stato provato in maniera indipendente da Berry [3] ed Esseen [19]. Riportiamo il Teorema di Berry-Esseen per variabili i.i.d. (cfr. [16, Teorema 3.4.17]).

Teorema 2 (Teorema di Berry-Esseen per variabili i.i.d.). *Siano X_1, \dots, X_n variabili aleatorie i.i.d. con $\mathbb{E}[X_1] = 0$, $\mathbb{E}[X_1^2] = \sigma^2$ e $\mathbb{E}[|X_1|^3] = \rho < \infty$. Se F_n è la funzione di ripartizione della variabile $\frac{\sum_{i=1}^n X_i}{\sigma\sqrt{n}}$ e Φ è quella della normale standard allora vale*

$$|F_n(x) - \Phi(x)| \leq 3 \frac{\rho}{\sigma^3 \sqrt{n}}.$$

Osservazione 1. Notiamo che sotto le ipotesi del Teorema 2 il tasso di $n^{-\frac{1}{2}}$ non può essere migliorato [16, Sezione 3.4.4].

Il metodo di Stein per l'approssimazione gaussiana, introdotto per la prima volta in [39], permette di ottenere versioni quantitative del teorema del limite centrale rispetto ad una classe di distanze tra le variabili aleatorie. Più precisamente, data una classe di funzioni boreliane \mathcal{H} e una variabile aleatoria S , utilizzando il metodo di Stein possiamo ottenere stime quantitative per $d_{\mathcal{H}}(S, Z)$ dove $Z \sim \mathcal{N}(0, 1)$ e

$$d_{\mathcal{H}}(X, Y) := \sup_{h \in \mathcal{H}} |\mathbb{E}[h(X)] - \mathbb{E}[h(Y)]|.$$

Nel Capitolo 2 mostreremo che, sotto alcune ipotesi su \mathcal{H} e sulle variabili aleatorie in gioco, la funzione $d_{\mathcal{H}}$ è una metrica. In tale capitolo ci soffermeremo su tre differenti metriche:

- la distanza indotta da

$$\mathcal{H} = \{1_A \mid A \in \mathcal{B}(\mathbb{R})\},$$

prende il nome di distanza della variazione totale e viene denotata con d_{TV} ,

- la distanza indotta da

$$\mathcal{H} = \{1_{(-\infty, x]} \mid x \in \mathbb{R}\}$$

prende il nome di distanza di Kolmogorov e viene denotata con d_K e

- la distanza indotta da

$$\mathcal{H}_W = \{f : \mathbb{R} \rightarrow \mathbb{R} \mid f \text{ è } 1\text{-lipschitziana}\}$$

prende il nome di distanza di Wasserstein e viene denotata con d_W .

Grazie a queste definizioni il Teorema 2 può essere riformulato nel seguente modo:

Teorema 3. Siano X_1, \dots, X_n variabili aleatorie indipendenti con $\mathbb{E}[X_1] = 0$, $\mathbb{E}[X_1^2] = \sigma^2$ e $\mathbb{E}[|X_1|^3] = \rho < \infty$. Se $S = \frac{\sum_{i=1}^n X_i}{\sigma\sqrt{n}}$ e $Z \sim \mathcal{N}(0, 1)$, allora

$$d_K(S, Z) \leq 3 \frac{\rho}{\sigma^3 \sqrt{n}}.$$

In generale, fissata una distribuzione target Z e una variabile aleatoria S , il metodo di Stein ha due parti fondamentali: la prima consiste nel trasformare il problema della limitazione $d_{\mathcal{H}}(S, Z)$ in quello della limitazione di $\mathbb{E}[\mathcal{A}(S)]$ dove \mathcal{A} è un funzionale che dipende unicamente da Z . Tipicamente per la scelta del

funzionale si fa ricorso ad un operatore caratteristico ovvero ad una funzione per la quale la distribuzione target è l'unico punto fisso. Ad esempio, per l'approssimazione gaussiana utilizziamo il seguente risultato noto come Lemma di Stein (cfr. [9, Lemma 2.1]).

Lemma 4 (Lemma di Stein).

- (i) Se $S \sim \mathcal{N}(0, 1)$ allora per ogni funzione assolutamente continua $f : \mathbb{R} \rightarrow \mathbb{R}$ con $\mathbb{E}[|f'(S)|], \mathbb{E}[|Sf(S)|] < \infty$ vale $\mathbb{E}[f'(S)] = \mathbb{E}[Sf(S)]$.
- (ii) Sia S una variabile aleatoria. Se per ogni funzione C^1 a tratti e limitata $f : \mathbb{R} \rightarrow \mathbb{R}$ con $\mathbb{E}[|f'(S)|], \mathbb{E}[|Sf(S)|] < \infty$ vale $\mathbb{E}[f'(S)] = \mathbb{E}[Sf(S)]$, allora $S \sim \mathcal{N}(0, 1)$.

Nella seconda parte, utilizzando la struttura di S e le proprietà di \mathcal{A} , viene stimato $\mathbb{E}[\mathcal{A}(S)]$. Per ottenere tale stima sono state sviluppate, nella letteratura, diverse tecniche.

Nella prima sezione del Capitolo 3 abbiamo spiegato in dettaglio la prima componente del metodo di Stein nel caso in cui $Z \sim \mathcal{N}(0, 1)$. Nelle sezioni successive, riprendendo il lavoro presentato in [9] e [33], abbiamo generalizzato il risultato del Teorema 3 sostituendo alla distanza di Kolmogorov quella di Wasserstein .

La prima limitazione trovata utilizza la tecnica del “leave one out” (tale termine compare per la prima volta in [39]). Se X_1, \dots, X_n sono variabili aleatorie e $i \in [n]$ definiamo la i -esima variabile “leave one out” come

$$S^{(i)} := \sum_{\substack{j=1 \\ j \neq i}}^n X_j.$$

Grazie a questa tecnica, nella Sezione 3.2 abbiamo provato il seguente risultato:

Teorema 5 (Teorema 3.2.1). Siano X_1, \dots, X_n variabili aleatorie indipendenti, centrate con momenti terzi assoluti finiti. Se $\sum_{i=1}^n \text{Var}(X_i) = 1$ allora vale

$$d_W(S, Z) \leq 4 \sum_{i=1}^n \mathbb{E}[|X_i|^3]$$

dove $S = \sum_{i=1}^n X_i$ e $Z \sim \mathcal{N}(0, 1)$.

Tipicamente per mostrare risultati come il Teorema 3 si utilizzano metodi basati sull'analisi di Fourier o sul metodo dei momenti [23]. Per poter applicare tali metodi, tuttavia, l'ipotesi di indipendenza tra le variabili aleatorie in gioco risulta cruciale. Grazie all'applicazione del metodo di Stein abbiamo generalizzato il risultato del Teorema 5 sostituendo all'indipendenza delle variabili una condizione meno restrittiva. Più precisamente, riprendendo la dimostrazione in [1, Teorema 3.4] e dimostrando alcuni dettagli lasciati come esercizio, otteniamo il seguente risultato: .

Teorema 6 (Teorema 3.2.2). *Sia \mathcal{I} un insieme di indici e sia $(\xi_i)_{i \in \mathcal{I}}$ una \mathcal{I} -successione di variabili aleatorie centrate con momenti terzi assoluti finiti. Supponiamo che per ogni $i \in \mathcal{I}$ esistano degli insiemi $S_i \subseteq T_i \subseteq \mathcal{I}$ tali che per ogni i si abbia*

$$\begin{aligned} & \xi_i \prod_{j \in S_i^c} (\xi_j), \\ & (\xi_j)_{j \in S_i} \prod_{j \in T_i^c} (\xi_j). \end{aligned}$$

Posto $W = \sum_{i \in \mathcal{I}} \xi_i$, se $\text{Var}(W) = 1$ allora vale

$$d_W(W, Z) \leq \sum_{i \in \mathcal{I}} \left\{ 2\mathbb{E}[|\xi_i \eta_i \tau_i|] + 2|\mathbb{E}[\xi_i \eta_i]| \mathbb{E}[|\tau_i|] + \mathbb{E}[|\xi_i \eta_i^2|] \right\}$$

dove $Z \sim \mathcal{N}(0, 1)$, $\eta_i = \sum_{j \in S_i} \xi_j$ e $\tau_i = \sum_{j \in T_i} \xi_j$.

Un'altra tecnica utilizzata in letteratura è quella dell'exchangeable pair [40]. Due variabili aleatorie S ed S' , definite sullo stesso spazio di probabilità, formano un'exchangeable pair se $(S, S') \sim (S', S)$. Inoltre, se $\mathbb{E}[|S'|] < \infty$ ed esiste $\lambda \in (0, 1)$ tale che

$$\mathbb{E}[S'|S] = (1 - \lambda)S$$

diremo che (S, S') è una λ -Stein coppia.

Grazie a questa definizione abbiamo dimostrato, nella Sezione 3.3, un teorema simile al precedente ma che migliora la costante assoluta.

Teorema 7 (Teorema 3.3.3). *Siano X_1, \dots, X_n variabili aleatorie indipendenti, centrate e con momenti terzi assoluti finiti. Se $\sum_{i=1}^n \text{Var}(X_i) = 1$ allora*

$$d_W(S, Z) \leq \sqrt{\frac{1}{2\pi}} \mathbb{E} \left[\left| 1 - \mathbb{E} \left[\sum_{i=1}^n X_i^2 \middle| S \right] \right| \right] + \frac{8}{3} \sum_{i=1}^n \mathbb{E}[|X_i|^3]$$

dove $S = \sum_{i=1}^n X_i$ e $Z \sim \mathcal{N}(0, 1)$.

Il prossimo metodo utilizzato fa uso della trasformata zero-bias introdotta da Goldstein e Reinert in [25]. Se S è una variabile aleatoria centrata con varianza σ^2 , diremo che una variabile aleatoria S^* ha la distribuzione zero-bias associata ad S se per ogni funzione assolutamente continua $f : \mathbb{R} \rightarrow \mathbb{R}$ con $\mathbb{E}[|Sf(S)|], \mathbb{E}[|f'(S^*)|] < \infty$ vale

$$\mathbb{E}[Sf(S)] = \sigma^2 \mathbb{E}[f'(S^*)].$$

Nella Sezione 3.4 abbiamo provato alcuni risultati sulla distribuzione zero-bias associata a variabili aleatorie definite mediante somme. In particolare, abbiamo applicato questi risultato al caso di variabili aleatorie di Bernoulli indipendenti ottenendo il seguente risultato:

Proposizione 8 (Proposizione 3.4.7). *Siano X_1, \dots, X_n i.i.d. con distribuzione di Bernoulli di parametro p . Allora vale allora*

$$d_W(\bar{S}, Z) \leq \frac{1}{\sqrt{np(1-p)}}$$

dove $\bar{S} = \frac{\sum_{i=1}^n X_i - np}{\sqrt{np(1-p)}}$ e $Z \sim \mathcal{N}(0, 1)$.

Un'altra tecnica utilizzata in letteratura è quello della distribuzione size-bias per variabili aleatorie non-negative (introdotta per la prima volta in [26]). Se S è una variabile aleatoria non-negativa con media $\mu > 0$ diremo che la variabile aleatoria S^s ha la distribuzione size-bias associata ad S se per ogni funzione limitata e continua $f : \mathbb{R} \rightarrow \mathbb{R}$ vale

$$\mathbb{E}[Sf(S)] = \mu \mathbb{E}[f(S^s)].$$

Nella Sezione 3.5 abbiamo caratterizzato la distribuzione di probabilità della distribuzione size-bias associata a variabili aleatorie discrete (cfr. Lemma 3.5.1). Tale caratterizzazione è stata utilizzata per stimare la distanza di Wasserstein tra la distribuzione data dalla somma di variabili aleatorie di Poisson indipendenti e la distribuzione normale standard. In particolare, abbiamo ottenuto il seguente risultato:

Proposizione 9 (Proposizione 3.5.6). *Siano X_1, \dots, X_n i.i.d. con distribuzione di Poisson di parametro λ . Allora vale*

$$d_W(\bar{S}, Z) \leq \frac{1}{\sqrt{\lambda n}}$$

dove $\bar{S} = \frac{\sum_{i=1}^n X_i - \lambda n}{\sqrt{n\lambda}}$ e $Z \sim \mathcal{N}(0, 1)$.

Nella seconda parte della tesi abbiamo introdotto il problema del change-point detection. I modelli di change-point detection sono ampiamente utilizzati in svariati campi per rilevare la non omogeneità in una sequenza di osservazioni. Tipicamente, data una sequenza di osservazioni $(y_i)_{i=1}^n$ siamo interessati a capire se i dati abbiamo tutti una medesima distribuzione F_0 oppure, al contrario, se esiste un tempo τ (detto change-point) tale per cui le osservazioni precedenti a τ seguono una distribuzione F_0 e quelle successive una distribuzione F_1 differente da F_0 . Più precisamente, data una sequenza di osservazioni $(y_i)_{i=1}^n$ indicizzate in un qualche modo (spaziale e/o temporale) e dati $1 \leq n_0 \leq n_1 \leq n$ siamo interessati a testare l'ipotesi nulla:

$$\mathcal{H}_0 : y_i \sim F_0 \quad \text{per } i = 1, \dots, n$$

contro l'ipotesi alternativa:

$$\mathcal{H}_1 : \exists \tau : \quad n_0 \leq \tau \leq n_1 \leq n \text{ per cui } y_i \sim \begin{cases} F_1 & \text{se } i > \tau \\ F_0 & \text{altrimenti} \end{cases}$$

per alcune distribuzioni di probabilità F_0 e F_1 distinte (non fissate a priori).

Nella tesi, riprendendo e ampliando il lavoro svolto in [8], abbiamo descritto un approccio non parametrico per tale problema. Andiamo ad illustrare schematicamente tale approccio. Per poter applicare tale metodo a dati in spazi molto generali (anche non euclidei) abbiamo associato alle osservazioni un grafo. In particolare, nelle sperimentazioni condotte, vengono utilizzate tre differenti costruzioni. Sia \mathcal{X} un insieme e sia $d : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$ una funzione assegnata pensata come funzione di dissimilarità. Dato $V \subseteq \mathcal{X}$ finito introduciamo i seguenti grafi.

- Se \mathcal{T} è l'insieme degli alberi su V definiamo l'**albero di copertura di dissimilarità minima** (*minimum spanning tree*) come \bar{T} dove

$$\bar{T} = \operatorname{argmin}_{T=(V,E) \in \mathcal{T}} \left\{ \sum_{\{x,y\} \in E} d(x,y) \right\}.$$

- Se \mathfrak{P} l'insieme delle partizioni di V formate da coppie, cioè

$$\mathfrak{P} = \left\{ P \subseteq \binom{V}{2} \mid P \text{ partizione di } V \right\},$$

definiamo la partizione di costo minimo \bar{P} come

$$\bar{P} := \operatorname{argmin}_{P \in \mathfrak{P}} \left\{ \sum_{\{x,y\} \in P} d(x,y) \right\}.$$

Il **grafo di minimo accoppiamento** (*minimum distance pairing*) è (V, \overline{P}) .

- Sia $n : V \rightarrow V$ la mappa definita da

$$n(x) := \operatorname{argmin}_{\substack{y \in V \\ y \neq x}} d(x, y).$$

Il grafo del **vicino più vicino** (*nearest neighbor graph*) è definito come (V, E) dove $E := \{\{x, n(x)\} \mid x \in V\}$.

Data una sequenza di osservazioni distinte (y_1, \dots, y_n) e posto $V = \{y_1, \dots, y_n\}$ sia $G = (V, E)$ uno dei grafi definiti sopra.

Sia \mathcal{S}_n il gruppo delle permutazioni su $[n]$ e denotiamo con \mathbb{P} la distribuzione uniforme su \mathcal{S}_n . Fissato $t \in \mathbb{N}$ definiamo la variabile aleatoria $R_G(t) : \mathcal{S}_n \rightarrow \mathbb{N}$ dove $R_G(t)(\pi)$ conta il numero di lati $\{i, j\} \in E$ per cui $\pi(i) \leq t < \pi(j)$ o $\pi(j) \leq t < \pi(i)$. Definiamo, inoltre, la variabile aleatoria

$$Z_G(t) := -\frac{R_G(t) - \mathbb{E}[R_G(t)]}{\sqrt{\operatorname{Var}(R_G(t))}}.$$

Il test statistico utilizzato nella tesi prevede di rigettare l'ipotesi nulla (dunque la non esistenza di change-point) se

$$\max_{n_0 \leq t \leq n_1} Z_G(t) > b$$

dove la soglia b viene scelta in modo da avere un controllo sull'errore di primo tipo, cioè, sul valore di $\alpha := \mathbb{P}\left(\max_{n_0 \leq t \leq n_1} Z_G(t) > b\right)$. Poiché \mathbb{P} è la distribuzione uniforme su \mathcal{S}_n , il calcolo di α richiede un costo dell'ordine di $n!$ operazioni. Al crescere di n tale calcolo diventa computazionalmente pesante. Per ovviare a tale problema, sotto opportune ipotesi sul grafo G , abbiamo discusso circa l'approssimazione di Z_G ad un processo gaussiano. Nel nostro lavoro (così come in [8]) non abbiamo dimostrato formalmente tale approssimazione, questa ipotesi di lavoro viene però legittimata da un risultato di approssimazione (cfr. Sezione 5.3) e confermata sperimentalmente [8, Sezione 3.5]. La dimostrazione di tale risultato segue quella di [8] ma vari passaggi lasciati al lettore sono stati elaborati e aggiunti. Precisiamo, inoltre, che nella dimostrazione presentata in [8] si assume implicitamente un ansatz. Nel Capitolo 6 forniremo una discussione più dettagliata su questa assunzione verificandone la validità in casi particolari (i risultati del Capitolo 6 sono originali e non fanno parte di [8])

2 | Metriche di probabilità

Definizione 2.1. Sia \mathcal{H} una collezione di funzioni boreliane. Diremo che \mathcal{H} è **separante** se per ogni coppia di variabili aleatorie (X, Y) si ha $X \sim Y$ (ovvero X e Y hanno la stessa legge) qualora per ogni $h \in \mathcal{H}$ con $\mathbb{E}[|h(X)|], \mathbb{E}[|h(Y)|] < \infty$ valga $\mathbb{E}[h(X)] = \mathbb{E}[h(Y)]$.

Definizione 2.2. Sia \mathcal{H} una collezione di funzioni boreliane separante. Se X, Y sono variabili aleatorie con $\mathbb{E}[|h(X)|], \mathbb{E}[|h(Y)|] < \infty$ per ogni $h \in \mathcal{H}$, definiamo la **distanza indotta dalla classe \mathcal{H}** tra le leggi di X e Y come

$$d_{\mathcal{H}}(X, Y) := \sup_{h \in \mathcal{H}} |\mathbb{E}[h(X)] - \mathbb{E}[h(Y)]|. \quad (2.0.1)$$

Proposizione 2.1. Fissata una collezione separante \mathcal{H} , la funzione definita da (2.0.1) è una pseudo-metrica sullo spazio delle variabili aleatorie tali per cui $\mathbb{E}[|h(X)|] < \infty$ per ogni $h \in \mathcal{H}$. Se identifichiamo le variabili aleatorie uguali quasi ovunque otteniamo una metrica.

Dimostrazione. Siano X, Y, Z tre variabili aleatorie per cui $\mathbb{E}[|h(X)|], \mathbb{E}[|h(Y)|], \mathbb{E}[|h(Z)|] < \infty$ per ogni $h \in \mathcal{H}$.

- Poiché $d_{\mathcal{H}}$ è definito come estremo superiore di quantità finite e positive, $d_{\mathcal{H}}(X, Y) \geq 0$. Inoltre, se $d_{\mathcal{H}}(X, Y) = 0$ allora $\mathbb{E}[h(X)] = \mathbb{E}[h(Y)]$ per ogni $h \in \mathcal{H}$. Poiché \mathcal{H} è separante si ha $X \sim Y$.
- Notiamo che (2.0.1) è una funzione simmetrica in X e Y .
- Proviamo la disuguaglianza triangolare

$$\begin{aligned} d_{\mathcal{H}}(X, Y) &= \sup_{h \in \mathcal{H}} |\mathbb{E}[h(X)] - \mathbb{E}[h(Y)]| \\ &= \sup_{h \in \mathcal{H}} |\mathbb{E}[h(X)] - \mathbb{E}[h(Z)] + \mathbb{E}[h(Z)] - \mathbb{E}[h(Y)]| \\ &\leq \sup_{h \in \mathcal{H}} \{|\mathbb{E}[h(X)] - \mathbb{E}[h(Z)]| + |\mathbb{E}[h(Z)] - \mathbb{E}[h(Y)]|\} \\ &\leq \sup_{h \in \mathcal{H}} |\mathbb{E}[h(X)] - \mathbb{E}[h(Z)]| + \sup_{h \in \mathcal{H}} |\mathbb{E}[h(X)] - \mathbb{E}[h(Z)]| \\ &= d_{\mathcal{H}}(X, Z) + d_{\mathcal{H}}(Z, Y) \end{aligned}$$

dove per l'ultima disuguaglianza abbiamo usato che

$$\sup_{h \in \mathcal{H}} \{f(h) + g(h)\} \leq \sup_{h \in \mathcal{H}} f(h) + \sup_{h \in \mathcal{H}} g(h).$$

□

Andiamo ora a mostrare alcuni esempi di classi separanti.

Proposizione 2.2. *Le seguenti classi sono separanti:*

$$\mathcal{H}_{TV} = \{1_A \mid A \in \mathcal{B}(\mathbb{R})\}.$$

$$\mathcal{H}_K = \{1_{(-\infty, x]} \mid x \in \mathbb{R}\}.$$

$$\mathcal{H}_W = Lip_1(\mathbb{R}).$$

dove $Lip_1(\mathbb{R})$ è l'insieme delle funzioni definite su \mathbb{R} e a valori in \mathbb{R} 1-lipschitziane.

Dimostrazione. Siano X, Y due variabili aleatorie con $\mathbb{E}[|h(X)|], \mathbb{E}[|h(Y)|] < \infty$ per ogni funzione nella classe selezionata.

- Sia $A \in \mathcal{B}(\mathbb{R})$ e $h = 1_A \in \mathcal{H}_{TV}$. Se $\mathbb{E}[h(X)] = \mathbb{E}[h(Y)]$ allora

$$\mathbb{P}_X(A) := \mathbb{P}(X \in A) = \mathbb{E}[1_A(X)] = \mathbb{E}[1_A(Y)] = \mathbb{P}(Y \in A) := \mathbb{P}_Y(A).$$

Poiché la legge di una variabile aleatoria reale è definita sui boreliani, dall'arbitrarietà di A , otteniamo che $\mathbb{P}_X \equiv \mathbb{P}_Y$.

- Sia $z \in \mathbb{R}$ e $h = 1_{(-\infty, z]} \in \mathcal{H}_K$. Se $\mathbb{E}[h(X)] = \mathbb{E}[h(Y)]$ allora

$$\begin{aligned} \mathbb{P}_X((-\infty, z]) &= \mathbb{P}(X \in (-\infty, z]) = \mathbb{E}[1_{(-\infty, z]}(X)] = \\ &= \mathbb{E}[1_{(-\infty, z]}(Y)] = \mathbb{P}(Y \in (-\infty, z]) = \mathbb{P}_Y((-\infty, z]). \end{aligned}$$

Se \mathcal{S} è come nel Lemma B.1 allora abbiamo provato che \mathbb{P}_X e \mathbb{P}_Y coincidono su \mathcal{S} . Essendo \mathcal{S} chiusa per intersezione finita otteniamo $\mathbb{P}_X \equiv \mathbb{P}_Y$ su $\sigma(\mathcal{S})$ (cfr. Teorema B.5). Poiché $\sigma(\mathcal{S})$ è la σ -algebra dei boreliani di \mathbb{R} (cfr. Lemma B.1) otteniamo la tesi.

- Sia $h \in Lip_C(\mathbb{R})$ con $C > 1$ allora $\frac{h}{C} \in Lip_1(\mathbb{R})$ e dunque chiedere che $\mathbb{E}[h(X)] = \mathbb{E}[h(Y)]$ per ogni $h \in Lip_1(\mathbb{R})$ è equivalente a richiedere che $\mathbb{E}[h(X)] = \mathbb{E}[h(Y)]$ per ogni funzione lipschitziana.

Fissato $a \in \mathbb{R}$, per ogni $n \in \mathbb{N}$ sia $h_n : \mathbb{R} \rightarrow \mathbb{R}$ la funzione definita da

$$h_n(x) = \begin{cases} 1 & \text{se } x \leq a \\ -nx + na + 1 & \text{se } a \leq x \leq a + \frac{1}{n} \\ 0 & \text{altrimenti} \end{cases}.$$

Presi $x \geq y \in \mathbb{R}$ si ha

$$|f(x) - f(y)| = \begin{cases} 0 & \text{se } y, x \leq a \vee y, x \geq a + \frac{1}{n} \\ 1 & \text{se } y \leq a \wedge x \geq a + \frac{1}{n} \\ n(x - y) & \text{altrimenti} \end{cases}$$

Notando che

$$y \leq a \wedge x \geq a + \frac{1}{n} \Rightarrow n(x - y) \geq n \left(a + \frac{1}{n} - a \right) = 1$$

otteniamo che $|f(x) - f(y)| \leq n(x - y)$ e dunque h_n è n -lipschitziana.

Utilizzando il teorema di convergenza dominata di Lebesgue e il fatto che $h_n \rightarrow \mathbb{1}_{(-\infty, a]}$ puntualmente otteniamo

$$\begin{aligned} \mathbb{P}(X \leq a) &= \lim_{n \rightarrow +\infty} \mathbb{E}[h_n(X)], \\ \mathbb{P}(Y \leq a) &= \lim_{n \rightarrow +\infty} \mathbb{E}[h_n(Y)]. \end{aligned}$$

Poiché h_n è n -lipschitziana vale $\mathbb{E}[h_n(X)] = \mathbb{E}[h_n(Y)]$ per ogni $n \in \mathbb{N}$. Mettendo insieme le identità precedenti si ha

$$\mathbb{P}(X \leq a) = \mathbb{P}(Y \leq a).$$

Si conclude ripercorrendo la dimostrazione del fatto che \mathcal{H}_K è separante. \square

Grazie alla proposizione precedente e alla Proposizione 2.1 possiamo definire alcune metriche:

- la metrica della **variazione totale** (denotata con d_{TV}) è la metrica ottenuta prendendo in (2.0.1) la classe \mathcal{H}_{TV} ;
- la metrica di **Kolmogorov** (denotata con d_K) è la metrica ottenuta prendendo in (2.0.1) la classe \mathcal{H}_K ;
- la metrica di **Wasserstein** (denotata con d_W) è la metrica ottenuta prendendo in (2.0.1) la classe \mathcal{H}_W .

La proposizione, la cui dimostrazione si può trovare in [33, Proposizione 1.2, p.241], fornisce un confronto (anche se parziale) tra le varie distanze.

Proposizione 2.3. *Siano W, Z due variabili aleatorie. Allora vale*

$$d_K(W, Z) \leq d_{TV}(W, Z).$$

Inoltre, se Z ammette una densità limitata da C assolutamente continua rispetto alla misura di Lebesgue allora vale

$$d_K(W, Z) \leq \sqrt{2Cd_W(W, Z)}.$$

La seguente proposizione, la cui dimostrazione può essere trovata in [42, Teorema 5.10, p. 70], mostra una differente caratterizzazione della distanza della variazione totale e di quella di Wasserstein.

Proposizione 2.4.

$$d_{TV}(X, Y) = \inf \{ \mathbb{P}(X_1 = Y_1) \mid (X_1, Y_1) \text{ coupling di } X \text{ e } Y \}.$$

$$d_W(X, Y) = \inf \{ \mathbb{E} [|X_1 - Y_1|] \mid (X_1, Y_1) \text{ coupling di } X \text{ e } Y \}.$$

Teorema 2.0.1. *Le tre distanze (variazione totale, Kolmogorov e Wasserstein) sono più forti della convergenza debole:*

$$\lim_{n \rightarrow +\infty} d(W_n, Z) = 0 \quad \Rightarrow \quad W_n \Rightarrow Z \text{ per } n \rightarrow +\infty$$

dove \Rightarrow denota la convergenza in distribuzione e d è una qualunque tra le tre distanze menzionate sopra.

3 | Il metodo di Stein

3.1 | L'equazione di Stein

Tipicamente, per mostrare la convergenza di una successione di variabili alla distribuzione gaussiana standard si utilizzano le funzioni caratteristiche. Il metodo di Stein, presentato per la prima volta in [40], sostituisce alle funzioni caratteristiche un **operatore caratteristico**. Il seguente lemma chiarisce il significato di questo termine.

Lemma 3.1.1 (Lemma di Stein).

- (i) Se $S \sim \mathcal{N}(0, 1)$ allora per ogni $f \in AC_{loc}(\mathbb{R})$ (cfr. Definizione A.3) con $\mathbb{E}[|f'(S)|], \mathbb{E}[|Sf(S)|] < \infty$ vale $\mathbb{E}[f'(S)] = \mathbb{E}[Sf(S)]$.
- (ii) Sia S una variabile aleatoria. Se per ogni funzione $f \in C_{tratti}^1(\mathbb{R})$ (cfr. Definizione A.4) limitata con $f'(S)$ ben definita quasi certamente e con $\mathbb{E}[|f'(S)|], \mathbb{E}[|Sf(S)|] < \infty$ vale $\mathbb{E}[f'(S)] = \mathbb{E}[Sf(S)]$, allora $S \sim \mathcal{N}(0, 1)$.

Dimostrazione.

- (i) Ricordando la definizione di distribuzione normale, otteniamo:

$$\mathbb{E}[f'(S)] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} f'(s) e^{-\frac{s^2}{2}} ds.$$

Notiamo che per ogni $s \leq 0$ vale

$$e^{-\frac{s^2}{2}} = \int_{-\infty}^s -xe^{-\frac{x^2}{2}} dx$$

e per $s \geq 0$

$$e^{-\frac{s^2}{2}} = \int_s^{+\infty} xe^{-\frac{x^2}{2}} dx.$$

Mettendo insieme le tre equazioni precedenti otteniamo

$$\begin{aligned} \mathbb{E}[f'(S)] &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^0 f'(s) \left(\int_{-\infty}^s -xe^{-\frac{x^2}{2}} dx \right) ds \\ &\quad + \frac{1}{\sqrt{2\pi}} \int_0^{+\infty} f'(s) \left(\int_s^{+\infty} xe^{-\frac{x^2}{2}} dx \right) ds. \end{aligned}$$

Usando il teorema di Fubini e che $\int_{-\infty}^{+\infty} x e^{-\frac{x^2}{2}} dx = 0$ le precedenti identità diventano

$$\begin{aligned}\mathbb{E}[f'(S)] &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^0 \left(\int_x^0 f'(s) ds \right) (-x) e^{-\frac{x^2}{2}} dx \\ &\quad + \frac{1}{\sqrt{2\pi}} \int_0^{+\infty} \left(\int_0^x f'(s) ds \right) x e^{-\frac{x^2}{2}} dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} (f(x) - f(0)) x e^{-\frac{x^2}{2}} dx = \mathbb{E}[Sf(S)].\end{aligned}$$

(ii) Per la dimostrazione di questo punto si rimanda a [9, Lemma 2.1 p.13]. □

Osservazione 2. Notiamo che dal Corollario A.3 $f \in AC_{loc}(\mathbb{R})$ ammette derivata quasi ovunque. Inoltre, quando S è normale la variabile aleatoria $f'(S)$ è ben definita a meno di un evento di misura nulla. Nel punto (ii), con $f'(S)$ ben definita quasi certamente intendiamo che f' è ben definito in $\{x \mid \mathbb{P}(S = x) > 0\}$.

Definizione 3.1. Sia $Z \sim \mathcal{N}(0, 1)$ e $h : \mathbb{R} \rightarrow \mathbb{R}$ una funzione di Borel con $\mathbb{E}[|h(Z)|] < \infty$. L'**equazione di Stein** associata ad h è la seguente equazione differenziale ordinaria:

$$f'(x) - xf(x) = h(x) - \mathbb{E}[h(Z)]. \quad (3.1.1)$$

Diremo che f è soluzione di (3.1.1) se è localmente assolutamente continua e (3.1.1) vale per tutti i punti in cui è definita la derivata di f (il Corollario A.3 prova che f ammette una derivata quasi ovunque).

Osservazione 3. Nel seguito supporremo che la derivata di una soluzione di (3.1.1) sia definita in ogni punto. Estenderemo la derivata della soluzione nei punti in cui non è definita ponendo

$$f'_h(x) := xf_h(x) + h(x) - \mathbb{E}[h(Z)].$$

Nella proposizione seguente, andremo a caratterizzare la soluzione dell'equazione di Stein ed enunceremo alcune sue proprietà.

Proposizione 3.1.2. Sia $Z \sim \mathcal{N}(0, 1)$ e sia $h : \mathbb{R} \rightarrow \mathbb{R}$ boreliana con $\mathbb{E}[|h(Z)|] < \infty$.

(i) Tutte e sole le soluzioni di (3.1.1) sono della forma:

$$f(x) = ce^{\frac{x^2}{2}} + e^{\frac{x^2}{2}} \int_{-\infty}^x (h(y) - \mathbb{E}[h(Z)]) e^{-\frac{y^2}{2}} dy \quad (3.1.2)$$

dove $c \in \mathbb{R}$.

(ii) L'unica soluzione f di (3.1.1) tale che $\lim_{x \rightarrow \pm\infty} f(x)e^{-\frac{x^2}{2}} = 0$ è

$$f_h(x) := e^{\frac{x^2}{2}} \int_{-\infty}^x (h(y) - \mathbb{E}[h(Z)]) e^{-\frac{y^2}{2}} dy. \quad (3.1.3)$$

(iii) Se h è limitata allora

$$\|f_h\|_\infty \leq \sqrt{\frac{\pi}{2}} \|h - \mathbb{E}[h(Z)]\|_\infty \quad e \quad \|f'_h\|_\infty \leq 2 \|h - \mathbb{E}[h(Z)]\|_\infty. \quad (3.1.4)$$

(iv) Se h è assolutamente continua allora

$$\|f_h\|_\infty \leq 2 \|h'\|_\infty, \quad \|f'_h\|_\infty \leq \sqrt{\frac{2}{\pi}} \|h'\|_\infty \quad e \quad \|f''_h\|_\infty \leq 2 \|h'\|_\infty. \quad (3.1.5)$$

(v) Se $h(x) = h_z(x) = \mathbb{1}_{(-\infty, z]}(x)$ allora

$$\|f_{h_z}\|_\infty \leq \frac{\sqrt{2\pi}}{4} \quad e \quad \|f'_{h_z}\|_\infty \leq 1. \quad (3.1.6)$$

Inoltre, per ogni $u, v, w \in \mathbb{R}$ vale

$$|(w+u)f_{h_z}(w+u) - (w+v)f_{h_z}(w+v)| \leq \left(|w| + \frac{\sqrt{2\pi}}{4} \right) (|u| + |v|). \quad (3.1.7)$$

Nelle precedenti disuguaglianze con $\|f'\|_\infty$ e $\|f''\|_\infty$ denotiamo la norma del sup essenziale mentre con $\|f\|_\infty$ la norma infinito. Notiamo che f_h è assolutamente quindi ammette derivata quasi ovunque (cfr. Corollario A.3).

Dimostrazione. Proveremo solamente (i) e (ii). Essendo le dimostrazioni dei rimanenti punti molto “tecniche”, si rimanda il lettore a [9, Lemma 2.3 e Lemma 2.4, p.16].

(i) Mostriamo che la funzione definita in (3.1.2) è $AC_{loc}(\mathbb{R})$. Siano a, b due numeri reali fissati con $a < b$. Poiché la funzione $ce^{\frac{x^2}{2}}$ è continua non decrescente e possiede una derivata in $L^1([a, b])$ allora $ce^{\frac{x^2}{2}} \in AC([a, b])$ (cfr. Teorema A.4). Dall'arbitrarietà di a, b si conclude che $ce^{\frac{x^2}{2}} \in AC_{loc}(\mathbb{R})$.

Notiamo che

$$\begin{aligned}
& \int_{-\infty}^x |h(y) - \mathbb{E}[h(Z)]| e^{-\frac{y^2}{2}} dy \\
& \leq \int_{-\infty}^x \left\{ |h(y)| e^{-\frac{y^2}{2}} + |\mathbb{E}[h(Z)]| e^{-\frac{y^2}{2}} \right\} dy \\
& \leq \int_{\mathbb{R}} \left\{ |h(y)| e^{-\frac{y^2}{2}} + |\mathbb{E}[h(Z)]| e^{-\frac{y^2}{2}} \right\} dy \\
& = \sqrt{2\pi} \int_{\mathbb{R}} \left\{ \frac{1}{\sqrt{2\pi}} |h(y)| e^{-\frac{y^2}{2}} + \frac{1}{\sqrt{2\pi}} |\mathbb{E}[h(Z)]| e^{-\frac{y^2}{2}} \right\} dy \\
& = \sqrt{2\pi} (\mathbb{E}[|h(Z)|] + |\mathbb{E}[h(Z)]|) \leq 2\sqrt{2\pi} \mathbb{E}[|h(Z)|] < \infty
\end{aligned}$$

ovvero $s(y) := (h(y) - \mathbb{E}[h(Z)]) e^{-\frac{y^2}{2}} \in L^1(\mathbb{R})$. Dal Teorema A.5 si ha che

$$g(x) := \int_{-\infty}^x (h(y) - \mathbb{E}[h(Z)]) e^{-\frac{y^2}{2}} dy \in AC_{loc}(\mathbb{R}).$$

Utilizzando il Corollario A.8 e la Proposizione A.9 otteniamo che f come in (3.1.2) è una funzione localmente assolutamente continua.

Proviamo ora che f come in (3.1.2) è soluzione di (3.1.1). Dal Teorema A.5 e dalla regola di derivazione di Leibnitz otteniamo che

$$\begin{aligned}
f'(x) &= cxe^{\frac{x^2}{2}} + xe^{\frac{x^2}{2}} \int_{-\infty}^x (h(y) - \mathbb{E}[h(Z)]) e^{-\frac{y^2}{2}} dy \\
&\quad + e^{\frac{x^2}{2}} (h(x) - \mathbb{E}[h(Z)]) e^{-\frac{x^2}{2}} \\
&= x \left(ce^{\frac{x^2}{2}} + e^{\frac{x^2}{2}} \int_{-\infty}^x (h(y) - \mathbb{E}[h(Z)]) e^{-\frac{y^2}{2}} dy \right) \\
&\quad + h(x) - \mathbb{E}[h(Z)] \\
&= xf(x) + h(x) - \mathbb{E}[h(Z)].
\end{aligned}$$

Resta da provare che tutte le soluzioni sono della forma data da (3.1.2). Moltiplicando (3.1.1) per il fattore integrante $e^{-\frac{x^2}{2}}$ e applicando la regola di derivazione di Leibnitz si ha

$$\left[e^{-\frac{x^2}{2}} f(x) \right]' = e^{-\frac{x^2}{2}} (h(x) - \mathbb{E}[h(Z)]).$$

Integrando otteniamo (3.1.2).

(ii) Dal teorema di convergenza dominata di Lebesgue otteniamo

$$\int_{-\infty}^x (h(y) - \mathbb{E}[h(Z)]) e^{-\frac{y^2}{2}} dy \rightarrow 0 \text{ per } x \rightarrow +\infty.$$

Dunque, $\lim_{x \rightarrow \pm\infty} f(x)e^{-\frac{x^2}{2}} = 0$ se e solo se $c = 0$.

(iv) Precisiamo solo un dettaglio non presente in [9]. Poiché f_h è soluzione di (3.1.1) otteniamo

$$f'_h(x) = x f_h(x) + h(x) - \mathbb{E}[h(Z)]. \quad (3.1.8)$$

Ora f_h e h sono funzioni assolutamente continue, dunque, ammettono derivata quasi certa (cfr. Corollario A.3). Derivando (3.1.8) otteniamo che f_h ammette una derivata seconda quasi certa. □

Grazie all'introduzione dell'equazione di Stein, possiamo calcolare la distanza tra una variabile aleatoria e una variabile normale in una maniera differente rispetto a (2.0.1).

Proposizione 3.1.3. *Sia \mathcal{H} una classe separante di funzioni boreliane. Siano S una variabile aleatoria e $Z \sim \mathcal{N}(0, 1)$ tali per cui $\mathbb{E}[|h(S)|], \mathbb{E}[|h(Z)|] < \infty$. Allora vale*

$$d_{\mathcal{H}}(S, Z) = \sup_{h \in \mathcal{H}} |\mathbb{E}[f'_h(S) - S f_h(S)]| \quad (3.1.9)$$

dove f_h è la funzione definita in (3.1.3).

Dimostrazione. Essendo f_h soluzione dell'equazione di Stein otteniamo

$$f'_h(S) - S f_h(S) = h(S) - \mathbb{E}[h(Z)]. \quad (3.1.10)$$

Siccome $\mathbb{E}[|h(S)|], \mathbb{E}[|h(Z)|] < \infty$ il membro di destra è una variabile aleatoria ben definita ed integrabile. Quindi $\mathbb{E}[|f'_h(S) - S f_h(S)|] < \infty$. Utilizzando la definizione di $d_H(S, Z)$ (cfr. Definizione 2.2) si ha

$$\begin{aligned} d_{\mathcal{H}}(S, Z) &:= \sup_{h \in \mathcal{H}} |\mathbb{E}[h(S)] - \mathbb{E}[h(Z)]| \\ &= \sup_{h \in \mathcal{H}} |\mathbb{E}[h(S)] - \mathbb{E}[f'_h(S) - S f_h(S) - h(S)]| \\ &= \sup_{h \in \mathcal{H}} |\mathbb{E}[f'_h(S) - S f_h(S)]| \end{aligned}$$

dove la penultima equazione segue da (3.1.10). □

Grazie alle proprietà della soluzione dell'equazione di Stein (Proposizione 3.1.2) possiamo specializzare la Proposizione 3.1.3 al caso della distanza di Wasserstein (Teorema 3.1.4) e a quella di Kolmogorov (Teorema 3.1.5).

Teorema 3.1.4. *Sia S una variabile aleatoria integrabile e $Z \sim \mathcal{N}(0, 1)$. Allora si ha*

$$d_W(S, Z) \leq \sup_{f \in \mathcal{F}_W} |\mathbb{E}[f'(S) - Sf(S)]| \quad (3.1.11)$$

dove

$$\mathcal{F}_W = \left\{ f \text{ derivabile due volte con derivate q.o.} \right. \\ \left. \|f\|_\infty, \|f''\|_\infty \leq 2 \text{ e } \|f'\|_\infty \leq \sqrt{\frac{2}{\pi}} \right\}. \quad (3.1.12)$$

Dimostrazione. Notiamo che per ogni $f \in \mathcal{F}_W$, f è limitata e dunque $\mathbb{E}[|f(S)|] < \infty$. Usando la Proposizione 3.1.3 con $\mathcal{H} = \mathcal{H}_W = Lip_1(\mathbb{R})$ otteniamo

$$d_W(S, Z) = \sup_{h \in \mathcal{H}_W} |\mathbb{E}[f'_h(S) - Sf_h(S)]| = \sup_{f \in \tilde{\mathcal{F}}_W} |\mathbb{E}[f'(S) - Sf(S)]|$$

dove

$$\tilde{\mathcal{F}}_W = \{f_h \mid h \in Lip_1(\mathbb{R})\}.$$

Se proviamo che $\tilde{\mathcal{F}}_W \subseteq \mathcal{F}_W$, la tesi segue dalla monotonia dell'esterno superiore. Se $f_h \in \tilde{\mathcal{F}}_W$ allora f_h è soluzione dell'equazione di Stein per qualche $h \in Lip_1(\mathbb{R})$. Poiché $h \in Lip_1(\mathbb{R})$ allora h è assolutamente continua (cfr. Lemma A.1). Dalla Proposizione 3.1.2(iv) sappiamo che f è derivabile con derivata seconda definita quasi certamente ed inoltre

$$\|f\|_\infty, \|f''\|_\infty \leq 2 \|h'\|_\infty \leq 2 \text{ e } \|f'\|_\infty \leq \sqrt{\frac{2}{\pi}} \|h'\|_\infty \leq \sqrt{\frac{2}{\pi}}$$

dove abbiamo usato l'Osservazione 13 (cfr. Sezione A.3). □

Teorema 3.1.5. *Sia S una variabile aleatoria integrabile e $Z \sim \mathcal{N}(0, 1)$. Allora si ha*

$$d_K(S, Z) = \sup_{f \in \mathcal{F}_K} |\mathbb{E}[f'(S) - Sf(S)]| \quad (3.1.13)$$

dove

$$\mathcal{F}_K = \left\{ f \in AC_{loc}(\mathbb{R}) \mid \|f\|_\infty \leq \frac{\sqrt{2\pi}}{4} \text{ e } \|f'\|_\infty \leq 1 \right\}. \quad (3.1.14)$$

Dimostrazione. Notiamo che per ogni $f \in \mathcal{F}_K$, f è limitata e dunque $\mathbb{E}[|f(S)|] < \infty$. Usando la Proposizione 3.1.3 con $\mathcal{H} = \mathcal{H}_K$ (cfr. Capitolo 2) otteniamo

$$d_K(S, Z) \leq \sup_{h \in \mathcal{H}_K} |\mathbb{E}[f'_h(S) - Sf_h(S)]| = \sup_{f \in \tilde{\mathcal{F}}_K} |\mathbb{E}[f'(S) - Sf(S)]|$$

dove

$$\tilde{\mathcal{F}}_K = \{f_{h_z} \in AC_{loc}(\mathbb{R}) \mid h_z = \mathbb{1}_{(-\infty, z]}\}.$$

Se proviamo che $\tilde{\mathcal{F}}_K \subseteq \mathcal{F}_K$, la tesi segue dalla monotonia dell'estremo superiore. Se $f \in \tilde{\mathcal{F}}_K$ allora f è soluzione dell'equazione di Stein per qualche $h = \mathbb{1}_{(-\infty, z]}$. Dalla Proposizione 3.1.2 (v) otteniamo che

$$\|f_{h_z}\|_\infty \leq \frac{\sqrt{2\pi}}{4} \quad \text{e} \quad \|f'_{h_z}\|_\infty \leq 1.$$

La tesi segue dalla monotonia dell'estremo superiore. □

3.2 | Leave one out

Questo primo metodo si basa sull'idea presente nel lavoro di Stein [39] e prende il nome di metodo *leave one out*. Riprendendo le idee presentate in [9, sez. 1.3] possiamo ottenere il seguente risultato

Teorema 3.2.1. *Siano X_1, \dots, X_n variabili aleatorie indipendenti, centrate con momenti assoluti terzi finiti. Se $\sum_{i=1}^n \text{Var}(X_i) = 1$ allora*

$$d_W(S, Z) \leq 4 \sum_{i=1}^n \mathbb{E}[|X_i|^3]$$

dove $S = \sum_{i=1}^n X_i$ e $Z \sim \mathcal{N}(0, 1)$.

Dimostrazione. Per ogni $i \in [n]$ definiamo la seguente variabile aleatoria detta **leave one out**

$$S^{(i)} := S - X_i.$$

Sia \mathcal{F}_W la classe definita in (3.1.12) e fissiamo $f \in \mathcal{F}_W$. Abbiamo

$$\begin{aligned} \mathbb{E}[Sf(S)] &= \sum_{i=1}^n \mathbb{E}[X_i f(S)] = \sum_{i=1}^n \mathbb{E}\left[X_i f\left(S^{(i)} + X_i\right)\right] \\ &= \mathbb{E}\left[\sum_{i=1}^n X_i f(S^{(i)}) + X_i^2 \int_0^1 f'(S^{(i)} + uX_i) du\right] \\ &= \mathbb{E}\left[\sum_{i=1}^n X_i^2 \int_0^1 f'(S^{(i)} + uX_i) du\right] \end{aligned} \tag{3.2.1}$$

dove per l'ultima uguaglianza abbiamo usato $\mathbb{E}[X_i] = 0$ e $X_i \perp S^{(i)}$.

D'altra parte, denotando con $\sigma_i^2 = \text{Var}(X_i)$ otteniamo $\sum_{i=1}^n \sigma_i^2 = 1$. Dunque, si ha

$$\begin{aligned} \mathbb{E}[f'(S)] &= \mathbb{E} \left[\sum_{i=1}^n \sigma_i^2 f'(S) \right] = \mathbb{E} \left[\sum_{i=1}^n \sigma_i^2 \left(f'(S^{(i)}) + f'(S) - f'(S^{(i)}) \right) \right] \\ &= \mathbb{E} \left[\sum_{i=1}^n X_i^2 f'(S^{(i)}) \right] + \mathbb{E} \left[\sum_{i=1}^n \sigma_i^2 \left(f'(S) - f'(S^{(i)}) \right) \right] \end{aligned} \quad (3.2.2)$$

dove l'ultima disuguaglianza segue dall'indipendenza tra X_i e $S^{(i)}$. Sottraendo (3.2.1) da (3.2.2) abbiamo

$$\begin{aligned} \mathbb{E}[f'(S) - S f(S)] &= \mathbb{E} \left[\sum_{i=1}^n \left\{ X_i^2 f'(S^{(i)}) + \sigma_i^2 \left(f'(S) - f'(S^{(i)}) \right) \right\} \right] \\ &\quad - \mathbb{E} \left[\sum_{i=1}^n \left\{ X_i^2 \int_0^1 f'(S^{(i)} + uX_i) du \right\} \right] \\ &= \mathbb{E} \left[\sum_{i=1}^n \left\{ X_i^2 \int_0^1 \left(f'(S^{(i)}) - f'(S^{(i)} + uX_i) \right) du \right\} \right] \\ &\quad + \mathbb{E} \left[\sum_{i=1}^n \left\{ \sigma_i^2 \left(f'(S) - f'(S^{(i)}) \right) \right\} \right]. \end{aligned}$$

Poiché $f \in \mathcal{F}_W$, se $u \in [0, 1]$ allora si ha

$$\left| f'(S^{(i)}) - f'(S^{(i)} + uX_i) \right| \leq \|f''\|_\infty |uX_i| \leq 2|X_i|$$

e

$$\left| f'(S) - f'(S^{(i)}) \right| \leq \|f''\|_\infty |X_i| \leq 2|X_i|.$$

Usando questa limitazione e la disuguaglianza triangolare abbiamo

$$\begin{aligned}
|\mathbb{E}[f'(S) - Sf(S)]| &\leq \mathbb{E} \left[\sum_{i=1}^n \left\{ |X_i|^2 \int_0^1 |f'(S^{(i)}) - f'(S^{(i)} + uX_i)| du \right\} \right] \\
&\quad + \mathbb{E} \left[\sum_{i=1}^n \left\{ \sigma_i^2 |f'(S) - f'(S^{(i)})| \right\} \right] \\
&\leq \mathbb{E} \left[\sum_{i=1}^n \left\{ |X_i|^2 \int_0^1 2|X_i| du \right\} \right] + \mathbb{E} \left[\sum_{i=1}^n \left\{ \sigma_i^2 2|X_i| \right\} \right] \\
&= 2 \sum_{i=1}^n \mathbb{E} [|X_i|^2 |X_i|] + 2 \sum_{i=1}^n \mathbb{E} [\sigma_i^2 |X_i|] \\
&= 2 \sum_{i=1}^n \mathbb{E} [|X_i|^3] + 2 \sum_{i=1}^n \mathbb{E} [\sigma_i^2 |X_i|].
\end{aligned}$$

Ma per ogni $i \in [n]$ vale

$$\mathbb{E} [\sigma_i^2 |X_i|] = \mathbb{E} [|X_i|^2] \mathbb{E} [|X_i|] \leq \mathbb{E} [|X_i|^3]^{\frac{2}{3}} \mathbb{E} [|X_i|^3]^{\frac{1}{3}} = \mathbb{E} [|X_i|^3]$$

dove per l'ultima disuguaglianza abbiamo utilizzato la disuguaglianza di Hölder. Mettendo insieme le due disuguaglianze precedenti otteniamo la tesi. \square

L'idea alla base del teorema precedente può essere usata anche sostituendo all'indipendenza della variabili aleatorie una condizione di indipendenza più debole. Ad esempio abbiamo ottenuto il seguente risultato [1, Teorema 3.4].

Teorema 3.2.2. *Sia \mathcal{I} un insieme di indici e sia $(\xi_i)_{i \in \mathcal{I}}$ una \mathcal{I} -successione di variabili aleatorie centrate con momenti assoluti terzi finiti. Supponiamo che per ogni $i \in \mathcal{I}$ esistano degli insiemi $S_i \subseteq T_i \subseteq \mathcal{I}$ tali che per ogni i si abbia*

$$\xi_i \perp (\xi_j)_{j \in S_i^c}, \quad (3.2.3a)$$

$$(\xi_j)_{j \in S_i} \perp (\xi_j)_{j \in T_i^c}. \quad (3.2.3b)$$

Posto $W = \sum_{i \in \mathcal{I}} \xi_i$, se $\text{Var}(W) = 1$ allora vale

$$d_W(W, Z) \leq \sum_{i \in \mathcal{I}} \left\{ 2\mathbb{E} [|\xi_i \eta_i \tau_i|] + 2|\mathbb{E}[\xi_i \eta_i]| \mathbb{E}[|\tau_i|] + \mathbb{E} [|\xi_i \eta_i^2|] \right\}$$

dove $Z \sim \mathcal{N}(0, 1)$, $\eta_i = \sum_{j \in S_i} \xi_j$ e $\tau_i = \sum_{j \in T_i} \xi_j$.

Dimostrazione. Fissato $h \in Lip_1(\mathbb{R})$ sia $f = f_h$ la soluzione dell'equazione di Stein associata ad h . Allora

$$\begin{aligned}\mathbb{E}[Wf(W)] &= \sum_{i \in \mathcal{I}} \mathbb{E}[\xi_i f(W)] \\ &= \sum_{i \in \mathcal{I}} \mathbb{E}[\xi_i (f(W) - f(W - \eta_i))]\end{aligned}\tag{3.2.4}$$

dove per l'ultima uguaglianza abbiamo usato che $\xi_i \perp (\xi_j)_{j \in S_i^c}$ e dunque $\xi_i \perp (W - \eta_i)$. Quindi, essendo ξ_i centrato, otteniamo

$$\mathbb{E}[\xi_i f(W - \eta_i)] = \mathbb{E}[\xi_i] \mathbb{E}[f(W - \eta_i)] = 0.$$

La (3.2.4) può essere riscritta come

$$\begin{aligned}\mathbb{E}[Wf(W)] &= \sum_{i \in \mathcal{I}} \mathbb{E}[\xi_i (f(W) - f(W - \eta_i) - \eta_i f'(W))] \\ &\quad + \mathbb{E}\left[\sum_{i \in \mathcal{I}} \{\xi_i \eta_i f'(W)\}\right]\end{aligned}\tag{3.2.5}$$

Ricordando che $\text{Var}(W) = 1$ otteniamo

$$\begin{aligned}1 = \text{Var}(W) &= \sum_{i, j \in \mathcal{I}} \mathbb{E}[\xi_i \xi_j] = \sum_{i \in \mathcal{I}} \left(\sum_{j \in S_i} \mathbb{E}[\xi_i \xi_j] + \sum_{j \notin S_i} \mathbb{E}[\xi_i \xi_j] \right) \\ &= \sum_{i \in \mathcal{I}} \sum_{j \in S_i} \mathbb{E}[\xi_i \xi_j] = \sum_{i \in \mathcal{I}} \mathbb{E}[\xi_i \eta_i]\end{aligned}\tag{3.2.6}$$

dove la penultima identità segue dal fatto che per ogni $i \in \mathcal{I}$ vale $\xi_i \perp (\xi_j)_{j \in S_i^c}$. Quindi, essendo ξ_i centrato, otteniamo

$$\sum_{j \notin S_i} \mathbb{E}[\xi_i \xi_j] = \sum_{j \notin S_i} \mathbb{E}[\xi_i] \mathbb{E}[\xi_j] = 0.$$

Utilizzando (3.2.6) si ha

$$\mathbb{E}[f'(W)] = \left(\sum_{i \in \mathcal{I}} \mathbb{E}[\xi_i \eta_i] \right) \mathbb{E}[f'(W)]\tag{3.2.7}$$

Combinando (3.2.5) e (3.2.7) otteniamo

$$\begin{aligned}\mathbb{E}[f'(W) - Wf(W)] &= \mathbb{E}\left[\sum_{i \in \mathcal{I}} f'(W) (\mathbb{E}[\xi_i \eta_i] - \xi_i \eta_i)\right] \\ &\quad - \sum_{i \in \mathcal{I}} \mathbb{E}[\xi_i (f(W) - f(W - \eta_i) - \eta_i f'(W))].\end{aligned}\tag{3.2.8}$$

Ora

$$\mathbb{E}[f'(W - \tau_i)(\mathbb{E}[\xi_i \eta_i] - \xi_i \eta_i)] = \mathbb{E}[f'(W - \tau_i)] \mathbb{E}[\xi_i \eta_i] - \mathbb{E}[f'(W - \tau_i) \xi_i \eta_i]$$

Poiché $(\xi_j)_{j \in S_i} \perp (\xi_j)_{j \in T_i^c}$ allora $\eta_i \perp (W - \tau_i)$. Da $\xi_i \perp (\xi_j)_{j \in S_i^c}$, poiché $S_i \subseteq T_i$, otteniamo $\xi_i \perp (\xi_j)_{j \in T_i^c}$ e quindi $\xi_i \perp (W - \tau_i)$. Mettendo insieme quanto detto otteniamo $\xi_i \eta_i \perp (W - \tau_i)$ da cui

$$\mathbb{E}[f'(W - \tau_i)(\mathbb{E}[\xi_i \eta_i] - \xi_i \eta_i)] = 0. \quad (3.2.9)$$

Utilizzando (3.2.9), (3.2.8) diventa

$$\begin{aligned} & \mathbb{E}[f'(W) - W f(W)] \\ &= \mathbb{E} \left[\sum_{i \in \mathcal{I}} (f'(W) - f'(W - \tau_i)) (\mathbb{E}[\xi_i \eta_i] - \xi_i \eta_i) \right. \\ & \quad \left. - \sum_{i \in \mathcal{I}} \mathbb{E}[\xi_i (f(W) - f(W - \eta_i) - \eta_i f'(W))] \right] \\ &= \mathbb{E} \left[\sum_{i \in \mathcal{I}} \{ f''(W_i^1) \tau_i (\mathbb{E}[\xi_i \eta_i] - \xi_i \eta_i) \} \right] - \sum_{i \in \mathcal{I}} \mathbb{E} \left[\xi_i \frac{\eta_i^2}{2} f''(W_i^2) \right]. \end{aligned}$$

dove per ogni $i \in \mathcal{I}$ W_i^1 è una variabile aleatoria con valori compresi tra W e $W - \tau_i$ mentre W_i^2 ha valori compresi tra W e $W - \eta_i$.

Applicando il valore assoluto e ricordando la Proposizione 3.1.2 (iv) otteniamo

$$|\mathbb{E}[f'(W) - W f(W)]| \leq \sum_{i \in \mathcal{I}} \{ 2\mathbb{E}[\xi_i \tau_i \eta_i] + 2|\mathbb{E}[\xi_i \eta_i]| \mathbb{E}[|\tau_i|] + \mathbb{E}[|\xi_i \eta_i^2|] \}.$$

□

3.3 | Exchangeable pairs

In questa sezione, introdurremo la definizione di coppia scambiabile (in inglese *exchangeable pair*). Tale definizione è stata presentata per la prima volta da Stein in [40].

Definizione 3.2. Siano S, S' due variabili aleatorie definite sullo stesso spazio di probabilità. Diremo che (S, S') è una **coppia scambiabile** se $(S, S') \sim (S', S)$. Inoltre, se $\mathbb{E}[|S'|] < \infty$ ed esiste $\lambda \in (0, 1)$ tale che

$$\mathbb{E}[S'|S] = (1 - \lambda)S \quad (3.3.1)$$

diremo che (S, S') è una λ -Stein coppia.

Il Teorema 3.3.1, che segue, mostra come attraverso l'utilizzo delle coppie scambiabili possiamo calcolare in maniera differente il membro destro di (3.1.11). Per la dimostrazione di questo teorema, così come pabbiamo utilizzato alcune idee prese da [9, sez. 2.3.2] e [33, sez. 3.3].

Teorema 3.3.1. *Sia S una variabile aleatoria con momento assoluto terzo finito. Se (S, S') è una λ -Stein coppia allora*

$$d_W(S, Z) \leq \sqrt{\frac{2}{\pi}} \mathbb{E} \left[\left| 1 - \frac{\mathbb{E}[(S' - S)^2 | S]}{2\lambda} \right| \right] + \frac{1}{3\lambda} \mathbb{E}[|(S - S')^3|]$$

dove $Z \sim \mathcal{N}(0, 1)$.

Dimostrazione. Sia \mathcal{F}_W è la classe definita in (3.1.12) e fissiamo $f \in \mathcal{F}_W$. Posta

$$g : \mathbb{R}^2 \rightarrow \mathbb{R}, \quad g(x, y) := (x - y)(f(y) + f(x))$$

notiamo che

$$\begin{aligned} \mathbb{E}[|g(S, S')|] &= \mathbb{E}[|(S - S')(f(S) + f(S'))|] \\ &\leq \mathbb{E}[|Sf(S)| + |Sf'(S)| + |S'f(S)| + |S'f(S')|] \\ &\leq \left(2 + \sqrt{\frac{2}{\pi}} \right) (\mathbb{E}[|S|] + \mathbb{E}[|S'|]) \end{aligned}$$

dove l'ultima disuguaglianza segue dalla definizione di \mathcal{F}_W .

Poiché $\mathbb{E}[|S|], \mathbb{E}[|S'|] < \infty$ otteniamo che $\mathbb{E}[|g(S, S')|] < \infty$. Ma $(S, S') \sim (S', S)$ dunque

$$\mathbb{E}[g(S, S')] = \mathbb{E}[g(S', S)] = -\mathbb{E}[g(S, S')]$$

dove abbiamo usato, per l'ultima uguaglianza, che g è antisimmetrica. Abbiamo quindi

$$\begin{aligned} 0 &= \mathbb{E}[g(S, S')] = \mathbb{E}[(S - S')(f(S) + f(S'))] \\ &= \mathbb{E}[(S - S')(f(S') - f(S))] + 2\mathbb{E}[f(S)(S - S')] \\ &= \mathbb{E}[(S - S')(f(S') - f(S))] + 2\mathbb{E}[\mathbb{E}[f(S)(S - S') | S]] \\ &= \mathbb{E}[(S - S')(f(S') - f(S))] + 2\mathbb{E}[f(S)\mathbb{E}[(S - S') | S]] \end{aligned} \tag{3.3.2}$$

dove la penultima uguaglianza segue da (B.3.1) mentre l'ultima da (B.3.3). Utilizzando (3.3.1), (3.3.2) diventa

$$0 = \mathbb{E}[(S - S')(f(S') - f(S))] + 2\lambda \mathbb{E}[Sf(S)]$$

cioè,

$$\mathbb{E}[Sf(S)] = \frac{1}{2\lambda} \mathbb{E}[(S' - S)(f(S') - f(S))]. \tag{3.3.3}$$

Utilizzando lo sviluppo di Taylor si ha

$$\mathbb{E}[Sf(S)] = \frac{1}{2\lambda} \mathbb{E} \left[f'(S)(S' - S)^2 + \frac{f''(\tilde{S})(S' - S)^3}{3} \right] \quad (3.3.4)$$

dove \tilde{S} è una variabile aleatoria che assume valori fra S e S' . Da (3.3.4) otteniamo

$$\mathbb{E}[f'(S) - Sf(S)] = \mathbb{E} \left[f'(S) \left(1 - \frac{(S' - S)^2}{2\lambda} \right) \right] + \frac{1}{6\lambda} \mathbb{E}[f''(\tilde{S})(S' - S)^3].$$

Combinando (B.3.1) e (B.3.3) otteniamo

$$\begin{aligned} \mathbb{E} [f'(S)(S' - S)^2] &= \mathbb{E} [\mathbb{E} [f'(S)(S' - S)^2 \mid S]] \\ &= \mathbb{E} [f'(S) \mathbb{E} [(S' - S)^2 \mid S]]. \end{aligned}$$

Mettendo insieme le equazioni ottenute si ha

$$\begin{aligned} |\mathbb{E}[f'(S) - Sf(S)]| &\leq \mathbb{E} \left[|f'(S)| \left| 1 - \frac{\mathbb{E}[(S' - S)^2 \mid S]}{2\lambda} \right| \right] \\ &\quad + \frac{1}{6\lambda} \mathbb{E} [|f''(\tilde{S})| |S' - S|^3]. \end{aligned}$$

Ricordando le proprietà di \mathcal{F}_W otteniamo la tesi. □

Grazie al Lemma 3.3.2, che segue, specializzeremo il teorema precedente al caso di somme di variabili aleatorie indipendenti.

Lemma 3.3.2. *Siano X_1, \dots, X_n variabili aleatorie indipendenti centrate con media zero tali che $\sum_{i=1}^n \text{Var}(X_i) = 1$. Sia (Y_1, \dots, Y_n) una copia indipendente di (X_1, \dots, X_n) e sia I una variabile aleatoria uniforme su $\{1, \dots, n\}$ indipendente da $(X_1, \dots, X_n, Y_1, \dots, Y_n)$. Posto*

$$S = \sum_{i=1}^n X_i \quad e \quad S' := S - X_I + Y_I$$

si ha che (S, S') è una $\frac{1}{n}$ -Stein coppia.

Dimostrazione. Poiché per ogni $i \in [n]$ si ha $X_i \sim Y_i$ allora $(S, S') \sim (S', S)$. Dunque (S, S') è una coppia scambiabile. Notiamo che

$$\begin{aligned} \mathbb{E}[S' - S \mid S] &= \mathbb{E}[Y_I - X_I \mid S] = \sum_{i=1}^n \mathbb{E} [\mathbb{E}[Y_I - X_I \mid S] \mid I = i] \mathbb{P}(I = i) \\ &= \sum_{i=1}^n \frac{1}{n} \mathbb{E}[Y_i - X_i \mid S] \end{aligned}$$

dove abbiamo usato la legge dell'aspettazione totale (cfr. Teorema B.9). Poiché $Y_i \perp\!\!\!\perp S$ si ha $\mathbb{E}[Y_i | S] = \mathbb{E}[Y_i] = 0$ (cfr. (B.3.4)) e dunque

$$\mathbb{E}[S' - S | S] = -\frac{1}{n} \sum_{i=1}^n \mathbb{E}[X_i | S] = -\frac{1}{n} \mathbb{E}[S | S] = -\frac{S}{n}.$$

□

Siamo ora pronti per applicare i risultati alla somma di variabili aleatorie indipendenti.

Teorema 3.3.3. *Siano X_1, \dots, X_n variabili aleatorie indipendenti, centrate e con momenti assoluti terzi finiti. Se $\sum_{i=1}^n \text{Var}(X_i) = 1$ allora*

$$d_W(S, Z) \leq \sqrt{\frac{1}{2\pi}} \mathbb{E} \left[\left| 1 - \mathbb{E} \left[\sum_{i=1}^n X_i^2 \mid S \right] \right| \right] + \frac{8}{3} \sum_{i=1}^n \mathbb{E}[|X_i|^3]$$

dove $S = \sum_{i=1}^n X_i$ e $Z \sim \mathcal{N}(0, 1)$.

Dimostrazione. Sia S' come nel lemma precedente, allora (S, S') è una $\frac{1}{n}$ -Stein coppia. Applicando il Teorema 3.3.1 con $\lambda = \frac{1}{n}$ otteniamo

$$d_W(S, Z) \leq \sqrt{\frac{2}{\pi}} \mathbb{E} \left[\left| 1 - \frac{n}{2} \mathbb{E}[(S - S')^2 | S] \right| \right] + \frac{n}{3} \mathbb{E}[|(S - S')^3|]. \quad (3.3.5)$$

Ora

$$\begin{aligned} \mathbb{E}[(S - S')^2 | S] &= \mathbb{E}[(X_I - Y_I)^2 | S] \\ &= \sum_{i=1}^n \mathbb{E} \left[\mathbb{E}[(X_i - Y_i)^2 | S] \mid I = i \right] \mathbb{P}(I = i) \\ &= \sum_{i=1}^n \frac{1}{n} \mathbb{E}[X_i^2 + Y_i^2 - 2X_i Y_i | S] \end{aligned}$$

dove abbiamo usato la legge dell'aspettazione totale. Da (B.3.4), poiché $Y_i \perp\!\!\!\perp S$ otteniamo

$$\begin{aligned} \mathbb{E}[Y_i | S] &= \mathbb{E}[Y_i] = 0, \\ \mathbb{E}[Y_i^2 | S] &= \mathbb{E}[Y_i^2]. \end{aligned}$$

Da (B.3.3) segue

$$\mathbb{E}[X_i Y_i | S] = X_i \mathbb{E}[Y_i | S] = 0.$$

Abbiamo quindi che

$$\mathbb{E}[(S - S')^2 | S] = \frac{1}{n} \left(1 + \mathbb{E} \left[\sum_{i=1}^n X_i^2 \middle| S \right] \right).$$

Dunque, il primo addendo nel membro sinistro di (3.3.5) si può scrivere come

$$\sqrt{\frac{2}{\pi}} \mathbb{E} \left[\left| 1 - \frac{n}{2} \mathbb{E}[(S - S')^2 | S] \right| \right] = \sqrt{\frac{1}{2\pi}} \mathbb{E} \left[\left| 1 - \mathbb{E} \left[\sum_{i=1}^n X_i^2 \middle| S \right] \right| \right]. \quad (3.3.6)$$

Per quanto riguarda il secondo addendo in (3.3.5), utilizzando la legge dell'aspettazione totale, si ha

$$\begin{aligned} \mathbb{E} [|S - S'|^3] &= \mathbb{E} [|X_I - Y_I|^3] = \sum_{i=1}^n \mathbb{E} [|X_I - Y_I|^3 | I = i] \mathbb{P}(I = i) \\ &= \frac{1}{n} \sum_{i=1}^n \mathbb{E} [|X_i - Y_i|^3] \leq \frac{1}{n} \sum_{i=1}^n \mathbb{E} [(|X_i| + |Y_i|)^3] \\ &= \frac{1}{n} \sum_{i=1}^n \{ \mathbb{E}[|X_i|^3] + 3\mathbb{E}[|X_i|^2|Y_i|] + 3\mathbb{E}[|X_i||Y_i|^2] + \mathbb{E}[|Y_i|^3] \}. \end{aligned}$$

Poiché X_i e Y_i sono indipendenti con la stessa legge vale

$$\begin{aligned} \mathbb{E} [|S - S'|^3] &\leq \frac{1}{n} \sum_{i=1}^n \{ 2\mathbb{E}[|X_i|^3] + 6\mathbb{E}[|X_i|^2]\mathbb{E}[|X_i|] \} \\ &\leq \frac{1}{n} \sum_{i=1}^n \{ 2\mathbb{E}[|X_i|^3] + 6\mathbb{E}[|X_i|^3] \} = \frac{8}{n} \sum_{i=1}^n \mathbb{E}[|X_i|^3]. \end{aligned} \quad (3.3.7)$$

Infatti

$$\mathbb{E} [|X_i|^2] \mathbb{E} [|X_i|] \leq \mathbb{E} [|X_i|^3]^{\frac{2}{3}} \mathbb{E} [|X_i|^3]^{\frac{1}{3}} = \mathbb{E} [|X_i|^3]$$

dove abbiamo utilizzato la disuguaglianza di Hölder.

Sostituendo in (3.3.5), le espressioni ottenute in (3.3.6) e (3.3.7) si ottiene

$$d_W(S, Z) \leq \sqrt{\frac{1}{2\pi}} \mathbb{E} \left[\left| 1 - \mathbb{E} \left[\sum_{i=1}^n X_i^2 \middle| S \right] \right| \right] + \frac{8}{3} \sum_{i=1}^n \mathbb{E}[|X_i|^3].$$

□

3.4 | Zero-bias

Il prossimo metodo che presenteremo fa ricorso alla trasformata zero-bias. Questo concetto è stato introdotto per la prima volta da Goldstein e Reinert in [25].

Definizione 3.3. Sia S una variabile aleatoria con media zero e $\text{Var}(S) = \sigma^2$. Diremo che una variabile aleatoria S^* ha la distribuzione **zero-bias** associata ad S se non ha atomi e per ogni $f \in \text{AC}_{loc}(\mathbb{R})$ (cfr. Definizione A.3) con $\mathbb{E}[|Sf(S)|], \mathbb{E}[|f'(S^*)|] < \infty$ vale

$$\mathbb{E}[Sf(S)] = \sigma^2 \mathbb{E}[f'(S^*)].$$

Lemma 3.4.1.

- (i) Se $S \sim \mathcal{N}(0, \sigma^2)$ allora per ogni $f \in \text{AC}_{loc}(\mathbb{R})$ con $\mathbb{E}[|f'(S)|], \mathbb{E}[|Sf(S)|] < \infty$ vale $\sigma^2 \mathbb{E}[f'(S)] = \mathbb{E}[Sf(S)]$.
- (ii) Sia S una variabile aleatoria senza atomi. Se per ogni funzione $f \in C_{tratti}^1(\mathbb{R})$ (cfr. Definizione A.4) limitata con $\mathbb{E}[|f'(S)|], \mathbb{E}[|Sf(S)|] < \infty$ vale $\sigma^2 \mathbb{E}[f'(S)] = \mathbb{E}[Sf(S)]$, allora $S \sim \mathcal{N}(0, \sigma^2)$.

Dimostrazione. Il lemma segue dal Lemma 3.1.1 con un semplice cambio di variabili. □

Il seguente teorema, oltre a provare l'esistenza e unicità della distribuzione zero-bias, mostra un modo per calcolarla esplicitamente. Per la dimostrazione del teorema si guardi [9, Proposizione 2.1, p. 27].

Teorema 3.4.2. Sia S una variabile aleatoria con media zero e varianza σ^2 . La distribuzione zero-bias associata ad S esiste unica. In particolare, S^* è una variabile continua con densità di probabilità data da

$$p^*(x) = \frac{1}{\sigma^2} \mathbb{E}[S 1_{(x, +\infty)}(S)], \quad x \in \mathbb{R}. \quad (3.4.1)$$

Il Teorema 3.4.3, che segue, mostra come attraverso l'uso della distribuzione zero-bias possiamo limitare il membro di destra di (3.1.11).

Teorema 3.4.3. Sia S una variabile aleatoria con media zero e varianza 1 e sia S^* una variabile aleatoria definita sullo stesso spazio di S con distribuzione zero-bias associata ad S . Allora vale

$$d_W(S, Z) \leq 2 \mathbb{E}[|S^* - S|]$$

dove $Z \sim \mathcal{N}(0, 1)$.

Dimostrazione. Sia \mathcal{F}_W la classe definita in (3.1.12). Dal Teorema 3.1.4 otteniamo

$$d_W(S, Z) \leq \sup_{f \in \mathcal{F}_W} |\mathbb{E}[f'(S) - Sf(S)]|.$$

Utilizzando la definizione di distribuzione zero-bias e sviluppando in serie di Taylor otteniamo

$$d_W(S, Z) \leq \sup_{f \in \mathcal{F}_W} |\mathbb{E}[f'(S) - f(S^*)]| = \sup_{f \in \mathcal{F}_W} \left| \mathbb{E} \left[f''(\tilde{S})(S - S^*) \right] \right|$$

dove abbiamo usato che f ammette una derivata seconda quasi certa e \tilde{S} è una variabile aleatoria che assume valori tra S e S^* . Ricordando che $\|f''\|_\infty \leq 2$ otteniamo la tesi. \square

Grazie al Lemma 3.4.4, nel Corollario 3.4.5 specializzeremo il Teorema 3.4.3 al caso di somme di variabili aleatorie indipendenti.

Lemma 3.4.4. *Sia $X = (X_1, \dots, X_n, X_1^*, \dots, X_n^*)$ un vettore aleatorio di variabili indipendenti tale che $\forall i \in [n]$ X_i^* ha la distribuzione zero-bias associata a X_i . Supponiamo che le X_i siano centrate con varianza σ_i^2 finita. Supponiamo che $\sum_{i=1}^n \sigma_i^2 = 1$ e sia I una variabile aleatoria indipendente da X con*

$$\mathbb{P}(I = i) = \begin{cases} \sigma_i^2 & \text{se } i \in [n], \\ 0 & \text{altrimenti.} \end{cases}$$

Allora vale

$$S^* \sim S - X_I + X_I^*$$

dove $S = \sum_{i=1}^n X_i$.

Dimostrazione. Sia $f \in \text{AC}_{loc}(\mathbb{R})$ con $\mathbb{E}[|f'(S^*)|], \mathbb{E}[|Sf(S)|] < \infty$. $\forall i \in [n]$ poniamo $S^{(i)} := S - X_i$ allora

$$\begin{aligned} \mathbb{E}[Sf(S)] &= \mathbb{E} \left[\sum_{i=1}^n X_i f(S) \right] = \sum_{i=1}^n \mathbb{E} \left[X_i f(S^{(i)} + X_i) \right] \\ &= \sum_{i=1}^n \mathbb{E} \left[\mathbb{E} \left[X_i f(S^{(i)} + X_i) \mid S^{(i)} \right] \right] \\ &= \sum_{i=1}^n \sigma_i^2 \mathbb{E} \left[\mathbb{E} \left[f'(S^{(i)} + X_i^*) \mid S^{(i)} \right] \right] \\ &= \sum_{i=1}^n \sigma_i^2 \mathbb{E} \left[f'(S - X_i + X_i^*) \right] \end{aligned}$$

dove per la terza uguaglianza abbiamo utilizzato (B.3.1) e per la quarta la definizione di distribuzione zero-bias. Poiché $\mathbb{P}(I = i) = \sigma_i^2$ e $I \perp\!\!\!\perp X$ otteniamo

$$\begin{aligned} \sum_{i=1}^n \sigma_i^2 \mathbb{E} [f'(S - X_i + X_i^*)] &= \sum_{i=1}^n \mathbb{E} [\mathbb{1}_{\{i\}}(I) f'(S - X_i + X_i^*)] \\ &= \mathbb{E} [f'(S - X_I + X_I^*)]. \end{aligned}$$

Ma

$$\text{Var}(S) = \sum_{i=1}^n \text{Var}(X_i) = 1$$

e dunque abbiamo provato che $\mathbb{E}[Sf(S)] = \text{Var}(S)\mathbb{E}[f'(S - X_I + X_I^*)]$. La tesi segue dall'unicità della distribuzione zero-bias. \square

Corollario 3.4.5. *Siano X_1, \dots, X_n variabili aleatorie indipendenti con media nulla e varianza σ_i^2 finita. Se $\sum_{i=1}^n \sigma_i^2 = 1$ allora*

$$d_W(S, Z) \leq 2 \sum_{i=1}^n \sigma_i^2 \mathbb{E} [|X_i - X_i^*|]$$

dove $S = \sum X_i$, $Z \sim \mathcal{N}(0, 1)$ e per ogni $i \in [n]$, X_i^* è indipendente da X_i con la distribuzione zero-bias associata a X_i .

Dimostrazione. Sia I come nel Lemma 3.4.4 allora $S^* - S \sim X_I^* - X_I$. Applicando il Teorema 3.4.3 otteniamo

$$\begin{aligned} d_W(S, Z) &\leq 2\mathbb{E} [|S^* - S|] = 2\mathbb{E} [|X_I^* - X_I|] \\ &= 2 \sum_{i=1}^n \mathbb{E} [|X_i^* - X_i| \mid I = i] \mathbb{P}(I = i) \end{aligned}$$

dove abbiamo usato la legge dell'aspettazione totale (cfr. Teorema B.9). Ricordando che $\mathbb{P}(I = i) = \sigma_i^2$ otteniamo la tesi. \square

Nella Proposizione 3.4.7 mostreremo un'applicazione del corollario precedente alla somma di variabili aleatorie di Bernoulli i.i.d.. Prima di fare ciò, attraverso il seguente lemma, andremo ad esplicitare la distribuzione zero-bias associata ad una variabile di Bernoulli centrata e opportunamente normalizzata. Sia La Proposizione 3.4.7 che il Lemma 3.4.6 sono risultati sviluppati in maniera autonoma.

Lemma 3.4.6. Sia $p \in (0, 1)$ e $n \in \mathbb{N} \setminus \{0\}$ e $Y \sim \text{Bernoulli}(p)$. Se

$$X := \frac{Y - p}{\sqrt{np(1-p)}}$$

allora $X^* \sim \text{Uniform} \left(\left[-\sqrt{\frac{p}{n(1-p)}}, \sqrt{\frac{1-p}{np}} \right] \right)$.

In particolare, se X^* ha la zero-bias distribuzione associata ad X e $X^* \perp X$ otteniamo

$$\mathbb{E}[|X - X^*|] = \frac{1}{\sqrt{np(1-p)}}.$$

Dimostrazione. Essendo X centrata con $\text{Var}(X) = \frac{p(1-p)}{np(1-p)} = \frac{1}{n}$ applicando il Teorema 3.4.2 otteniamo che X^* è una variabile aleatoria continua con densità data dalla formula

$$p^*(x) = n\mathbb{E}[X \mathbb{1}_{(x, \infty)}(X)].$$

Poiché X assume il valore $x_1 = -\sqrt{\frac{p}{n(1-p)}}$ con probabilità $1-p$ e $x_2 = \sqrt{\frac{1-p}{np}}$ con probabilità p otteniamo

$$p^*(x) = n \begin{cases} \mathbb{E}[X] & \text{se } x < x_1, \\ p\sqrt{\frac{1-p}{np}} & \text{se } x_1 \leq x \leq x_2, \\ 0 & \text{se } x > x_2 \end{cases} = \begin{cases} 0 & \text{se } x < x_1, \\ \sqrt{np(1-p)} & \text{se } x_1 \leq x \leq x_2, \\ 0 & \text{se } x > x_2 \end{cases}$$

ciò prova che $X^* \sim \text{Uniform} \left(\left[-\sqrt{\frac{p}{n(1-p)}}, \sqrt{\frac{1-p}{np}} \right] \right)$.

Passiamo al calcolo di $\mathbb{E}[|X - X^*|]$. Dalla proprietà (B.3.1) otteniamo

$$\begin{aligned} \mathbb{E}[|X - X^*|] &= \mathbb{E}[|X - X^*| | X] = (1-p)\mathbb{E}[X^* - x_1] + p\mathbb{E}[x_2 - X^*] \\ &= (1-p)(\mathbb{E}[X^*] - x_1) + p(x_2 - \mathbb{E}[X^*]) \\ &= (1-p)\left(\frac{x_1 + x_2}{2} - x_1\right) + p\left(x_2 - \frac{x_1 + x_2}{2}\right) \\ &= \frac{x_2 - x_1}{2} = \frac{1}{2\sqrt{n}} \left(\sqrt{\frac{1-p}{p}} + \sqrt{\frac{p}{1-p}} \right) = \frac{1}{2\sqrt{np(1-p)}} \end{aligned}$$

dove la seconda uguaglianza segue dal fatto che $X \perp X^*$ e da $x_1 \leq X^* \leq x_2$. \square

Proposizione 3.4.7. Siano Y_1, \dots, Y_n i.i.d. con distribuzione di Bernoulli di parametro p . Se

$$X_i := \frac{Y_i - p}{\sqrt{np(1-p)}}$$

allora

$$d_W(S, Z) \leq \frac{1}{\sqrt{np(1-p)}}$$

dove $S = \sum_{i=1}^n X_i$ e $Z \sim \mathcal{N}(0, 1)$.

Dimostrazione. Notiamo che

$$\text{Var}(S) = \sum_{i=1}^n \text{Var}(X_i) = n \frac{p(1-p)}{np(1-p)} = 1.$$

Dunque, utilizzando il Corollario 3.4.5 otteniamo

$$d_W(S, Z) \leq 2 \sum_i^n \text{Var}(X_i) \mathbb{E}[|X_i - X_i^*|].$$

Ma $\text{Var}(X_1) = \frac{p(1-p)}{np(1-p)} = \frac{1}{n}$. Per il lemma precedente si ha

$$d_W(S, Z) \leq 2n \frac{1}{n} \frac{1}{2\sqrt{np(1-p)}} = \frac{1}{\sqrt{np(1-p)}}.$$

□

3.5 | Size-bias

Un altro metodo che si può utilizzare per limitare $\mathbb{E}[Sf(S) - f'(S)]$ è quello della size-bias. Questo metodo appare per la prima volta in [26].

Definizione 3.4. Sia S una variabile aleatoria non-negativa con media $\mu > 0$. Diremo che la variabile aleatoria S^s ha la distribuzione **size-bias** associata ad S se per ogni funzione f limitata e continua vale

$$\mathbb{E}[Sf(S)] = \mu \mathbb{E}[f(S^s)]. \quad (3.5.1)$$

Il seguente lemma ci fornisce una caratterizzazione per la distribuzione size-bias associata a variabili aleatorie discrete.

Lemma 3.5.1. Sia X una variabile aleatoria discreta non negativa con media $\mu > 0$ e che assume valori in $S \subseteq \mathbb{R}$. Se p_X è la densità di probabilità di X , una variabile aleatoria discreta con densità

$$p_{X^s}(x) = \frac{x}{\mu} p_X(x). \quad (3.5.2)$$

ha la distribuzione size-bias associata ad X . In particolare, X^s assume valori in S .

Dimostrazione. Per prima cosa notiamo che (3.5.2) è una distribuzione di probabilità su S . Poiché X assume valori non negativi $p_{X^s}(x) = 0$ per ogni $x < 0$ ed inoltre $p_{X^s}(x) \geq 0$ per ogni $x \geq 0$. Essendo $\mu = \mathbb{E}[X]$ vale

$$\mu = \sum_{x \in S} xp_X(x)$$

e dunque

$$\sum_{x \in S} \frac{x}{\mu} p_X(x) = 1.$$

Sia Y una variabile aleatoria discreta con densità data da (3.5.2). Fissata una funzione f continua e limitata si ha

$$\begin{aligned} \mathbb{E}[Xf(X)] &= \sum_{x \in S} xf(x)p_X(x) = \mu \sum_{x \in S} \frac{x}{\mu} f(x)p_X(x) \\ &= \mu \sum_{x \in S} xp_Y(x) = \mu \mathbb{E}[f(Y)]. \end{aligned}$$

Abbiamo dunque provato che per ogni funzione f continua e limitata vale $\mathbb{E}[Xf(X)] = \mu \mathbb{E}[f(Y)]$. Dalla definizione di size-bias, Y ha una distribuzione size-bias associata a X . □

Osservazione 4 (Come si arriva a (3.5.2)?). Supponiamo che X^s abbia la distribuzione size-bias associata a X . Fissato $x \in \mathbb{R}$ otteniamo

$$p_{X^s}(x) := \mathbb{P}(X^s = x) = \mathbb{E}[\mathbb{1}_{\{x\}}(X)] = \frac{1}{\mu} \mathbb{E}[X \mathbb{1}_{\{x\}}(X)] = \frac{x}{\mu} p_X(x)$$

dove la penultima disuguaglianza deriva da (3.5.1) dove abbiamo posto $f = \mathbb{1}_{\{x\}}$.

Il seguente teorema mostra come utilizzare la distribuzione size-bias per limitare il membro destro di (3.1.11).

Teorema 3.5.2. Sia S una variabile aleatoria non-negativa con media $\mu > 0$ e varianza σ^2 . Sia S^s una variabile, definita sullo stesso spazio di probabilità di S , con la distribuzione size-bias associata ad S . Posta

$$\bar{S} = \frac{S - \mu}{\sigma}$$

la standardizzazione di S vale

$$d_W(\bar{S}, Z) \leq \sqrt{\frac{2}{\pi}} \frac{1}{\sigma^2} \mathbb{E}[|\sigma^2 - \mu(S^s - S)|] + \frac{\mu}{\sigma^3} \mathbb{E}[|S^s - S|^2].$$

Dimostrazione. Sia \mathcal{F}_W la classe definita in (3.1.12) e fissiamo $f \in \mathcal{F}_W$. Chiamata

$$\tilde{f}(x) = f\left(\frac{x - \mu}{\sigma}\right)$$

si ha

$$\begin{aligned} \mathbb{E}[\bar{S}f(\bar{S})] &= \mathbb{E}\left[\frac{S - \mu}{\sigma} f\left(\frac{S - \mu}{\sigma}\right)\right] \\ &= \frac{1}{\sigma} \mathbb{E}\left[S\tilde{f}(S)\right] - \frac{\mu}{\sigma} \mathbb{E}\left[f\left(\frac{S - \mu}{\sigma}\right)\right] \\ &= \frac{\mu}{\sigma} \mathbb{E}\left[\tilde{f}(S^s)\right] - \frac{\mu}{\sigma} \mathbb{E}\left[f\left(\frac{S - \mu}{\sigma}\right)\right] \\ &= \frac{\mu}{\sigma} \mathbb{E}\left[f\left(\frac{S^s - \mu}{\sigma}\right) - f\left(\frac{S - \mu}{\sigma}\right)\right] \end{aligned} \tag{3.5.3}$$

dove la penultima uguaglianza segue da (3.5.1). Usando l'espansione di Taylor otteniamo

$$\mathbb{E}[\bar{S}f(\bar{S})] = \frac{\mu}{\sigma} \mathbb{E}\left[\frac{S^s - S}{\sigma} f'\left(\frac{S - \mu}{\sigma}\right) + \frac{(S^s - S)^2}{2\sigma^2} f''(\xi)\right] \tag{3.5.4}$$

dove ξ è una variabile aleatoria che assume valori compresi fra $\frac{S - \mu}{\sigma}$ e $\frac{S^s - \mu}{\sigma}$.

Sottraendo $\mathbb{E}[f'(\bar{S})]$ da (3.5.4) si ha

$$\mathbb{E}[\bar{S}f(\bar{S}) - f'(\bar{S})] = \mathbb{E}\left[\left(\frac{\mu(S^s - S)}{\sigma^2} - 1\right) f'(\bar{S}) + \frac{\mu(S^s - S)^2}{2\sigma^3} f''(\xi)\right].$$

Dunque

$$\begin{aligned} |\mathbb{E}[\bar{S}f(\bar{S}) - f'(\bar{S})]| &\leq \mathbb{E}\left[\left|\frac{\mu(S^s - S)}{\sigma^2} - 1\right| |f'(\bar{S})| + \left|\frac{\mu(S^s - S)^2}{2\sigma^3}\right| |f''(\xi)|\right] \\ &\leq \|f'\|_\infty \mathbb{E}\left[\left|\frac{\mu(S^s - S)}{\sigma^2} - 1\right|\right] + \|f''\|_\infty \mathbb{E}\left[\left|\frac{\mu(S^s - S)^2}{2\sigma^3}\right|\right] \\ &\leq \sqrt{\frac{2}{\pi}} \mathbb{E}\left[\left|\frac{\mu(S^s - S)}{\sigma^2} - 1\right|\right] + 2\mathbb{E}\left[\left|\frac{\mu(S^s - S)^2}{2\sigma^3}\right|\right] \\ &= \sqrt{\frac{2}{\pi}} \frac{1}{\sigma^2} \mathbb{E}[|\mu(S^s - S) - \sigma^2|] + \frac{\mu}{\sigma^3} \mathbb{E}[|S^s - S|^2] \end{aligned}$$

dove l'ultima disuguaglianza segue dalla proprietà della classe \mathcal{F}_W .

Attraverso il seguente lemma andremo a specializzare il Teorema 3.5.2 al caso di somme di variabili aleatorie indipendenti ottenendo il Corollario 3.5.4.

Lemma 3.5.3. Sia $X = (X_1, \dots, X_n, X_1^s, \dots, X_n^s)$ un vettore aleatorio di variabili indipendenti tale che $\forall i \in [n]$ X_i è non negativa con media $\mu_i > 0$ e X_i^s ha la distribuzione size-bias associata a X_i . Se I è una variabile aleatoria indipendente da X tale con

$$\mathbb{P}(I = i) = \begin{cases} \frac{\mu_i}{\mu} & \text{se } i \in [n], \\ 0 & \text{altrimenti,} \end{cases}$$

dove $\mu = \sum_{i=1}^n \mu_i$, allora

$$S^s \sim S - X_I + X_I^s$$

dove $S = \sum_{i=1}^n X_i$.

Dimostrazione. Sia f continua e limitata, $\forall i \in [n]$ posta $S^{(i)} := S - X_i$ otteniamo

$$\begin{aligned} \mathbb{E}[Sf(S)] &= \mathbb{E}\left[\sum_{i=1}^n X_i f(S)\right] = \sum_{i=1}^n \mathbb{E}[X_i f(S - X_i + X_i)] \\ &= \sum_{i=1}^n \mathbb{E}\left[\mathbb{E}\left[X_i f(S^{(i)} + X_i) \mid S^{(i)}\right]\right] \\ &= \sum_{i=1}^n \mu_i \mathbb{E}\left[\mathbb{E}\left[f(S^{(i)} + X_i^s) \mid S^{(i)}\right]\right] = \sum_{i=1}^n \mu_i \mathbb{E}\left[f(S - X_i + X_i^s)\right] \end{aligned}$$

dove abbiamo utilizzato (B.3.1). Poiché $\mu \mathbb{P}(I = i) = \mu_i$ e $I \perp X$ otteniamo

$$\begin{aligned} \sum_{i=1}^n \mu_i \mathbb{E}\left[f(S - X_i + X_i^s)\right] &= \mu \sum_{i=1}^n \mathbb{E}\left[\mathbb{1}_{\{i\}}(I) f(S - X_i + X_i^s)\right] \\ &= \mu \mathbb{E}\left[f(S - X_I + X_I^s)\right]. \end{aligned}$$

Abbiamo provato che $\mathbb{E}[Sf(S)] = \mu \mathbb{E}[f(S - X_I + X_I^s)]$. La tesi segue dalla caratterizzazione della distribuzione size-bias. □

Corollario 3.5.4. Siano X_1, \dots, X_n variabili aleatorie non-negative e indipendenti con $\mathbb{E}[X_i] = \mu_i > 0$ e $\sigma_i^2 = \text{Var}(X_i) < \infty$ per ogni $i \in [n]$. Allora

$$\begin{aligned} d_W\left(\frac{S - \mu}{\sigma}, Z\right) &\leq \sqrt{\frac{2}{\pi}} \frac{1}{\sigma^2} \sum_{i=1}^n \mu_i \mathbb{E}\left[\left|\frac{\sigma^2}{\mu} - (X_i^s - X_i)\right|\right] \\ &\quad + \frac{1}{\sigma^3} \sum_{i=1}^n \mu_i \mathbb{E}\left[|X_i^s - X_i|^2\right] \end{aligned}$$

dove $S = \sum X_i$, $\sigma^2 = \text{Var}(S)$, $\mu = \mathbb{E}[S]$, $Z \sim \mathcal{N}(0, 1)$ e per ogni $i \in [n]$ X_i^s ha la distribuzione size-bias associata a X_i con $X_i^s \perp X_i$.

Dimostrazione. Sia l come nel Lemma 3.5.3 allora $S^s - S \sim X_I^s - X_I$. Applicando il Teorema 3.5.2 otteniamo

$$\begin{aligned} d_W(\bar{S}, Z) &\leq \sqrt{\frac{2}{\pi}} \frac{1}{\sigma^2} \mathbb{E} [|\sigma^2 - \mu(S^s - S)|] + \frac{\mu}{\sigma^3} \mathbb{E} [|S^s - S|^2] \\ &\leq \sqrt{\frac{2}{\pi}} \frac{1}{\sigma^2} \mathbb{E} [|\sigma^2 - \mu(X_I^s - X_I)|] + \frac{\mu}{\sigma^3} \mathbb{E} [|X_I^s - X_I|^2]. \end{aligned} \quad (3.5.5)$$

Utilizzando la legge dell'aspettazione totale e poiché $I \perp (X_j, X_j^s)$ per ogni $j \in [n]$ otteniamo

$$\begin{aligned} \mathbb{E} [|\sigma^2 - \mu(X_I^s - X_I)|] &= \sum_{i=1}^n \mathbb{E} [|\sigma^2 - \mu(X_i^s - X_i)| \mid I = i] \mathbb{P}(I = i) \\ &= \sum_{i=1}^n \frac{\mu_i}{\mu} \mathbb{E} [|\sigma^2 - \mu(X_i^s - X_i)|] \\ &= \sum_{i=1}^n \mu_i \mathbb{E} \left[\left| \frac{\sigma^2}{\mu} - (X_i^s - X_i) \right| \right] \end{aligned} \quad (3.5.6)$$

dove per l'ultima disuguaglianza abbiamo usato che $\mu > 0$.

Similmente

$$\begin{aligned} \mu \mathbb{E} [|X_I^s - X_I|^2] &= \mu \sum_{i=1}^n \mathbb{E} [|X_i^s - X_i|^2 \mid I = i] \mathbb{P}(I = i) = \\ &= \mu \sum_{i=1}^n \frac{\mu_i}{\mu} \mathbb{E} [|X_i^s - X_i|^2] = \sum_{i=1}^n \mu_i \mathbb{E} [|X_i^s - X_i|^2]. \end{aligned} \quad (3.5.7)$$

Combinando (3.5.5), (3.5.6) e (3.5.7) otteniamo la tesi. □

Nella Proposizione 3.5.6 mostreremo un'applicazione del corollario precedente alla somma di variabili aleatorie di Poisson i.i.d.. Prima di fare ciò occorre, attraverso il Lemma 3.5.5, esplicitare la distribuzione size-bias associata ad una variabile di Poisson.

Lemma 3.5.5. *Sia X una distribuzione di Poisson allora vale*

$$X^s \sim X + 1.$$

Dimostrazione. Ricordiamo che una variabile aleatoria di Poisson di parametro λ ha media e varianza paria a λ e densità discreta data dalla formula

$$p_X(n) = e^{-\lambda} \frac{\lambda^n}{n!} \mathbb{1}_{\mathbb{N}}(n)$$

Dunque, utilizzando (3.5.2), otteniamo

$$p_{X^s}(n) = \frac{n}{\lambda} e^{-\lambda} \frac{\lambda^n}{n!} \mathbb{1}_{\mathbb{N}}(n) = e^{-\lambda} \frac{\lambda^{n-1}}{(n-1)!} \mathbb{1}_{\mathbb{N}}(n) = p_X(n-1).$$

Notando che $p_X(n-1)$ è la densità di $X+1$ si ha la tesi. □

Proposizione 3.5.6. *Siano X_1, \dots, X_n i.i.d. con distribuzione di Poisson di parametro λ allora*

$$d_W(\bar{S}, Z) \leq \frac{1}{\sqrt{\lambda n}}$$

dove

$$\bar{S} = \frac{\sum X_i - \lambda n}{\sqrt{n\lambda}}$$

e $Z \sim \mathcal{N}(0, 1)$.

Dimostrazione. Poiché $\mathbb{E}[X_i] = \text{Var}(X_i) = \lambda$. Applicando il Corollario 3.5.4 otteniamo

$$d_W(\bar{S}, Z) \leq \sqrt{\frac{2}{\pi}} \frac{1}{\lambda n} \sum_{i=1}^n \lambda \mathbb{E}[|1 - (X_i^s - X_i)|] + \frac{1}{\lambda n \sqrt{\lambda n}} \sum_{i=1}^n \lambda \mathbb{E}[|X_i^s - X_i|^2].$$

Dal Lemma 3.5.5 sappiamo che $X_i^s - X_i \sim 1$, dunque l'espressione di sopra diventa

$$d_W(\bar{S}, Z) \leq \frac{1}{\lambda n \sqrt{\lambda n}} \sum_{i=1}^n \lambda = \frac{1}{\sqrt{\lambda n}}.$$

□

4 | Change-point detection con metodi grafici

4.1 | Il change-point detection

L'analisi delle serie temporali è diventata sempre più importante in svariati campi tra cui la medicina, l'astronomia, la finanza, la meteorologia, l'intrattenimento Possiamo pensare i dati delle serie temporali come sequenze di misurazioni nel tempo che descrivono il comportamento del sistema in esame. Questi comportamenti possono, a causa di eventi esterni e/o cambiamenti sistematici interni, cambiare nel tempo. In letteratura va sotto il nome di change-point detection il problema della determinazione dell'effettivo cambiamento di comportamento del sistema ed eventualmente della sua localizzazione temporale. L'applicazione del change-point è molto ampia e copre problemi di natura molto differenti. Andiamo a presentare alcune applicazioni motivanti con la relativa letteratura.

- **Monitoraggio delle condizioni mediche.** Il monitoraggio continuo della salute del paziente consiste nella misurazione delle variabili fisiologiche (e.g. frequenza cardiaca, elettroencefalogramma, elettrocardiogramma ...). In questo ambito il change-point detection viene utilizzato per eseguire un monitoraggio automatizzato e in tempo reale del paziente al fine di rendere le cure più efficaci [5, 31, 38, 44].
- **Rilevamento del cambiamento climatico.** A causa del possibile verificarsi di cambiamenti climatici, dell'aumento dei gas serra nell'atmosfera e della siccità, negli ultimi decenni il problema del change-point detection viene applicato anche su dati di natura meteorologica [11, 15, 32].
- **Riconoscimento vocale.** Il riconoscimento vocale è il processo di conversione delle espressioni vocali pronunciate dall'uomo in testo. I metodi di change-point detection in questo ambito vengono applicati per la segmentazione audio cioè per separare la traccia del silenzio, da quella del rumore ambientale per estrarre solamente le parole [10, 36].
- **Analisi delle immagini.** Il change-point detection può essere utile anche per l'analisi dei dati video; in questo ambito il problema è data una serie di immagini della stessa scena scattate in momenti diversi identificare gli insiemi di pixel significativamente diversi. Tale tecniche possono essere

utilizzate in diversi ambiti tra cui, ad esempio la videosorveglianza [12], il telerilevamento [6], il rilevamento subacqueo [18] e i sistemi a guida autonoma [21].

- **Analisi dell'attività umana.** Il rilevamento dei cambiamenti delle attività rilevate attraverso i sensori presenti nelle “smart home” o tramite i dispositivi mobili può essere formulato come problema di change-point. Questi punti di cambiamento sono utili per segmentare le attività umane con il fine di fornire servizi sempre più personalizzati ma anche per rilevare cambiamenti nel comportamento che diano informazioni sullo stato di salute [29, 41].

Tutti questi problemi possono essere descritti mediante un approccio statistico. Data una sequenza di osservazioni $(y_i)_{i=1}^n$ indicizzate in un qualche modo (spaziale e/o temporale) e dati $1 \leq n_0 \leq n_1 \leq n$ siamo interessati a testare l'ipotesi nulla:

$$\mathcal{H}_0 : y_i \sim F_0 \quad \text{per } i = 1, \dots, n \quad (4.1.1)$$

contro l'ipotesi alternativa:

$$\mathcal{H}_1 : \exists \tau : \quad n_0 \leq \tau \leq n_1 \leq n \text{ per cui } y_i \sim \begin{cases} F_1 & \text{se } i > \tau \\ F_0 & \text{altrimenti} \end{cases} \quad (4.1.2)$$

per alcune distribuzioni di probabilità F_0 e F_1 distinte (non fissate a priori). In altre parole, ci dobbiamo chiedere se le osservazioni seguono tutte una comune distribuzione (ipotesi nulla) oppure se esista un istante temporale per cui le osservazioni seguono una certa distribuzione prima di tale e una differente dopo questo tempo (ipotesi alternativa).

Nella sezione 4.6 verrà presentato, sotto opportune ipotesi sui dati, un test statistico per le ipotesi (4.1.1) e (4.1.2)

4.2 | Alcune definizioni sui grafi

Nelle sezioni successive utilizzeremo i grafi. Fissiamo, in questa sezione, alcune notazioni e definizioni utili nel seguito. Le proprietà sui grafi che utilizzeremo saranno provate nell'Appendice C.

Se X è un insieme e $n \in \mathbb{N}$ poniamo

$$\binom{X}{n} := \{A \subseteq X \mid |A| = n\}$$

ovvero $\binom{X}{n}$ è l'insieme dei sottoinsiemi di X aventi cardinalità n .

Definizione 4.1. Un **grafo semplice** è una coppia $G = (V, E)$ dove

- V è un insieme finito e
- $E \subseteq \binom{V}{2}$.

Gli elementi di V si chiamano **vertici** mentre quelli di E si chiamano **archi**.

Per non appesantire ulteriormente la trattazione ometteremo il termine semplice e parleremo di grafi.

Definizione 4.2. Diremo che $G = (V, E)$ è **connesso** se per ogni coppia $x \neq y \in V$ esistono dei vertici $x = x_1, x_2, \dots, x_{l+1} = y$ tali che $\forall i \in [l]$ vale $\{x_i, x_{i+1}\} \in E$.

Definizione 4.3. Diremo che $G = (V, E)$ ammette un **ciclo** se esiste una successione di vertici $x = x_1, \dots, x_{l+1} = x$ tali che per ogni $i \in [l]$ vale $\{x_i, x_{i+1}\} \in E$ e se $i \neq j$ allora $\{x_i, x_{i+1}\} \neq \{x_j, x_{j+1}\}$.

Definizione 4.4. Diremo che $G = (V, E)$ è un **albero** se è connesso e senza cicli. Equivalentemente, diremo che un grafo è un albero se è connesso e soddisfa la relazione $|E| = |V| - 1$.

L'equivalenza tra le due definizioni di albero viene provata in [27, Teorema 4.1, p.32].

Definizione 4.5. Dato $G = (V, E)$ diremo che due archi distinti a e b sono **adiacenti** se $a \cap b \neq \emptyset$.

Definizione 4.6. Dato un grafo $G = (V, E)$ e un suo vertice x , definiamo l'**intorno** di x in G come l'insieme

$$\Gamma_G(x) := \{y \in V \mid \{x, y\} \in E\}$$

e il **grado** in G del vertice x come la quantità

$$d_G(x) := |\Gamma_G(x)|.$$

Diremo che un vertice è **isolato** se $d_G(i) = 0$.

Definizione 4.7. Per ogni arco e del grafo sia

$$A_e := \{l \in E \mid l \cap e \neq \emptyset\}, \quad (4.2.1)$$

$$B_e := \bigcup_{e' \in A_e} A_{e'}, \quad (4.2.2)$$

$$C_e := \{j \in [n] \mid \exists l \in A_e, j \in l\}. \quad (4.2.3)$$

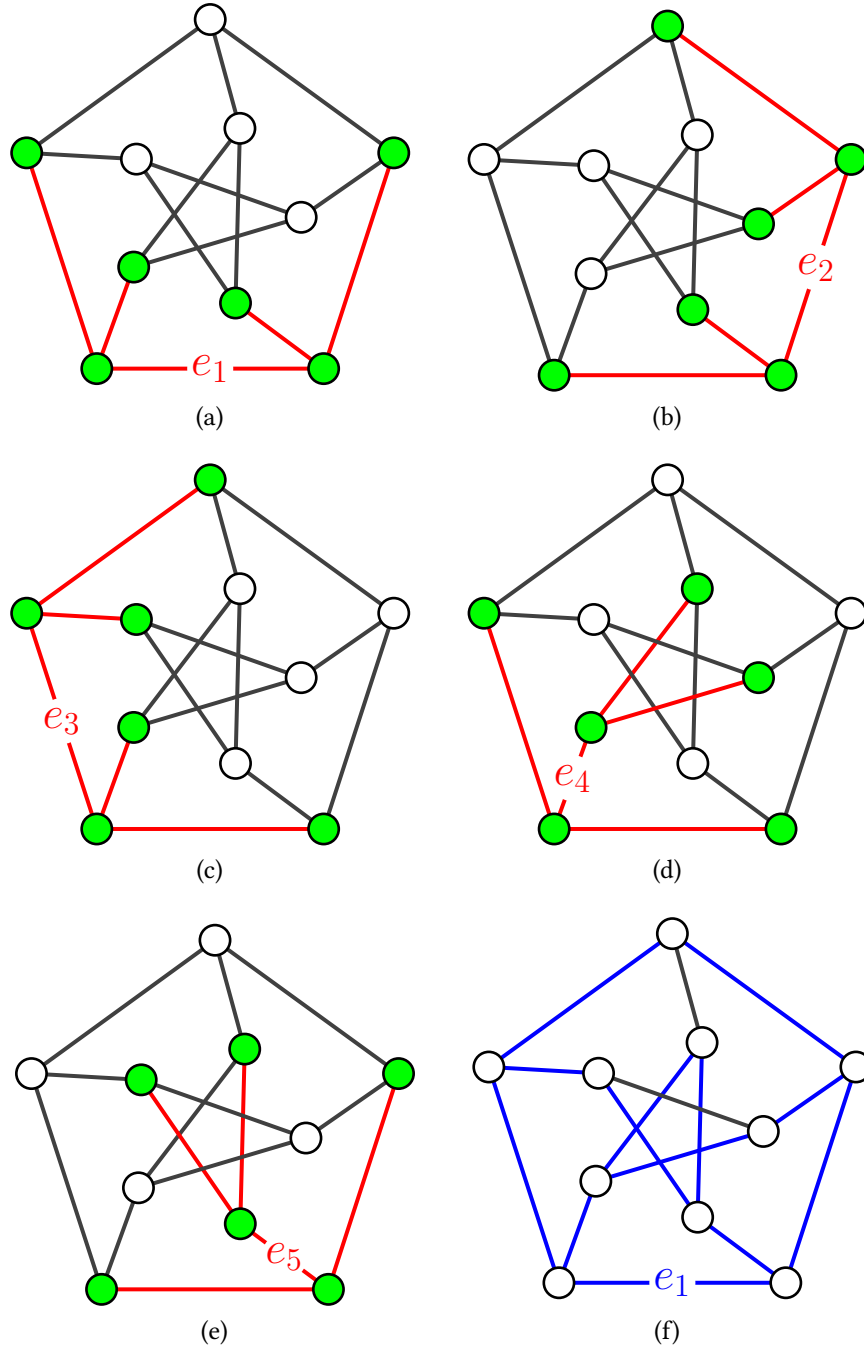


Figura 4.1: Il grafo di Petersen in cui sono evidenziati in rosso gli archi di A_e e in verde i vertici di C_e ordinatamente per $e = e_1, \dots, e_5$. In (f) sono evidenziati in blu gli archi di B_{e_1} .

Si veda la Figura 4.1 per un esempio di tali insiemi su un particolare grafo.

Definizione 4.8. Per ogni vertice i del grafo poniamo

$$\Gamma_G^2(i) := \{e \in E \mid \exists j \in \Gamma_G(i), j \in e\}. \quad (4.2.4)$$

Definizione 4.9. Diremo che due grafi $G_1 = (V_1, E_1)$ e $G_2 = (V_2, E_2)$ sono

isomorfi se esiste una mappa invertibile $f : V_1 \rightarrow V_2$ tale che

$$\{x, y\} \in E_1 \quad \Leftrightarrow \quad \{f(x), f(y)\} \in E_2.$$

accoppiamento

4.3 | Grafi associati a quantità osservate

Sia \mathcal{X} un insieme e sia $d : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$ una funzione assegnata pensata come funzione di dissimilarità. Dato $V \subseteq \mathcal{X}$ finito possiamo introdurre i seguenti grafi.

- Se \mathcal{T} è l'insieme degli alberi su V definiamo l'**albero di copertura di dissimilarità minima** (*minimum spanning tree*) come \bar{T} dove

$$\bar{T} = \operatorname{argmin}_{T=(V,E) \in \mathcal{T}} \left\{ \sum_{\{x,y\} \in E} d(x,y) \right\}.$$

- Se \mathfrak{P} l'insieme delle partizioni di V formate da coppie, cioè

$$\mathfrak{P} = \left\{ P \subseteq \binom{V}{2} \mid P \text{ partizione di } V \right\},$$

definiamo la partizione di costo minimo \bar{P} come

$$\bar{P} := \operatorname{argmin}_{P \in \mathfrak{P}} \left\{ \sum_{\{x,y\} \in P} d(x,y) \right\}.$$

Il **grafo di minimo accoppiamento** (*minimum distance pairing*) è (V, \bar{P}) .

- Sia $n : V \rightarrow V$ la mappa definita da

$$n(x) := \operatorname{argmin}_{\substack{y \in V \\ y \neq x}} d(x, y).$$

Il grafo del **vicino più vicino** (*nearest neighbor graph*) è definito come (V, E) dove $E := \{\{x, n(x)\} \mid x \in V\}$.

In Figura 4.2 si possono notare le rappresentazioni dei tre grafi presentati. Per realizzarli abbiamo preso $V \subseteq \mathbb{R}^2$ formato da 30 punti randomici: 15 campionati dalla distribuzione $\mathcal{N}((0, 0), I_2)$ e i rimanenti dalla distribuzione $\mathcal{N}((2, 2), I_2)$. Abbiamo, inoltre, considerato la distanza euclidea di \mathbb{R}^2 come funzione di dissimilarità. Per l'implementazione si rimanda il lettore all'Appendice C.1.1.

Nelle applicazioni supporremo di avere una sequenza di osservazioni (y_1, \dots, y_n) e posto $V = \{y_1, \dots, y_n\}$ costruiremo un grafo $G = (V, E)$ (es. uno tra quelli definiti sopra). Possiamo associare a $G = (V, E)$ un grafo $\bar{G} = (\bar{V}, \bar{E})$ su $\{1, \dots, n\} := [n]$ dove $\bar{E} = \{\{i, j\} \mid \{y_i, y_j\} \in E\}$.

Notiamo che se y_1, \dots, y_n sono distinti c'è un naturale isomorfismo tra i grafi G e \bar{G} . Di seguito supponemmo, senza ulteriore menzione, che i valori osservati corrispondano a variabili aleatorie che assumono quasi certamente valori distinti. Questa assunzione porta ad utilizzare il simbolo G anche per il grafo \bar{G} .

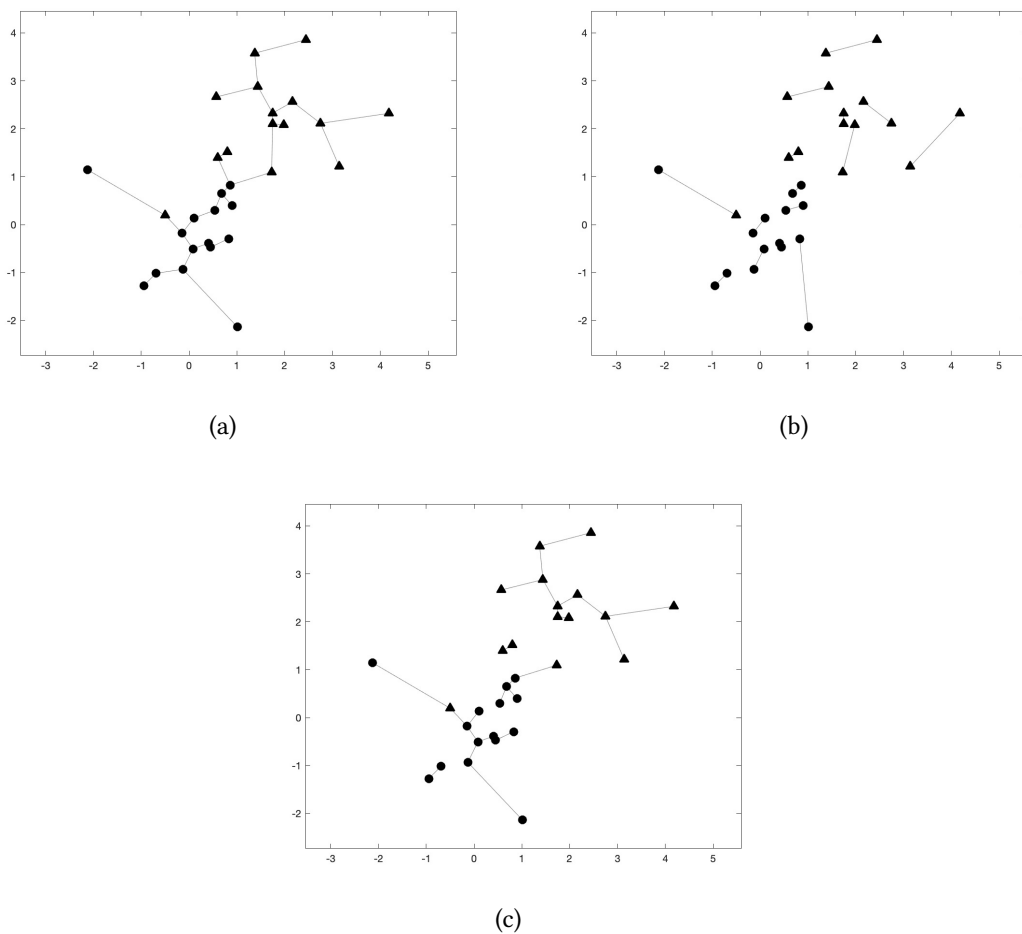


Figura 4.2: (a) L'albero di copertura di dissimilarità minima; (b) il grafo di minimo accoppiamento; (c) il grafo del vicino più vicino. I dati sono punti in \mathbb{R}^2 : 15 punti sono estratti da $\mathcal{N}((0, 0), I_2)$ (disegnati con un cerchio) e 15 punti sono estratti da $\mathcal{N}((2, 2), I_2)$ (disegnati con un triangolo).

4.4 | Le variabili aleatorie $R_G(t)$ e $Z_G(t)$

Definizione 4.10. Denotiamo con \mathcal{S}_n il gruppo delle permutazioni su $[n]$ e definiamo la distribuzione **permutation null** come la distribuzione uniforme su \mathcal{S}_n . Nel

seguito indicheremo con \mathbb{P} tale distribuzione e con \mathbb{E} , Var rispettivamente il valore atteso e la varianza sotto questa distribuzione.

Fissato un grafo $G = G(n) = ([n], E)$ dove $E = E(n)$ e fissato $t \in \mathbb{N}$ definiamo il contatore $R_G(t)$ come la variabile aleatoria

$$R_G(t) : \mathcal{S}_n \rightarrow \mathbb{N} \quad R_G(t)(\pi) = \sum_{\{i,j\} \in E} \mathbb{1}_{\{\sigma \in \mathcal{S}_n \mid g_i(t)(\sigma) \neq g_j(t)(\sigma)\}}(\pi) \quad (4.4.1)$$

dove

$$g_i(t)(\sigma) = \begin{cases} 1 & \text{se } \sigma(i) > t \\ 0 & \text{altrimenti} \end{cases}.$$

Definiamo, inoltre, la variabile aleatoria

$$Z_G(t) := -\frac{R_G(t) - \mathbb{E}[R_G(t)]}{\sqrt{\text{Var}(R_G(t))}}.$$

Per la definizione di R_G e Z_G abbiamo ripreso la notazione presente in [8]. Notiamo che $R_G(t)(\pi)$ conta il numero di lati $\{i, j\} \in E$ per cui $\pi(i) \leq t < \pi(j)$ o $\pi(j) \leq t < \pi(i)$.

Il prossimo lemma ci permette di calcolare esplicitamente media e varianza di $R_G(t)$ e dunque di ottenere un'espressione semplificata per $Z_G(t)$.

Lemma 4.4.1. *Sia $G = ([n], E)$ un grafo parametrizzato da $n \in \mathbb{N}$ (quindi $G = G(n)$ e $E = E(n)$). Fissato $t \in [n]$ vale*

$$\mathbb{E}[R_G(t)] = p_1(t) |E| \quad (4.4.2)$$

e

$$\begin{aligned} \text{Var}(R_G(t)) = p_2(t) |E| + \sum_{i=1}^n d_G(i)^2 \left(\frac{1}{2} p_1(t) - p_2(t) \right) \\ + |E|^2 (p_2(t) - p_1(t)^2) \end{aligned} \quad (4.4.3)$$

dove

$$p_1(t) := 2 \frac{t(n-t)}{n(n-1)},$$

$$p_2(t) := 4 \frac{t(t-1)(n-t)(n-t-1)}{n(n-1)(n-2)(n-3)}.$$

Tale lemma appare enunciato come Lemma 2.1 in [8] e un'idea della dimostrazione appare nel supplemento a tale articolo. La dimostrazione presentata nella sezione finale di questo capitolo chiarisce alcuni dettagli lasciati impliciti in [8].

4.5 | Teorema di approssimazione del processo $Z_G(\cdot)$

Ricordiamo due importanti definizioni per il comportamento di una successione all'infinito.

Definizione 4.11. Diremo $f(x) \asymp g(x)$ per $x \rightarrow +\infty$ se esistono due costanti positive C_1, C_2 indipendenti da x tali che, definitivamente in x , valga

$$C_1 g(x) \leq f(x) \leq C_2 g(x).$$

Definizione 4.12. Diremo che $f(x) = o(g(x))$ per $x \rightarrow +\infty$ se

$$\lim_{x \rightarrow +\infty} \frac{f(x)}{g(x)} = 0$$

con la convenzione $\frac{f(x)}{g(x)} = 0$ quando $f(x) = g(x) = 0$.

Fissato un numero reale α con $0 < \alpha < 1.125$ e fissata una successione di grafi $(G_n)_{n \in \mathbb{N}}$ dove $G_n = ([n], E_n)$ consideriamo le seguenti condizioni

Condizione 1. $|E_n| \asymp n^\alpha$.

Condizione 2. $\sum_{e \in E_n} |A_e| |B_e| = o\left(n^{1.5(\alpha \wedge 1)}\right)$.

Presentiamo ora un teorema di approssimazione che ci permetterà di presentare un test statistico per il problema del change-point detection.

Teorema 4.5.1. [8, Teorema 3.1] Sia $G = ([n], E)$ un grafo parametrizzato da $n \in \mathbb{N}$ (quindi $G = G(n)$ e $E = E(n)$). Se la Condizione 1 e la Condizione 2 sono verificate allora le distribuzioni delle marginali finite del processo stocastico $(Z_G(\lfloor nu \rfloor))_{u \in (0,1)}$ sono ben approssimate, per n grande, da quelle del processo gaussiano centrato $(Z_G^*(u))_{u \in (0,1)}$ con funzione di covarianza

$$\rho_G^*(u, v) = \frac{2(u \wedge v)^2(1 - (u \vee v))^2 |E|}{\sigma_G^*(u) \sigma_G^*(v)} \cdot \left((1 - (u \vee v))^2 |E| + (1 - 2u)(1 - 2v) \sum_{i=1}^n d_G(i)^2 \right)$$

dove

$$\sigma_G^*(u) = \sqrt{2u^2(1 - u)^2 |E| + u(1 - u)(1 - 2u)^2 \sum_{i=1}^n d_G(i)^2}$$

Per la dimostrazione di questo teorema necessitiamo di alcuni strumenti che verranno presentati nella prossima sezione. La formulazione dell'approssimazione nel Teorema 4.5.1 verrà presentata nella Sezione 5.5 dove verrà data una dimostrazione (in parte euristica ma supportata da un'argomentazione matematica).

4.6 | Il test statistico

Dalla Figura 4.3 e dalla Figura 4.4 notiamo che in assenza di change-point $Z_G(t)$ raggiunge un valore massimo molto inferiore a quello raggiunto nei dati con change-point.

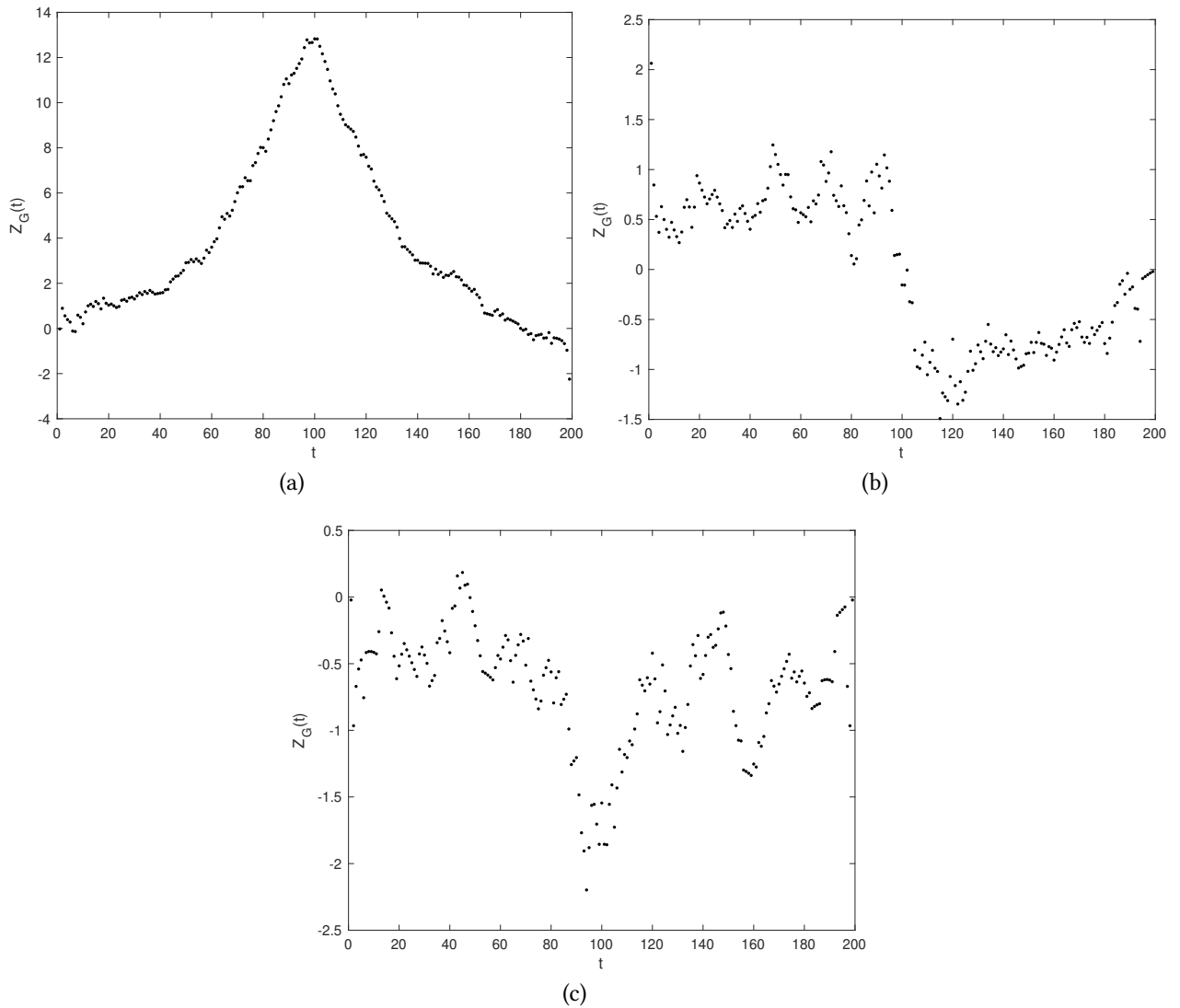


Figura 4.3: Il grafico della funzione $Z_G(t)$ dove G è l'albero di minimo accoppiamento dei punti in \mathbb{R}^2 estratti in tre differenti modi. (a) 100 punti estratti da $\mathcal{N}((0, 0), I_2)$ e poi 100 da $\mathcal{N}((2, 2), I_2)$. (b) 200 punti estratti da $\mathcal{N}((0, 0), I_2)$. (c) 200 punti estratti da $\mathcal{N}((2, 2), I_2)$.

Questa osservazione ci porta alla formulazione del seguente test statistico. Per testare l'ipotesi nulla (4.1.1) \mathcal{H}_0 contro quella alternativa (4.1.2) \mathcal{H}_1 , usiamo come statistica del test

$$T_{G, n_0, n_1} := \max_{n_0 \leq t \leq n_1} Z_G(t).$$

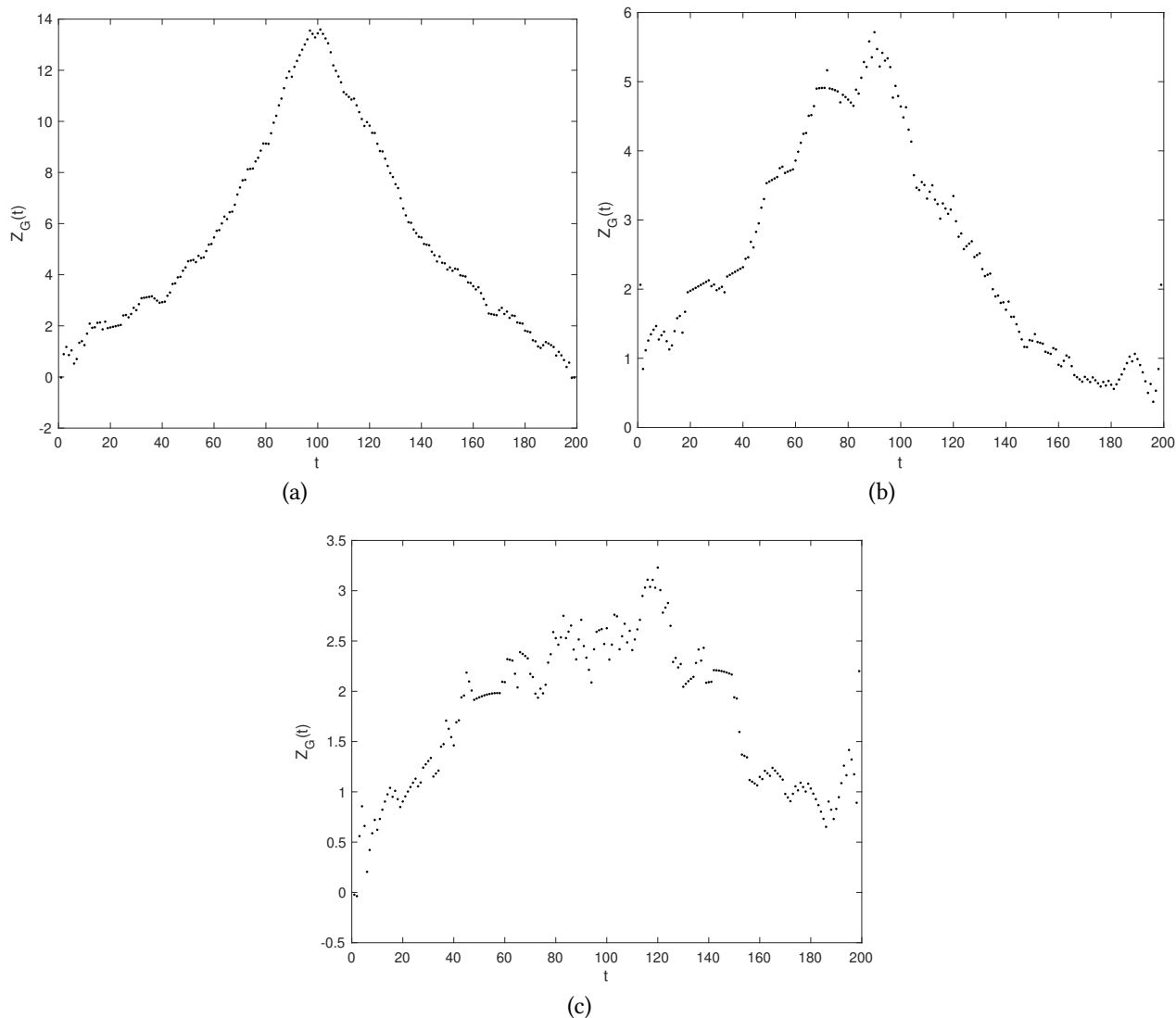


Figura 4.4: Il grafico della funzione $Z_G(t)$ dove G è il grafo del vicino più vicino dei punti in \mathbb{R}^2 estratti in tre differenti modi. (a) 100 punti estratti da $\mathcal{N}((0, 0), I_2)$ e poi 100 da $\mathcal{N}((2, 2), I_2)$. (b) 200 punti estratti da $\mathcal{N}((0, 0), I_2)$. (c) 200 punti estratti da $\mathcal{N}(2, 2), I_2)$.

L'ipotesi nulla viene rifiutata se il valore di T_{G,n_0,n_1} è maggiore di una certa soglia b . La soglia viene scelta in modo da avere un controllo sull'errore di primo tipo:

Definizione 4.13. *Dato un test d'ipotesi si definisce l'errore di primo tipo o significatività del test come la probabilità di rigettare l'ipotesi nulla quando essa è vera.*

Supponiamo di aver osservato i valori y_1, \dots, y_n . Se vale l'ipotesi nulla, osservare la stringa (y_1, \dots, y_n) ha la stessa probabilità di osservare la stringa $(y_{\pi(1)}, \dots, y_{\pi(n)})$ per ogni $\pi \in \mathcal{S}_n$. Assumendo che y_1, \dots, y_n siano valori distinti quasi certamente, possiamo identificare $(y_{\pi(1)}, \dots, y_{\pi(n)})$ con π . È naturale, quindi, considerare lo spazio di probabilità dato da \mathcal{S}_n con la distribuzione uniforme \mathbb{P}

(cioè la null distribution). La probabilità nella Definizione 4.13 è quella di $(\mathcal{S}_n, \mathbb{P})$ e la statistica del test T_{G,n_0,n_1} è una variabile aleatoria su tale spazio. Si noti, inoltre, che il grafo G costruito con la funzione di dissimilarità a partire da (y_1, \dots, y_n) è lo stesso di quello costruito a partire da $(y_{\pi(1)}, \dots, y_{\pi(n)})$.

Essendo sotto la distribuzione permutation null, il calcolo di $\mathbb{P}(T_{G,n_0,n_1} > b)$ richiede un costo dell'ordine di $n!$ operazioni. Al crescere di n tale calcolo diventa computazionalmente pesante. Grazie al Teorema 4.5.1 possiamo approssimare il processo Z_G con un processo gaussiano con una ben nota funzione di covarianza. A partire da tale approssimazione, in [8] viene ottenuto il seguente risultato:

Proposizione 4.6.1. [8, Proposizione 3.4] *Siano $n_0 = n_0(n)$, $n_1 = n_1(n)$ e $b = b(n)$. Supponiamo che esistano $0 < x_0 < x_1 < 1$ e $b_0 > 0$ tali che*

$$\lim_{n \rightarrow +\infty} \frac{n_i}{n} = x_i \quad \text{per } i = 0, 1 \quad \text{e} \quad \lim_{n \rightarrow +\infty} \frac{b}{\sqrt{n}} = b_0.$$

Sia $G = ([n], E)$ un grafo parametrizzato da $n \in \mathbb{N}$ (quindi $G = G(n)$ e $E = E(n)$) tale che

$$\sum_{e \in E} |A_e| |B_e| \sim o\left(n^{\frac{3}{2}}\right)$$

e per cui esistono finiti e positivi i seguenti due limiti

$$r_0 = \lim_{n \rightarrow +\infty} \frac{|E|}{n} \quad \text{e} \quad r_1 = \lim_{n \rightarrow +\infty} \frac{\sum_{i=1}^n d_G(i)^2}{|E|}.$$

Allora, per $n \rightarrow +\infty$, vale

$$\mathbb{P}(T_{G,n_0,n_1} > b) \approx b \phi(b) \int_{x_0}^{x_1} h^*(x) \nu\left(b_0 \sqrt{2h^*(x)}\right) dx$$

dove ϕ è la funzione di densità della gaussiana standard,

$$\nu(x) = \frac{2}{x^2} \exp\left(-2 \sum_{m=1}^{\infty} \frac{1}{m} \Phi\left(-\frac{1}{n} x \sqrt{m}\right)\right),$$

Φ è la funzione di ripartizione della gaussiana standard e

$$h^*(x) = \frac{1}{2x(1-x)} + \frac{2}{4x(1-x) + (1-2x)^2(r_1 - 4r_0)}.$$

Osservazione 5. *Per il calcolo numerico di ν possiamo utilizzare la seguente approssimazione [37]*

$$\nu(x) \approx \frac{\frac{2}{x} \left(\Phi\left(\frac{x}{2}\right) - \frac{1}{2}\right)}{\frac{x}{2} \Phi\left(\frac{x}{2}\right) + \phi\left(\frac{x}{2}\right)}.$$

Notiamo che sotto le ipotesi della Proposizione 4.6.1 le Condizione 1 e le Condizioni 2 sono banalmente verificate ponendo $\alpha = 1$.

4.7 | Dimostrazione del Lemma 4.4.1

Per non appesantire ulteriormente la dimostrazione abbrevieremo la scrittura dell'evento $\{\sigma \in \mathcal{S}_n \mid g_i(t)(\sigma) \neq g_j(t)(\sigma)\}$ con $g_i(t) \neq g_j(t)$.

Dalla linearità del valore atteso e dalla definizione di $R_G(t)$ otteniamo

$$\mathbb{E}[R_G(t)] = \sum_{\{i,j\} \in E} \mathbb{E}[\mathbb{1}_{g_i(t) \neq g_j(t)}] = \sum_{\{i,j\} \in E} \mathbb{P}(g_i(t) \neq g_j(t)). \quad (4.7.1)$$

Fissati $i \neq j \in [n]$ vale

$$g_i(t) \neq g_j(t) = A \cup B \quad (4.7.2)$$

dove

$$A := \{\pi \in \mathcal{S}_n \mid \pi(i) > t \text{ e } \pi(j) \leq t\}, \quad (4.7.3)$$

$$B := \{\pi \in \mathcal{S}_n \mid \pi(i) \leq t \text{ e } \pi(j) > t\}. \quad (4.7.4)$$

Notando che

$$A = \bigcup_{\substack{t < I \leq n \\ 1 \leq J \leq t}} \{\pi \in \mathcal{S}_n \mid \pi(i) = I \text{ e } \pi(j) = J\}$$

e usando che il numero di permutazioni di \mathcal{S}_n che fissano due elementi distinti è $(n-2)!$ otteniamo

$$|A| = \sum_{\substack{t < I \leq n \\ 1 \leq J \leq t}} (n-2)! = t(n-t)(n-2)!.$$

Per ogni coppia $a, b \in [n]$ definiamo la mappa $\psi^{(a,b)} : \mathcal{S}_n \rightarrow \mathcal{S}_n$ come

$$\psi^{(a,b)}(\pi)(h) = \begin{cases} \pi(a) & \text{se } h = b \\ \pi(b) & \text{se } h = a \\ \pi(h) & \text{altrimenti} \end{cases}. \quad (4.7.5)$$

Si verifica banalmente che ψ è una biiezione e poiché $\psi^{(i,j)}(A) = B$ vale $|A| = |B|$. Dalla definizione di distribuzione permutation null otteniamo

$$\mathbb{P}(B) = \mathbb{P}(A) = \frac{|A|}{|\mathcal{S}_n|} = \frac{t(n-t)(n-2)!}{n!} = \frac{t(n-t)}{n(n-1)} = \frac{1}{2}p_1(t). \quad (4.7.6)$$

Da (4.7.2) e (4.7.6), poiché gli eventi A e B sono disgiunti, si ha

$$\mathbb{P}(g_i(t) \neq g_j(t)) = \mathbb{P}(A) + \mathbb{P}(B) = p_1(t). \quad (4.7.7)$$

Usando (4.7.1) otteniamo l'espressione richiesta per la media.

Per il calcolo della varianza, calcoliamo prima il momento secondo:

$$\mathbb{E}[R_G(t)^2] = \sum_{\{i,j\},\{h,k\} \in E} \mathbb{P}(g_i(t) \neq g_j(t) \text{ e } g_h(t) \neq g_k(t)). \quad (4.7.8)$$

Fissati gli archi $\{i, j\}$ e $\{h, k\}$ andiamo a distinguere alcuni casi.

- Se $\{i, j\} = \{h, k\}$ allora

$$\mathbb{P}(g_i(t) \neq g_j(t) \text{ e } g_h(t) \neq g_k(t)) = \mathbb{P}(g_i(t) \neq g_j(t)) = p_1(t)$$

dove l'ultima uguaglianza deriva da (4.7.7).

- Se $h \in \{i, j\}$ e $k \notin \{i, j\}$ allora possiamo supporre, senza perdita di generalità, che $h = i$. In questo caso otteniamo

$$\{g_i(t) \neq g_j(t) \text{ e } g_i(t) \neq g_k(t)\} = C \cup D \quad (4.7.9)$$

dove

$$C := \{\pi \in \mathcal{S}_n \mid \pi(i) > t, \pi(j) \leq t \text{ e } \pi(k) \leq t\},$$

$$D := \{\pi \in \mathcal{S}_n \mid \pi(i) \leq t, \pi(j) > t \text{ e } \pi(k) > t\}.$$

Notando che

$$C = \bigcup_{\substack{t < I \leq n \\ 1 \leq J, K \leq t \\ J \neq K}} \{\pi \in \mathcal{S}_n \mid \pi(i) = I, \pi(j) = J \text{ e } \pi(k) = K\},$$

$$D = \bigcup_{\substack{1 \leq I \leq t \\ t < J, K \leq n \\ J \neq K}} \{\pi \in \mathcal{S}_n \mid \pi(i) = I, \pi(j) = J \text{ e } \pi(k) = K\}$$

e usando che il numero di permutazioni di \mathcal{S}_n che fissano tre elementi è $(n-3)!$ si ha

$$|C| = \sum_{\substack{t < I \leq n \\ 1 \leq J, K \leq t \\ J \neq K}} (n-3)! = (n-t)t(t-1)(n-3)!,$$

$$|D| = \sum_{\substack{1 \leq I \leq t \\ t < J, K \leq n \\ J \neq K}} (n-3)! = t(n-t)(n-t-1)(n-3)!.$$

Applicando la definizione di distribuzione permutation null otteniamo

$$\mathbb{P}(C) = \frac{|C|}{|\mathcal{S}_n|} = \frac{t(t-1)(n-t)}{n(n-1)(n-2)}, \quad (4.7.10)$$

$$\mathbb{P}(D) = \frac{|D|}{|\mathcal{S}_n|} = \frac{t(n-t)(n-t-1)}{n(n-1)(n-2)}. \quad (4.7.11)$$

Da (4.7.9), (4.7.10) e (4.7.11), poiché gli eventi C e D sono disgiunti, abbiamo

$$\mathbb{P}(g_i(t) \neq g_j(t) \text{ e } g_i(t) \neq g_k(t)) = \frac{t(n-t)}{n(n-1)} = \frac{1}{2}p_1(t).$$

- Se $h \notin \{i, j\}$ e $k \in \{i, j\}$ possiamo procedere come nel caso precedente ottenendo lo stesso risultato.
- Se $h, k \notin \{i, j\}$ allora

$$\begin{aligned} & \{g_i(t) \neq g_j(t) \text{ e } g_h(t) \neq g_k(t)\} \\ &= (A \cap A') \cup (A \cap B') \cup (B \cap A') \cup (B \cap B') \end{aligned} \quad (4.7.12)$$

dove A e B sono gli insiemi definiti in (4.7.3) e (4.7.4) mentre

$$\begin{aligned} A' &:= \{\pi \in \mathcal{S}_n \mid \pi(h) > t \text{ e } \pi(k) \leq t\}, \\ B' &:= \{\pi \in \mathcal{S}_n \mid \pi(h) \leq t \text{ e } \pi(k) > t\}. \end{aligned}$$

Notando che

$$\begin{aligned} A \cap A' &= \{\pi \in \mathcal{S}_n \mid \pi(i), \pi(h) > t \text{ e } \pi(j), \pi(k) \leq t\} \\ &= \bigcup_{\substack{t < I, H \leq n \\ 1 \leq J, H \leq t \\ I \neq H \\ J \neq K}} \{\pi \in \mathcal{S}_n \mid \pi(i) = I, \pi(j) = J, \pi(h) = H \text{ e } \pi(k) = K\}. \end{aligned}$$

e usando che il numero di permutazioni che fissano 4 elementi è $(n-4)!$ otteniamo

$$|A \cap A'| = \sum_{\substack{t < I, H \leq n \\ 1 \leq J, H \leq t \\ I \neq H \\ J \neq K}} (n-4)! = t(t-1)(n-t)(n-t-1)(n-4)!.$$

Notiamo che

$$\begin{aligned} \psi^{(h,k)}(A \cap A') &= A \cap B', \\ \psi^{(i,j)}(A \cap A') &= B \cap A', \\ \left(\psi^{(i,j)} \circ \psi^{(h,k)}\right)(A \cap A') &= B \cap B' \end{aligned}$$

dove $\psi^{(a,b)}$ è la biiezione definita in (4.7.5). Dunque si ha

$$|A \cap A'| = |A \cap B'| = |B \cap A'| = |B \cap B'|.$$

Dalla definizione di distribuzione permutation null si ha

$$\begin{aligned} \mathbb{P}(A \cap A') &= \mathbb{P}(A \cap B') = \mathbb{P}(B \cap A') = \mathbb{P}(B \cap B') \\ &= \frac{|A \cap A'|}{|\mathcal{S}_n|} = \frac{t(t-1)(n-t)(n-t-1)}{n(n-1)(n-2)(n-3)}. \end{aligned} \quad (4.7.13)$$

Da (4.7.12) e (4.7.13), notando che i quattro eventi sono a due a due disgiunti, otteniamo

$$\mathbb{P}(g_i(t) \neq g_j(t) \text{ e } g_h(t) \neq g_k(t)) = p_2(t).$$

Mettendo insieme i quattro punti precedenti possiamo riscrivere (4.7.8) come

$$\mathbb{E}[R_G^2(t)] = \sum_{e \in E} \left\{ p_1(t) + \sum_{\substack{e' \in E \\ |e \cap e'| = 1}} \frac{1}{2} p_1(t) + \sum_{\substack{e' \in E \\ |e' \cap e| = 0}} p_2(t) \right\}. \quad (4.7.14)$$

Ora

$$\{(e, e') \in E^2 \mid |e \cap e'| = 1\} = \bigcup_{i=1}^n \{(e, e') \in E^2 \mid e \cap e' = \{i\}\}$$

e dunque

$$|\{(e, e') \in E^2 \mid |e \cap e'| = 1\}| = \sum_{i=1}^n d_G(i) (d_G(i) - 1) \quad (4.7.15)$$

dove abbiamo usato che l'insieme degli archi che incidono in un vertice sono in bigezione con $\Gamma_G(i)$ e dunque sono $d_G(i)$. Inoltre si ha

$$\{(e, e') \in E^2 \mid e \cap e' = \emptyset\} = E^2 \setminus \{(e, e') \in E^2 \mid e \cap e' \neq \emptyset\}.$$

Utilizzando il Lemma C.2 otteniamo

$$|\{(e, e') \in E^2 \mid e \cap e' = \emptyset\}| = |E|^2 - \sum_{\{i,j\} \in E} (d_G(i) + d_G(j) - 1). \quad (4.7.16)$$

Utilizzando (4.7.15) e (4.7.16), possiamo riscrivere (4.7.14) come

$$\begin{aligned} \mathbb{E}[R_G^2(t)] &= p_1(t) |E| + \frac{1}{2} p_1(t) \sum_{i=1}^n d_G(i) (d_G(i) - 1) \\ &\quad + p_2(t) \left(|E|^2 - \sum_{\{i,j\} \in E} (d_G(i) + d_G(j) - 1) \right) \\ &= \cancel{p_1(t) |E|} + \frac{1}{2} p_1(t) \sum_{i \in V} d_G(i)^2 - \cancel{\frac{1}{2} p_1(t) |E|^2} \\ &\quad + p_2(t) \left(|E|^2 + |E| - \sum_{i=1}^n d_G(i)^2 \right) \\ &= \sum_{i=1}^n d_G(i)^2 \left(\frac{1}{2} p_1(t) - p_2(t) \right) + p_2(t) (|E|^2 + |E|). \end{aligned}$$

dove la penultima ultima uguaglianza segue dal Lemma C.1 e dal Lemma C.3. La tesi segue ricordando che per una variabile aleatoria X vale

$$\text{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2.$$

5 | Derivazione del Teorema 4.5.1

In questo capitolo, dopo aver definito una distribuzione di probabilità sulle n -uple, mimando la definizioni di R_G e Z_G , definiremo le variabili aleatorie R_G^B e Z_G^B sulle n -uple caratterizzando la loro media e varianza. Nella sezione successiva introdurremo due nuove variabili aleatorie n^B e X^B e caratterizzeremo la loro media e varianza. Le variabili aleatorie definite nelle sezioni precedenti verranno utilizzate per dimostrare formalmente un risultato di approssimazione (Teorema 5.3.1). Nella Sezione 5.4 andremo a collegare le distribuzioni bootstrap con la permutation null. Questo collegamento ci sarà utile per la dimostrazione del Teorema 4.5.1 (Sezione 5.5).

5.1 | Le variabili $R_G^B(t)$ e $Z_G^B(t)$ definito sulle n -uple

Denotiamo con $[n]^n$ l'insieme delle n -uple ordinate. Dato $i \in [n]$ sia $\pi_i : [n]^n \rightarrow [n]$ la proiezione sulla i -esima componente della n -upla.

Definizione 5.1. *Definiamo la distribuzione **bootstrap** come la distribuzione uniforme su $[n]^n$. Nel seguito indicheremo con \mathbb{P}_B tale distribuzione e con $\mathbb{E}_B, \text{Var}_B$ rispettivamente il valore atteso e la varianza sotto questa distribuzione.*

Mimando la definizione di $R_G(t)$, fissato un grafo $G = G(n) = ([n], E)$ dove $E = E(n)$ e fissato $t \in \mathbb{N}$ definiamo il contatore $R_G^B(t)$ come la variabile aleatoria

$$R_G^B(t) : [n]^n \rightarrow \mathbb{N} \quad R_G^B(t)(x) := \sum_{\{i,j\} \in E} \mathbb{1}_{\{y \in [n]^n \mid g_i(t)(y) \neq g_j(t)(y)\}}(x) \quad (5.1.1)$$

dove

$$g_i(t)(x) := \begin{cases} 1 & \text{se } \pi_i(x) > t \\ 0 & \text{altrimenti} \end{cases}.$$

Definiamo, inoltre,

$$Z_G^B(t) := -\frac{R_G^B(t) - \mathbb{E}_B [R_G^B(t)]}{\sqrt{\text{Var}_B (R_G^B(t))}}. \quad (5.1.2)$$

Lemma 5.1.1. Sia $G = G(n) = ([n], E)$ un grafo con $E = E(n)$. Fissato $t \in [n]$ vale

$$\mathbb{E}_B[R_G^B(t)] = p_1^B(t) |E| \quad (5.1.3)$$

e

$$\text{Var}_B(R_G^B(t)) = p_2^B(t) |E| + \sum_{i \in V} d_G(i)^2 \left(\frac{1}{2} p_1^B(t) - p_2^B(t) \right) \quad (5.1.4)$$

dove

$$p_1^B(t) := 2 \frac{t(n-t)}{n^2}, \quad (5.1.5)$$

$$p_2^B(t) := 4 \frac{t^2(n-t)^2}{n^4}. \quad (5.1.6)$$

Dimostrazione. Dalla linearità del valore atteso e dalla definizione di $R_G^B(t)$ otteniamo

$$\mathbb{E}_B[R_G^B(t)] = \sum_{\{i,j\} \in E} \mathbb{E}_B[\mathbb{1}_{g_i(t) \neq g_j(t)}] = \sum_{\{i,j\} \in E} \mathbb{P}_B(g_i(t) \neq g_j(t)). \quad (5.1.7)$$

Fissati $i \neq j \in [n]$ vale

$$g_i(t) \neq g_j(t) = A \cup B \quad (5.1.8)$$

dove

$$A := \{x \in [n]^n \mid \pi_i(x) > t \text{ e } \pi_i(j) \leq t\}, \quad (5.1.9)$$

$$B := \{x \in [n]^n \mid \pi_i(x) \leq t \text{ e } \pi_i(j) > t\}. \quad (5.1.10)$$

Notando che

$$A = \bigcup_{\substack{t < I \leq n \\ 1 \leq J \leq t}} \{x \in [n]^n \mid \pi_i(x) = I \text{ e } \pi_j(x) = J\}$$

e usando che il numero di sequenze n -arie con due componenti fissate è n^{n-2} otteniamo

$$|A| = \sum_{\substack{t < I \leq n \\ 1 \leq J \leq t}} n^{n-2} = t(n-t)n^{n-2}.$$

Per ogni $a, b \in [n]$ definiamo la mappa $\phi^{(a,b)} : [n]^n \rightarrow [n]^n$ dove

$$\phi^{(a,b)}(x) = y \text{ dove } \forall h \in [n] \quad \pi_h(y) := \begin{cases} \pi_a(x) & \text{se } h = b \\ \pi_b(x) & \text{se } h = a \\ \pi_h(x) & \text{altrimenti} \end{cases}. \quad (5.1.11)$$

Si verifica banalmente che $\psi^{(a,b)}$ è una biiezione e poiché $\psi^{(i,j)}(A) = B$ otteniamo $|A| = |B|$. Dalla definizione di distribuzione uniforme otteniamo

$$\mathbb{P}_B(B) = \mathbb{P}_B(A) := \frac{|A|}{|[n]^n|} = \frac{t(n-t)n^{n-2}}{n^n} = \frac{t(n-t)}{n^2}. \quad (5.1.12)$$

Da (5.1.8) e (5.1.12), poiché gli eventi A e B sono disgiunti, otteniamo

$$\mathbb{P}_B(g_i(t) \neq g_j(t)) = \mathbb{P}_B(A) + \mathbb{P}_B(B) = p_1^B(t) \quad (5.1.13)$$

Utilizzando (5.1.7) e l'identità precedente otteniamo (5.1.3).

Per il calcolo della varianza, calcoliamo prima il momento secondo:

$$\mathbb{E}_B [R_G^B(t)^2] = \sum_{\{i,j\},\{h,k\} \in E} \mathbb{P}_B(g_i(t) \neq g_j(t) \text{ e } g_h(t) \neq g_k(t)). \quad (5.1.14)$$

Fissati gli archi $\{i, j\}$ e $\{h, k\}$ andiamo a distinguere alcuni casi.

- Se $\{i, j\} = \{h, k\}$ allora

$$\mathbb{P}_B(g_i(t) \neq g_j(t) \text{ e } g_h(t) \neq g_k(t)) = \mathbb{P}_B(g_i(t) \neq g_j(t)) = p_1^B(t).$$

dove l'ultima uguaglianza segue da (5.1.13).

- Se $h \in \{i, j\}$ e $k \notin \{i, j\}$ allora possiamo supporre, senza perdita di generalità. In questo caso vale

$$\{g_i(t) \neq g_j(t) \text{ e } g_i(t) \neq g_k(t)\} = C \cup D \quad (5.1.15)$$

dove

$$C := \{x \in [n]^n \mid \pi_i(x) > t, \pi_j(x) \leq t \text{ e } \pi_k(x) \leq t\},$$

$$D := \{x \in [n]^n \mid \pi_i(x) \leq t, \pi_j(x) > t \text{ e } \pi_k(x) > t\}.$$

Notando che

$$C = \bigcup_{\substack{t < I \leq n \\ 1 \leq J, K \leq t}} \{\pi \in S_n \mid \pi(i) = I, \pi(j) = J \text{ e } \pi(k) = K\},$$

$$D = \bigcup_{\substack{1 \leq I \leq t \\ t < J, K \leq n}} \{\pi \in S_n \mid \pi(i) = I, \pi(j) = J \text{ e } \pi(k) = K\}$$

e usando che il numero di sequenze n -arie con tre componenti fissate è n^{n-3} otteniamo

$$|C| = \sum_{\substack{t < I \leq n \\ 1 \leq J, K \leq t}} n^{n-3} = (n-t)t^2 n^{n-3},$$

$$|D| = \sum_{\substack{1 \leq I \leq t \\ t < J, K \leq n}} n^{n-3} = t(n-t)^2 n^{n-3}.$$

Dalla definizione di distribuzione bootsrap otteniamo

$$\mathbb{P}_B(C) = \frac{|C|}{|[n]^n} = \frac{t^2(n-t)}{n^3}, \quad (5.1.16)$$

$$\mathbb{P}_B(D) = \frac{|D|}{|[n]^n} = \frac{t(n-t)^2}{n^3}. \quad (5.1.17)$$

Da (5.1.15), (5.1.16) e (5.1.17), notando che gli eventi C e D sono disgiunti, abbiamo

$$\mathbb{P}_B(g_i(t) \neq g_j(t) \text{ e } g_i(t) \neq g_k(t)) = \frac{1}{2}p_1^B(t).$$

- Se $h \notin \{i, j\}$ e $k \in \{i, j\}$ possiamo procedere come nel caso precedente ottenendo lo stesso risultato.
- Se $h, k \notin \{i, j\}$ allora

$$\{g_i(t) \neq g_j(t) \text{ e } g_h(t) \neq g_k(t)\} = (A \cap A') \cup (A \cap B') \cup (B \cap A') \cup (B \cap B') \quad (5.1.18)$$

dove A e B sono gli insiemi definiti in (5.1.9) e (5.1.10) mentre

$$A' := \{x \in [n]^n \mid \pi_h(x) > t \text{ e } \pi_k(x) \leq t\},$$

$$B' := \{x \in [n]^n \mid \pi_h(x) \leq t \text{ e } \pi_k(x) > t\}.$$

Notando che

$$\begin{aligned} A \cap A' &= \{x \in [n]^n \mid \pi_i(x), \pi_h(x) > t \text{ e } \pi_j(x), \pi_k(x) \leq t\} \\ &= \bigcup_{\substack{t < I, H \leq n \\ 1 \leq J, K \leq t}} \{x \in [n]^n \mid \pi_i(x) = I, \pi_j(x) = J, \pi_h(x) = H \text{ e } \pi_k(x) = K\}. \end{aligned}$$

e usando che il numero di sequenze n -arie con 4 componenti fissate è n^{n-4} otteniamo

$$|A \cap A'| = \sum_{\substack{t < I, H \leq n \\ 1 \leq J, K \leq t}} n^{n-4} = t^2(n-t)^2 n^{n-4}.$$

Poiché

$$\begin{aligned} \phi^{(h,k)}(A \cap A') &= A \cap B', \\ \phi^{(i,j)}(A \cap A') &= B \cap A', \\ \left(\phi^{(h,k)} \circ \phi^{(i,j)}\right)(A \cap A') &= B \cap B' \end{aligned}$$

dove $\phi^{(a,b)}$ è la biiezione definita in (5.1.11), abbiamo

$$|A \cap A'| = |A \cap B'| = |B \cap A'| = |B \cap B'|.$$

Dalla definizione di distribuzione bootstrap si ha

$$\begin{aligned}\mathbb{P}_B(A \cap A') &= \mathbb{P}_B(A \cap B') = \mathbb{P}_B(B \cap A') = \mathbb{P}_B(B \cap B') \\ &= \frac{|A \cap A'|}{|[n]^n} = \frac{t^2(n-t)^2}{n^4}.\end{aligned}\quad (5.1.19)$$

Da (5.1.18) e (5.1.19), notando che i quattro eventi sono a due a due disgiunti, otteniamo

$$\mathbb{P}_B(g_i(t) \neq g_j(t) \text{ e } g_h(t) \neq g_k(t)) = 4 \frac{t^2(n-t)^2}{n^4} = p_2^B(t).$$

Mettendo insieme i quattro punti precedenti possiamo riscrivere (5.1.14) come

$$\mathbb{E}_B [R_G^B(t)^2] = \sum_{e \in E} \left\{ p_1^B(t) + \sum_{\substack{e' \in E \\ |e \cap e'|=1}} \frac{1}{2} p_1^B(t) + \sum_{\substack{e' \in E \\ e' \cap e = \emptyset}} p_2^B(t) \right\}. \quad (5.1.20)$$

Utilizzando le identità (4.7.15) e (4.7.16), provate nel Lemma 4.4.1, otteniamo

$$\begin{aligned}\mathbb{E}_B [R_G^B(t)^2] &= p_1^B(t)|E| + \frac{1}{2} p_1^B(t) \sum_{i=1}^n d_G(i)(d_G(i) - 1) \\ &\quad + p_2^B(t) \left(|E|^2 - \sum_{\{i,j\} \in E} (d_G(i) + d_G(j) - 1) \right) \\ &= \cancel{p_1^B(t)|E|} + \frac{1}{2} p_1^B(t) \sum_{i=1}^n d_G(i)^2 - \cancel{\frac{1}{2} p_1^B(t) 2|E|} \\ &\quad + p_2^B(t) \left(|E|^2 + |E| - \sum_{i=1}^n d_G(i)^2 \right) \\ &= \sum_{i=1}^n d_G(i)^2 \left(\frac{1}{2} p_1^B(t) - p_2^B(t) \right) + p_2^B(t) \left(|E|^2 + |E| \right).\end{aligned}$$

dove l'ultima uguaglianza segue dal Lemma C.1 e dal Lemma C.3 mentre l'ultima dal fatto che $p_1^B(t)^2 = p_2^B(t)$. La tesi segue ricordando che per una variabile aleatoria X vale

$$\text{Var}(X) = \mathbb{E} [X^2] - \mathbb{E} [X]^2.$$

□

Osservazione 6. Un semplice calcolo prova che (5.1.4) può essere riscritto come

$$\text{Var}_B (R_G^B(t)) = \frac{t(n-t)}{n^4} \left(4t(n-t) |E| + \sum_{i=1}^n d_G(i)^2 (n-2t)^2 \right). \quad (5.1.21)$$

Poiché $t, n-t \geq 0$ allora vale

$$\text{Var}_B (R_G^B(t)) \geq p_2^B(t) |E|.$$

5.2 | Le variabili aleatorie $n^B(t)$ e $X^B(t)$

Fissato $t \in \mathbb{N}$ definiamo il contatore $n^B(t)$ come la variabile aleatoria

$$n^B(t) : [n]^n \rightarrow \mathbb{N} \quad n^B(t)(x) = \sum_{i=1}^n \mathbb{1}_{\{y \in [n]^n \mid \pi_i(y) \leq t\}}(x)$$

e la variabile aleatoria $X^B(t)$ come

$$X^B(t) := \frac{n^B(t) - \mathbb{E}_B[n^B(t)]}{\sqrt{\text{Var}_B(n^B(t))}}. \quad (5.2.1)$$

Il Lemma 5.2.1 ci fornirà un modo per scrivere $X^B(t)$ in maniera più semplice.

Lemma 5.2.1. Fissato $t \in [n]$ vale

$$\mathbb{E}_B[n^B(t)] = t \quad (5.2.2)$$

e

$$\text{Var}_B(n^B(t)) = \frac{t(n-t)}{n}. \quad (5.2.3)$$

Dimostrazione. Dalla definizione di $n^B(t)$ e dalla linearità del valore atteso otteniamo

$$\begin{aligned} \mathbb{E}_B[n^B(t)] &= \mathbb{E}_B \left[\sum_{i=1}^n \mathbb{1}_{\{x \in [n]^n \mid \pi_i(x) \leq t\}} \right] \\ &= \sum_{i=1}^n \mathbb{P}_B(\{x \in [n]^n \mid \pi_i(x) \leq t\}). \end{aligned} \quad (5.2.4)$$

Notando che

$$A := \{x \in [n]^n \mid \pi(x) \leq t\} = \bigcup_{1 \leq I \leq t} \{x \in [n]^n \mid \pi_i(x) = I\} \quad (5.2.5)$$

e usando che il numero di n -uple con una coordinata fissata è n^{n-1} otteniamo

$$|A| = \sum_{1 \leq I \leq t} n^{n-1} = tn^{n-1}.$$

Da (5.2.4) e dalla definizione di distribuzione bootstrap otteniamo

$$\mathbb{P}_B(A) = \sum_{1 \leq I \leq t} \frac{n^{n-1}}{n^n} = \frac{t}{n}. \quad (5.2.6)$$

Per il calcolo della varianza, calcoliamo prima il momento secondo:

$$\begin{aligned} \mathbb{E}_B [n^B(t)^2] &= \mathbb{E}_B \left[\left(\sum_{i=1}^n \mathbb{1}_{\{x \in [n]^n \mid \pi_i(x) \leq t\}} \right) \left(\sum_{j=1}^n \mathbb{1}_{\{x \in [n]^n \mid \pi_j(x) \leq t\}} \right) \right] \\ &= \mathbb{E}_B \left[\sum_{1 \leq i, j \leq n} \mathbb{1}_{\{x \in [n]^n \mid \pi_i(x) \leq t \text{ e } \pi_j(x) \leq t\}} \right] \\ &= \sum_{1 \leq i \leq n} \mathbb{P}_B(\{x \in [n]^n \mid \pi_i(x) \leq t\}) + \\ &\quad + \sum_{\substack{1 \leq i, j \leq n \\ i \neq j}} \mathbb{P}_B(\{x \in [n]^n \mid \pi_i(x) \leq t \text{ e } \pi_j(x) \leq t\}) \end{aligned} \quad (5.2.7)$$

Ora, se $i \neq j$ si ha

$$\{x \in [n]^n \mid \pi_i(x) \leq t \text{ e } \pi_j(x) \leq t\} = \bigcup_{1 \leq I, J \leq t} \{x \in [n]^n \mid \pi_i(x) = I \text{ e } \pi_j(x) = J\}.$$

Poiché il numero di n -uple con due coordinate fissate è n^{n-2} e gli eventi che uniamo sono disgiunti otteniamo

$$\mathbb{P}_B(\{x \in [n]^n \mid \pi_i(x) \leq t \text{ e } \pi_j(x) \leq t\}) = \sum_{1 \leq I, J \leq t} \frac{n^{n-2}}{n^n} = \frac{t^2}{n^2}.$$

Utilizzando (5.2.6), (5.2.7) e l'uguaglianza precedente otteniamo

$$\mathbb{E}_B [n^B(t)^2] = t + \frac{t^2}{n^2}n(n-1) = t + t^2 - \frac{t^2}{n}. \quad (5.2.8)$$

Da (5.2.2) e (5.2.8) si ha

$$\text{Var}_B(n^B(t)) = t + t^2 - \frac{t^2}{n} - t^2 = t \left(1 - \frac{t}{n}\right).$$

□

5.3 | Un risultato di approssimazione

Sia $G = ([n], E)$ un grafo parametrizzato da $n \in \mathbb{N}$ (quindi $G = G(n)$ e $E = E(n)$) e fissata una sequenza $0 < u_1 < u_2 < \dots < u_K < 1$, poniamo $t_i := \lfloor nu_i \rfloor$. Il Teorema 5.3.1, presentato sotto, contiene un risultato di approssimazione per il seguente vettore aleatorio

$$X = X(n) = (Z_G^B(t_1), \dots, Z_G^B(t_K), X^B(t_1), \dots, X^B(t_K)) \quad (5.3.1)$$

sotto la distribuzione bootstrap.

Nel Teorema mostreremo in modo rigoroso (attraverso il metodo di Stein) che, sotto un ansatz, le combinazioni lineari delle componenti di X , opportunamente normalizzate, sono vicine alla distribuzione normale standard. La dimostrazione segue quella di [8] ma vari passaggi lasciati al lettore sono stati elaborati e aggiunti.

Ansatz 1. Per ogni $K \in \mathbb{N}$, per ogni scelta di $0 < u_1 < \dots < u_K < 1$ e per ogni $a_1, \dots, a_K, b_1, \dots, b_K \in \mathbb{R}$, poniamo

$$\tilde{W}_n := \sum_{k=1}^K \{a_k Z_G^B(\lfloor nu_k \rfloor) + b_k X^B(\lfloor nu_k \rfloor)\}$$

e $\sigma_n := \text{Var}(\tilde{W}_n) > 0$. Supponiamo che, se per una sotto-successione n_j vale $\sigma_{n_j} > 0$, allora vale

$$\liminf_{j \rightarrow +\infty} \sigma_{n_j} > 0.$$

Equivalentemente, assumeremo che esistono $C, N > 0$ per cui $\sigma_n \geq C$ per ogni $n \geq N$ con $\sigma_n > 0$.

Nel Capitolo 6 forniremo una discussione più dettagliata su questa assunzione (assunta implicitamente in [8]).

Teorema 5.3.1. Sia $G = ([n], E)$ un grafo parametrizzato da $n \in \mathbb{N}$. Supponiamo che la Condizione 1 e la Condizione 2 (cfr. Sezione 6.0.1) siano verificate. Supponendo vero l'Ansatz 1, sotto la distribuzione bootstrap ogni combinazione lineare delle componenti di X , opportunamente normalizzata, è vicina ad una variabile aleatoria normale standard. Più precisamente $\forall K \in \mathbb{N}, \forall 0 < u_1 < \dots < u_K < 1$ e $\forall a_1, \dots, a_K, b_1, \dots, b_K \in \mathbb{R}$, posti \tilde{W}_n e σ_n come nell'Ansatz 1 vale

$$\lim_{\substack{n \rightarrow +\infty \\ \sigma_n \neq 0}} d_W \left(\frac{\tilde{W}_n}{\sqrt{\sigma_n}}, Z \right) = 0$$

dove $Z \sim \mathcal{N}(0, 1)$ e d_W è la distanza di Wasserstein (cfr. Capitolo 2).

Dimostrazione. Sia $k \in [K]$ fissato. Utilizzando il Lemma 4.4.1 e la definizione (5.1.2) di Z_G^B otteniamo

$$\begin{aligned} Z_G^B(t_k) &:= - \frac{R_G^B(t_k) - \mathbb{E}_B [R_G^B(t_k)]}{\sqrt{\text{Var}_B (R_G^B(t_k))}} \\ &= - \frac{1}{\sqrt{\text{Var}_B (R_G^B(t_k))}} \left(\sum_{\substack{e \in E \\ e = \{i,j\}}} \mathbb{1}_{g_i(t_k) \neq g_j(t_k)} - p_1^B(t_k) |E| \right) \\ &= - \frac{1}{\sqrt{\text{Var}_B (R_G^B(t_k))}} \left(\sum_{\substack{e \in E \\ e = \{i,j\}}} \{ \mathbb{1}_{g_i(t_k) \neq g_j(t_k)} - p_1^B(t_k) \} \right) = \sum_{e \in E} \xi_{e,k} \end{aligned}$$

dove l'ultima uguaglianza si ha definendo, per ogni arco $e = \{i, j\}$, la variabile $\xi_{e,k}$ come

$$\xi_{e,k} := - \frac{\mathbb{1}_{g_i(t_k) \neq g_j(t_k)} - p_1^B(t_k)}{\sqrt{\text{Var}_B (R_G^B(t_k))}}. \quad (5.3.2)$$

Dall'identità (5.1.13), provata nel Lemma 5.1.1, otteniamo che

$$\mathbb{E}_B [\mathbb{1}_{g_i(t_k) \neq g_j(t_k)}] = p_1^B(t_k)$$

dunque $\mathbb{E}_B[\xi_{e,k}] = 0$. Inoltre, da (5.3.2), si ha

$$|\xi_{e,k}| \leq \frac{1}{\sqrt{\text{Var}_B (R_G^B(t_k))}} \quad (5.3.3)$$

infatti le funzioni caratteristiche assumono valori in $\{0, 1\}$ e per ogni $k \in [K]$ vale $p_1^B(t_k) \in (0, 1)$.

Ricordando la definizione (5.2.1) di $X^B(t_k)$, grazie a (5.2.3), otteniamo che

$$\begin{aligned} X^B(t_k) &:= \frac{n^B(t_k) - t_k}{\sqrt{\text{Var}_B (n^B(t_k))}} = \frac{1}{\sqrt{\text{Var}_B (n^B(t_k))}} \left(\sum_{i=1}^n \mathbb{1}_{\{x \in [n]^n \mid \pi_i(x) \leq t_k\}} - t_k \right) \\ &= \frac{1}{\sqrt{t_k \left(1 - \frac{t_k}{n}\right)}} \sum_{i=1}^n \left\{ \mathbb{1}_{\{x \in [n]^n \mid \pi_i(x) \leq t_k\}} - \frac{t_k}{n} \right\} = \sum_{i=1}^n \xi_{i,k}. \end{aligned}$$

dove l'ultima uguaglianza si ha definendo, per ogni vertice $i \in [n]$, la variabile $\xi_{i,k}$ come

$$\xi_{i,k} := \frac{\mathbb{1}_{\{x \in [n]^n \mid \pi_i(x) \leq t_k\}} - \frac{t_k}{n}}{\sqrt{t_k \left(1 - \frac{t_k}{n}\right)}}. \quad (5.3.4)$$

Dall'identità (5.2.6), provata nel Lemma 5.2.1, otteniamo che

$$\mathbb{E} \left[\mathbb{1}_{\{x \in [n]^n \mid \pi_i(x) \leq t_k\}} \right] = \frac{t_k}{n}$$

dunque $\mathbb{E}_B[\xi_{i,k}] = 0$. Inoltre, da (5.3.4), otteniamo

$$|\xi_{i,k}| \leq \frac{1}{\sqrt{t_k \left(1 - \frac{t_k}{n}\right)}} \quad (5.3.5)$$

infatti per ogni $k \in [K]$ vale $\frac{t_k}{n} \in (0, 1)$.

Siano $a = (a_1, \dots, a_K), b = (b_1, \dots, b_K) \in \mathbb{R}^K$ e definiamo la variabile aleatoria

$$\tilde{W} := \sum_{k=1}^K \{a_k Z_G^B(t_k) + b_k X^B(t_k)\} = \sum_{k=1}^K \left\{ a_k \sum_{e \in E} \xi_{e,k} + b_k \sum_{i=1}^n \xi_{i,k} \right\}.$$

Notiamo che \tilde{W} , essendo somma di variabili aleatorie centrate, è centrata. Dunque, se $\sigma^2 := \text{Var}(\tilde{W}) = 0$ allora \tilde{W} è costante pari a 0. In tal caso W è normale degenera. Supponiamo, ora, che $\sigma^2 \neq 0$ e sia $W := \frac{\tilde{W}}{\sigma}$. Ricordando le definizioni di $\xi_{e,k}$ e $\xi_{i,k}$ otteniamo

$$\begin{aligned} W &= \frac{1}{\sigma} \sum_{k=1}^K \{a_k Z_G^B(t_k) + b_k X^B(t_k)\} = \frac{1}{\sigma} \sum_{k=1}^K \left\{ a_k \sum_{e \in E} \xi_{e,k} + b_k \sum_{i=1}^n \xi_{i,k} \right\} \\ &= \frac{1}{\sigma} \sum_{e \in E} \sum_{k=1}^K a_k \xi_{e,k} + \frac{1}{\sigma} \sum_{i=1}^n \sum_{k=1}^K b_k \xi_{i,k} = \sum_{e \in E} \xi_e + \sum_{i=1}^n \xi_i \end{aligned}$$

dove abbiamo posto per ogni $s \in E \cup [n] =: \mathcal{I}$

$$\xi_s := \frac{1}{\sigma} \sum_{k=1}^K c_{s,k} \xi_{s,k} \quad \text{dove} \quad c_{s,k} := \begin{cases} a_k & \text{se } s \in E \\ b_k & \text{se } s \in [n] \end{cases}. \quad (5.3.6)$$

In particolare posto

$$C := \max_{\substack{s \in \mathcal{I} \\ k \in [K]}} |c_{s,k}|$$

e

$$\sigma_1 := \min \left\{ \min_{k \in [K]} \sqrt{\text{Var}_B(R_G^B(t_k))}, \min_{k \in [K]} \sqrt{t_k \left(1 - \frac{t_k}{n}\right)} \right\} \quad (5.3.7)$$

utilizzando (5.3.3), (5.3.5) e (5.3.6) otteniamo che per ogni $s \in \mathcal{I}$ vale

$$|\xi_s| \leq \frac{KC}{\sigma \sigma_1}. \quad (5.3.8)$$

Fissato $s \in \mathcal{I}$ definiamo gli insiemi (cf. (4.2.1), (4.2.2) e (4.2.4))

$$S_s := \begin{cases} A_e \cup e & \text{se } s = e \in E \\ \{l \in E \mid i \in l\} & \text{se } s = i \in [n] \end{cases}, \quad (5.3.9)$$

$$T_s = \begin{cases} B_e \cup \{j \in [n] \mid \exists l \in A_e, j \in l\} & \text{se } s = e \in E \\ \Gamma_G^2(i) \cup \Gamma_G(i) \cup \{i\} & \text{se } s = i \in [n] \end{cases}. \quad (5.3.10)$$

Osserviamo che $A_e \cup \{e\}$ è da intendersi come $A_e \cup \{x, y\}$ se $e = \{x, y\}$.

Per non appesantire ulteriormente la dimostrazione, le verifiche che gli insiemi appena descritti verifichino le condizioni (3.2.3a) e (3.2.3b) richieste dal Teorema 3.2.2 verranno posticipate e svolte nel Lemma 5.3.2 e nel Lemma 5.3.3.

Siano η_s e τ_s come nel Teorema 3.2.2. Utilizzando (5.3.8) otteniamo

$$\eta_s := \sum_{j \in S_s} \xi_j \leq \frac{KC}{\sigma\sigma_1} |S_s|,$$

$$\tau_s := \sum_{j \in T_s} \xi_j \leq \frac{KC}{\sigma\sigma_1} |T_s|.$$

In particolare

$$\mathbb{E}[|\xi_s \eta_s \tau_s|] \leq \left(\frac{KC}{\sigma\sigma_1}\right)^3 |S_s| |T_s|, \quad (5.3.11a)$$

$$|\mathbb{E}[\xi_s \eta_s]| \leq \mathbb{E}[|\xi_s \eta_s|] \leq \left(\frac{KC}{\sigma\sigma_1}\right)^2 |S_s|, \quad (5.3.11b)$$

$$\mathbb{E}[|\tau_s|] \leq \frac{KC}{\sigma\sigma_1} |T_s|, \quad (5.3.11c)$$

$$\mathbb{E}[|\xi_s \eta_s^2|] \leq \left(\frac{KC}{\sigma\sigma_1}\right)^3 |S_s|^2. \quad (5.3.11d)$$

Poiché le variabili ξ_s sono somme di v.a. centrate, sono centrate e poiché $\text{Var}(\tilde{W}) = 1$, applicando il Teorema 3.2.2 otteniamo

$$\begin{aligned} \sup_{h \in \text{Lip}_1(\mathbb{R})} |\mathbb{E}[h(W)] - \mathbb{E}[h(Z)]| &\leq \left(\frac{KC}{\sigma\sigma_1}\right)^3 \sum_{s \in \mathcal{I}} \{4|S_s| |T_s| + |S_s|^2\} \\ &= \left(\frac{KC}{\sigma\sigma_1}\right)^3 \sum_{s \in \mathcal{I}} \{|S_s| (4|T_s| + |S_s|)\} \end{aligned} \quad (5.3.12)$$

dove $Z \sim \mathcal{N}(0, 1)$ e le limitazioni sono state ottenute utilizzando le disuguaglianze da (5.3.11a) a (5.3.11d).

Da (5.3.9) otteniamo

$$|S_s| = \begin{cases} |A_e| + 2 & \text{se } s = e \in E \\ d_G(i) & \text{se } s = i \in [n] \end{cases} \quad (5.3.13)$$

infatti l'insieme $\{l \in E \mid i \in l\}$ è in bigezione con $\Gamma_G(i)$.

Inoltre vale

$$|T_s| \leq \begin{cases} |A_e| + |B_e| + 1 & \text{se } s = e \in E \\ |\Gamma_G^2(i)| + d_G(i) + 1 & \text{se } s = i \in [n] \end{cases} \quad (5.3.14)$$

dove abbiamo usato il Lemma C.7.

Da (5.3.13) e (5.3.14) si ha che $|T_s| \geq |S_s|$ infatti $|B_e| \geq 1$ dato che $e \in B_e$. Possiamo, quindi, riscrivere (5.3.12) come

$$\sup_{h \in Lip_1(\mathbb{R})} |\mathbb{E}[h(W)] - \mathbb{E}[h(Z)]| \leq 5 \left(\frac{KC}{\sigma\sigma_1} \right)^3 \sum_{s \in \mathcal{I}} \{|S_s| |T_s|\} \quad (5.3.15)$$

Andiamo a stimare $|S_s| |T_s|$.

- Se $s = e \in E$ allora

$$\begin{aligned} |S_e| |T_e| &\leq (|A_e| + 2) (|A_e| + |B_e| + 1) \\ &= |A_e| |B_e| + |A_e|^2 + 3|A_e| + 2|B_e| + 2 \\ &\leq |A_e| |B_e| + |A_e| |B_e| + 3|A_e| |B_e| + 2|B_e| |A_e| + 2|B_e| \\ &= 9|A_e| |B_e| \end{aligned}$$

dove nell'ultima uguaglianza la disuguaglianza abbiamo usato $|A_e| \leq |B_e|$ e che $|A_e|, |B_e| \geq 1$.

- Se $s = i \in [n]$ allora siano e, e' gli archi le cui esistenze sono garantite rispettivamente dal Lemma C.5 e dal Lemma C.6. Per l'Osservazione 15 è sempre possibile scegliere i due archi in modo che $e = e'$. Da (5.3.13) e (5.3.14) otteniamo

$$\begin{aligned} |S_i| |T_i| &\leq d_G(i) (|\Gamma_G^2(i)| + d_G(i) + 1) \\ &= d_G(i) |\Gamma_G^2(i)| + d_G(i)^2 + d_G(i) \\ &\leq |A_e| |B_e| + |A_e|^2 + |A_e| \leq 3|A_e| |B_e| \end{aligned}$$

dove la penultima disuguaglianza segue dal Lemma C.5 e dal Lemma C.6 mentre l'ultima dal fatto che $|A_e| \leq |B_e|$.

Mettendo insieme i due punti, riscriviamo (5.3.15) come

$$\begin{aligned}
\sup_{h \in Lip_1(\mathbb{R})} |\mathbb{E}[h(W)] - \mathbb{E}[h(Z)]| &\leq 5 \left(\frac{KC}{\sigma\sigma_1} \right)^3 \left(\sum_{e \in E} 9 |A_e| |B_e| + \sum_{i=1}^n 3 |A_{e_i}| |B_{e_i}| \right) \\
&\leq 5 \left(\frac{KC}{\sigma\sigma_1} \right)^3 \left(\sum_{e \in E} 9 |A_e| |B_e| + 2 \sum_{e \in e} 3 |A_e| |B_e| \right) \\
&= 75 \left(\frac{KC}{\sigma\sigma_1} \right)^3 \sum_{e \in E} |A_e| |B_e|
\end{aligned}$$

dove l'ultima disuguaglianza deriva dal fatto che ogni arco può connettere al massimo due nodi.

Andiamo a stimare dall'alto $\sigma_1 = \sigma_1(n)$. Utilizzando l'Osservazione 6 e (5.3.7) otteniamo

$$\sigma_1 \geq \min \left\{ \min_{k \in [K]} \sqrt{p_2^B(t_k) |E|}, \min_{k \in [K]} \sqrt{t_k \left(1 - \frac{t_k}{n} \right)} \right\}$$

Ora

$$t_k \left(1 - \frac{t_k}{n} \right) = \lfloor nu_k \rfloor \left(1 - \frac{\lfloor nu_k \rfloor}{n} \right) \geq nu_k \left(1 - \frac{nu_k - 1}{n} \right) \geq nu_k(1 - u_k)$$

Abbiamo, dunque, che esiste una costante C_1 tale che

$$\min_{k \in [K]} \sqrt{t_k \left(1 - \frac{t_k}{n} \right)} \geq C_1 \sqrt{n} \quad \forall n \geq 1. \quad (5.3.16)$$

Mentre, per la stima di $p_2^B(t_k)$, abbiamo

$$\begin{aligned}
p_2^B(t_k) &= \frac{2 \lfloor nu_k \rfloor^2 (n - \lfloor nu_k \rfloor)^2}{n^4} \geq \frac{2 (nu_k)^2 (n - nu_k + 1)^2}{n^4} \\
&= \frac{2u_k^2 (n^2 + n^2 u_k^2 + 1 - 2n^2 u_k + 2n - 2nu_k)}{n^2} \geq 2u_k^2 (1 + u_k^2 - 2u_k)
\end{aligned}$$

dove per l'ultima disuguaglianza abbiamo usato che $u_k < 1$. Esiste, dunque, una costante C_2 (dipendente solo da u_1, \dots, u_K) tale che

$$\min_{k \in [K]} \sqrt{p_2^B(t_k)} \geq C_2.$$

Ricordando la Condizione 1 deve esistere una costante C_3 tale che

$$\min_{k \in [K]} \sqrt{p_2^B(t_k) |E|} \geq C_3 n^{\frac{\alpha}{2}}. \quad (5.3.17)$$

Mettendo insieme (5.3.16) e (5.3.17) otteniamo

$$\sigma_1 = \sigma_1(n) \geq Cn^{\frac{1}{2}(\alpha \wedge 1)} \quad (5.3.18)$$

dove C è una costante indipendente da n .

Possiamo riscrivere (5.3.15) come

$$\sup_{h \in Lip_1(\mathbb{R})} |\mathbb{E}[h(W)] - \mathbb{E}[h(Z)]| \leq 75 \left(\frac{KD}{\sigma} \right)^3 \frac{1}{n^{\frac{1}{2}(\alpha \wedge 1)}} \sum_{e \in E} |A_e| |B_e|$$

Supponendo vero l'Ansatz 1: σ è limitata dall'alto da una costante. L'equazione precedente diventa allora

$$\sup_{h \in Lip_1(\mathbb{R})} |\mathbb{E}[h(W)] - \mathbb{E}[h(Z)]| \leq 75E \frac{1}{n^{\frac{1}{2}(\alpha \wedge 1)}} \sum_{e \in E} |A_e| |B_e|.$$

Ricordando la Condizione 2 vale

$$\sup_{h \in Lip_1(\mathbb{R})} |\mathbb{E}[h(W) - h(Z)]| \rightarrow 0 \text{ per } n \rightarrow +\infty.$$

□

Prima di provare che gli insiemi S_s e T_s definiti rispettivamente in (5.3.9) e (5.3.10) verifichino le condizioni del Teorema 3.2.2 facciamo due osservazioni preliminari che seguono immediatamente dalle definizioni di $\xi_{s,k}$ e ξ_s con $s \in \mathcal{I}$.

Osservazione 7. Se $e = \{i, j\} \in E$ allora ξ_e dipende unicamente dalle entrate in posizione i e j .

Osservazione 8. Se $i \in [n]$ allora ξ_i dipende unicamente dall'entrata in posizione i .

Lemma 5.3.2. Siano S_s e T_s gli insiemi definiti rispettivamente in (5.3.9) e (5.3.10). Se ξ_s è definito come in (5.3.6) allora $\xi_s \Pi (\xi_p)_{p \in S_s^c}$.

Dimostrazione. Fissato $s \in \mathcal{I}$ andiamo a distinguere i due possibili casi

- Se $s = e = \{i, j\} \in E$ allora dall'Osservazione 7 abbiamo che ξ_s dipende dalle entrate in posizione i e j . Gli unici ξ_p dipendenti da ξ_e sono ξ_i, ξ_j e ξ_l con $l \in E$ tali che $l \cap e \neq \emptyset$. Ricordando la definizione 5.3.9 di S_e otteniamo che ξ_e è dipendente solo da ξ_p con $p \in S_e$.
- Se $s = i \in [n]$ allora dall'Osservazione 8 abbiamo che ξ_i dipende solo dall'entrata in posizione i . Gli unici ξ_p dipendenti da ξ_i sono quelli con $p \in E$ con $i \in p$. Ricordando la definizione 5.3.9 di S_i otteniamo che ξ_i è dipendente solo da ξ_p con $p \in S_i$.

□

Lemma 5.3.3. *Siano S_s e T_s gli insiemi definiti rispettivamente in (5.3.9) e (5.3.10). Se ξ_s è definito come in (5.3.6) allora $(\xi_j)_{j \in S_s} \amalg (\xi_p)_{p \in T_s^c}$.*

Dimostrazione. Fissato $s \in \mathcal{I}$ andiamo a distinguere i due possibili casi

- Sia $s = e = \{i, j\} \in E$. Dalle osservazioni 7 e 8 e dal lemma precedente sappiamo che $(\xi_j)_{j \in S_s}$ dipende unicamente dalle entrate in posizione \mathcal{C}_e dove

$$\mathcal{C}_e := \{p \in [n] \mid \exists l \in A_e \ p \in l\}.$$

- Sia $p \in E$ allora ξ_p dipende da un'entrata in una posizione presente nell'insieme \mathcal{C}_e se e solo se esiste $x \in \mathcal{C}_e$ con $x \in p$; ciò equivale a richiedere che $p \in B_e$.
- Sia $p \in [n]$ allora ξ_p dipende da un'entrata in una posizione presente nell'insieme \mathcal{C}_e se e solo se $p \in \mathcal{C}_e$.

Mettendo insieme queste due osservazioni si ha che $(\xi_j)_{j \in S_s}$ dipende solo dagli ξ_p dove $p \in B_e \cap \mathcal{C}_e = T_s$.

- Sia $s = i \in [n]$. Dalle osservazioni 7 e 8 e dal lemma precedente sappiamo che $(\xi_j)_{j \in S_s}$ dipende unicamente dalle entrate in posizione \mathcal{C}_i dove

$$\mathcal{C}_i := \{p \in [n] \mid \{i, p\} \in E\} \cup \{i\}.$$

- Sia $e \in E$ allora ξ_e dipende dall'entrate in una posizione in \mathcal{C}_i se e solo se $i \in e$ oppure esiste un $j \in \Gamma_G(i)$ con $j \in e$. In altre parole, ricordando la definizione (4.2.4) di $\Gamma_G^2(i)$, si ha che ξ_e dipende dall'entrate in una posizione in \mathcal{C}_i se e solo se $e \in \Gamma_G^2(i)$.
- Sia $p \in [n]$ allora ξ_p dipende da un'entrata in una posizione presente nell'insieme \mathcal{C}_i se e solo se $p \in \Gamma_G(i) \cup \{i\}$.

Mettendo insieme queste due osservazioni si ha che $(\xi_j)_{j \in S_s}$ dipende solo dagli ξ_p con $p \in \Gamma_G(i) \cup \Gamma_G^2(i) \cup \{i\}$.

□

Osservazione 9. *Il Teorema 5.3.1 non prova una convergenza del vettore X ad un vettore gaussiano. Possiamo solamente dire che il vettore $X = X(n)$ è ben approssimato da un certo vettore gaussiano.*

5.4 | Legami tra le due distribuzioni: bootstrap e permutation null

Il seguente lemma mette in collegamento $R_G(t)$ con $R_G^B(t)$ e ci permetterà di utilizzare i risultati provati precedentemente per $R_G^B(t)$ anche per $R_G(t)$.

Lemma 5.4.1. *Sia $G = ([n], E)$ un grafo parametrizzato da $n \in \mathbb{N}$ (quindi $G = G(n)$ e $E = E(n)$) che verifica la Condizione 1 e la Condizione 2. Se $t = ln$ con $l > 0$ fissato allora*

$$\frac{\text{Var}_B(R_G^B(t))}{\text{Var}(R_G(t))} \rightarrow 1 \text{ per } n \rightarrow +\infty, \quad (5.4.1)$$

$$\frac{\mathbb{E}_B[R_G^B(t)] - \mathbb{E}[R_G(t)]}{\sqrt{\text{Var}_B(R_G^B(t))}} \rightarrow 0 \text{ per } n \rightarrow +\infty. \quad (5.4.2)$$

Dimostrazione. Una facile verifica mostra che

$$\lim_{n \rightarrow +\infty} p_1(t) = \lim_{n \rightarrow +\infty} p_1^B(t) = l(1-l),$$

$$\lim_{n \rightarrow +\infty} p_2(t) = \lim_{n \rightarrow +\infty} p_2^B(t) = l^2(1-l)^2.$$

Ricordando le espressioni (4.4.3) e (5.1.4) calcolate rispettivamente per la varianza di R_G e R_G^B otteniamo (5.4.1).

Andiamo a provare (5.4.2). Da (4.4.2) e (5.1.3) otteniamo

$$\begin{aligned} \mathbb{E}_B[R_G^B(t)] - \mathbb{E}[R_G(t)] &= |E| (p_1^B(t) - p_1(t)) = -|E| \frac{2t(n-t)}{n^2(n-1)} \\ &= |E| \frac{2l(l-1)}{n-1}. \end{aligned}$$

Utilizzando (5.1.4) e il fatto che $t = nl$ otteniamo

$$\begin{aligned} \text{Var}_B(R_G^B(t)) &= 4 \frac{n^2 l^2 (n-nl)^2}{n^4} |E| + \left(\frac{nl(n-nl)}{n^2} - 4 \frac{n^2 l^2 (n-nl)^2}{n^4} \right) \sum_{i \in V} d_G(i)^2 \\ &= 4l^2(1-l)^2 |E| + (l(1-l) - 4l^2(1-l)^2) \sum_{i \in V} d_G(i)^2. \end{aligned}$$

Dunque

$$\begin{aligned}
& \frac{\mathbb{E}_B [R_G^B(t)] - \mathbb{E} [R_G(t)]}{\sqrt{\text{Var}_B(R_G^B(t))}} \\
&= - \frac{2l(l-1)|E|}{(n-1) \sqrt{4l^2(1-l)^2|E| + l(1-l)(1-2l)^2 \sum_{i \in V} d_G(i)^2}} \\
&= - \frac{2\sqrt{l(1-l)}}{(n-1) \sqrt{4l(1-l)\frac{1}{|E|} + \frac{1}{|E|^2}(1-2l)^2 \sum_{i \in V} d_G(i)^2}}.
\end{aligned}$$

Utilizzando il Teorema C.4 e la Condizione 1 otteniamo

$$\begin{aligned}
& (n-1) \sqrt{4l(1-l)\frac{1}{|E|} + \frac{1}{|E|^2}(1-2l)^2 \sum_{i \in V} d_G(i)^2} \\
& \geq \sqrt{4l(1-l)\frac{(n-1)^2}{Cn^\alpha} + (1-2l)^2\frac{(n-1)^2}{n}} \rightarrow +\infty
\end{aligned}$$

dove abbiamo usato che $\alpha < 2$.

□

Il seguente lemma ci mostra come grazie ad un opportuno condizionamento possiamo collegare le distribuzioni permutation null e bootstrap.

Lemma 5.4.2. *Fissati $n \geq t \in \mathbb{N}$ definiamo la mappa $\varphi : [n]^n \rightarrow \{L, R\}^n$ dove $\varphi(x) = \tilde{x}$ è definito dall'equazione*

$$\pi_i(\tilde{x}) = \begin{cases} L & \text{se } \pi_i(x) \leq t \\ R & \text{altrimenti} \end{cases}. \quad (5.4.3)$$

Tramite l'identificazione

$$\psi : \mathcal{S}_n \rightarrow \{x \in [n]^n \mid x \text{ ha entrate distinte}\} \quad \psi(\sigma) = (\sigma_1, \dots, \sigma_n)$$

possiamo restringere la mappa φ ad \mathcal{S}_n (con un abuso identificheremo le mappe φ e $\varphi \circ \psi$).

Sia $A \subseteq \{L, N\}^n$, se

$$\mathbb{P}_\star(A) := \mathbb{P} \left(\varphi|_{\mathcal{S}_n}^{-1}(A) \right),$$

$$\mathbb{P}_\star^B(A) := \mathbb{P} \left(\varphi^{-1}(A) \mid n^B(t) = t \right)$$

allora \mathbb{P}_\star e \mathbb{P}_\star^B coincidono.

Dimostrazione. Sia $\sigma \in \mathcal{S}_n$, Poiché $\sigma(1) \neq \dots \neq \sigma(n)$ allora $\tilde{\sigma}$ avrà t elementi uguali ad L e $n - t$ elementi uguali a R . Inoltre fissato un elemento $a \in \{L, R\}^n$ con esattamente t entrate uguali ad L e poste x_1, \dots, x_t le posizioni di queste entrate definiamo

$$A(a) = \{\sigma \in \mathcal{S}_n \mid \sigma(i) \leq t, \sigma(j) > t \forall i \in \{x_1, \dots, x_t\} \text{ e } \forall j \notin \{x_1, \dots, x_t\}\}$$

Si noti che $A(a) = \varphi_{|\mathcal{S}_n}^{-1}(a)$ e dunque per calcolare $\mathbb{P}_*(\{a\})$ dobbiamo calcolare la sua cardinalità. Per lo studio della cardinalità di $A(a)$ possiamo assumere, senza perdita di generalità che $x_i = i$ per ogni $i \in [t]$. Sia

$$\Phi : A(a) \rightarrow \mathcal{S}_t \times \mathcal{S}_{n-t} \quad \Phi(\sigma) = ((\sigma_1, \dots, \sigma_t), (t - \sigma_{t+1}, \dots, t - \sigma_n)).$$

Notiamo che la mappa è ben definita infatti essendo $a_1 = \dots = a_t = L$ per ogni permutazione $\sigma \in A(a)$ si avrà $\sigma(1), \dots, \sigma(t) \in [t]$ e dunque $(\sigma_1, \dots, \sigma_t) \in \mathcal{S}_t$. Analogamente, essendo $a_{t+1} = \dots = a_n = R$ per ogni permutazione $\sigma \in A(a)$ si avrà $\sigma(t+1), \dots, \sigma(t+n) > t$ e dunque $(t - \sigma_{t+1}, \dots, t - \sigma_n) \in \mathcal{S}_{n-t}$. Inoltre Φ è banalmente una bigezione. Abbiamo dunque

$$|A(a)| = |\mathcal{S}_t \times \mathcal{S}_{n-t}| = t!(n-t)!$$

e dunque

$$\mathbb{P}_*(a) = \begin{cases} 0 & \text{se il numero di L in } a \neq t \\ \frac{t!(n-t)!}{n!} & \text{altrimenti} \end{cases}$$

Fissato $b \in \{L, R\}^n$ e $X = \{x_1 \neq \dots \neq x_k\}$ l'insieme delle posizioni dell'entrate di b uguali a L allora

$$\varphi^{-1}(b) = \{x \in [n]^n \mid \pi_i(x) \leq t, \pi_j(x) > t \forall i \in X \text{ e } \forall j \notin X\} =: B(b)$$

Per lo studio della cardinalità di $B(b)$ possiamo assumere, senza perdita di generalità, che $x_i = i$ per ogni $i \in [k]$. Per analoghi motivi visti per Φ , risulta ben definita la mappa

$$\Phi_2 : B(b) \rightarrow [k]^k \times [n-k]^{n-k}$$

$$\Phi_2(x) = ((\pi_1(x), \dots, \pi_k(x)), (t - \pi_{k+1}(x), \dots, t - \pi_n(x)))$$

e dunque $|B(b)| = k^k(n-k)^{n-k}$.

Ricordando la definizione di \mathbb{P}_*^B otteniamo

$$\mathbb{P}_*^B(b) := \mathbb{P}^B(B(b) \mid n_B(t) = t) = \frac{\mathbb{P}^B(B(b) \cap \{n_B(t) = t\})}{\mathbb{P}^B(n_B(t) = t)}. \quad (5.4.4)$$

Notiamo che il numeratore è nullo se il numero delle entrate di b uguali a L è diverso da t dunque

$$\mathbb{P}^B(B(b) \cap \{n_B(t) = t\}) = \begin{cases} t^t(n-t)^{n-t} & \text{se il numero di L in } b = t \\ 0 & \text{altrimenti} \end{cases}. \quad (5.4.5)$$

Ora

$$\begin{aligned} \{n_B(t) = t\} &= \{x \in [n]^n \mid x \text{ ha esattamente } t \text{ elementi } \leq t\} \\ &= \bigcup_{A \in \binom{[n]}{t}} \bigcup_{\substack{x \in [t]^t \\ x=(x_1, \dots, x_t)}} \bigcup_{\substack{y \in [n-t]^{n-t} \\ y=(y_1, \dots, y_{n-t})}} \{s_{A,x,y}\} \end{aligned} \quad (5.4.6)$$

dove

$$s_{A,x,y} = (s_1, \dots, s_n) \text{ con } s_i = \begin{cases} x_j & \text{se } i \in A \text{ e } |\{z \in A \mid z \leq i\}| = j \\ t + y_j & \text{se } i \notin A \text{ e } |\{z \notin A \mid z \leq i\}| = j \end{cases}$$

cioè $s_{A,x,y} \in [n]^n$ ha gli elementi di x nelle posizioni di A e gli elementi di $t + y$ in quelle di A^C .

Da (5.4.6) e Poiché gli eventi $\{s_{A,x,y}\}$ sono disgiunti otteniamo

$$|\{n^B(t) = t\}| = \sum_{A \in \binom{[n]}{t}} \sum_{\substack{x \in [t]^t \\ x=(x_1, \dots, x_t)}} \sum_{\substack{y \in [n-t]^{n-t} \\ y=(y_1, \dots, y_{n-t})}} 1 = \binom{n}{t} t^t (n-t)^{n-t}$$

e ricordando la definizione della definizione bootstrap otteniamo

$$\mathbb{P}^B(n^B(t) = t) = \frac{\binom{n}{t} t^t (n-t)^{n-t}}{n^n}. \quad (5.4.7)$$

Utilizzando (5.4.4), (5.4.7) e (5.4.5) otteniamo

$$\mathbb{P}_\star^B(b) = \begin{cases} \frac{t!(n-t)!}{n!} & \text{se il numero di L in } b = t \\ 0 & \text{altrimenti} \end{cases}.$$

Abbiamo provato che le due distribuzioni coincidono. □

5.5 | Derivazione del Teorema 4.5.1

La derivazione del Teorema 4.5.1 non è completamente rigorosa, sebbene sia supportata da fatti matematici ben posti.

Poiché le distribuzioni di $Z_G^B(t)$ e $Z_G(t)$ dipendono esclusivamente dalle entrate minori o uguali a t , utilizzando l'identificazione data dal Lemma 5.4.2 otteniamo che la distribuzione del vettore $(Z_G^B(t_1), \dots, Z_G^B(t_K))$ condizionata all'evento $H = \{X^B(t_1) = 0, \dots, X^B(t_K) = 0\}$ coincide con la distribuzione del vettore $(A_G(t_1), \dots, A_G(t_K))$ dove

$$A_G(t) = -\frac{R_G(t) - \mathbb{E}_B[R_G^B(t)]}{\sqrt{\text{Var}_B(R_G^B(t))}}.$$

Notiamo che

$$Z_G(t) = A_G(t) \sqrt{\frac{\text{Var}_B(R_G^B(t))}{\text{Var}(R_G(t))}} - \frac{\mathbb{E}_B [R_G^B(t)] - \mathbb{E}[R_G(t)]}{\sqrt{\text{Var}(R_G(t))}}.$$

Per il Lemma 5.4.1, nel limite $n \rightarrow +\infty$, $(A_G(t_1), \dots, A_G(t_K))$ approssima $(Z_G(t_1), \dots, Z_G(t_K))$. Concludiamo che, per n grande, il vettore aleatorio $(Z_G(t_1), \dots, Z_G(t_K))$ è ben approssimato dal vettore aleatorio $(Z_G^B(t_1), \dots, Z_G^B(t_K))$ condizionato all'evento H .

D'altro canto, grazie al Teorema 5.3.1, possiamo approssimare, per n grande, il vettore X definito in (5.3.1) dal vettore gaussiano

$$\tilde{X} := \left(\tilde{Z}_G(t_1), \dots, \tilde{Z}_G(t_K), \tilde{X}(t_1), \dots, \tilde{X}(t_K) \right).$$

Possiamo, quindi, approssimare il vettore aleatorio $(Z_G^B(t_1), \dots, Z_G^B(t_K))$ condizionato all'evento H con il vettore aleatorio $(\tilde{Z}_G(t_1), \dots, \tilde{Z}_G(t_K))$ condizionato all'evento $\tilde{H} = \left\{ \tilde{X}(t_1) = 0, \dots, \tilde{X}(t_K) = 0 \right\}$. Per il Teorema B.10 quest'ultimo è un vettore gaussiano. Mettendo insieme tutte le osservazioni abbiamo che la distribuzione del vettore $(Z_G(t_1), \dots, Z_G(t_K))$ è vicina alla distribuzione normale.

Per il calcolo della funzione di covarianza si guardi il calcolo svolto nel supplemento a [8].

6 | Discussione dell'Ansatz 1

Nelle prossime sezioni verificheremo l'Ansatz 1 per alcune particolari scelte dei coefficienti a_i e b_i . In particolare nella Proposizione 6.1.4 supporremo che solamente a_i e a_j siano non nulli per qualche $i \neq j$; similmente nella Proposizione 6.2.3 assumeremo che gli unici coefficienti non nulli siano b_i e b_j . Nella Proposizione 6.3.4 mostreremo la veridicità dell'ansatz se a_i e b_i sono gli unici coefficienti diversi da zero ed infine nella Proposizione prop-2 assumeremo che gli unici coefficienti non nulli siano a_i e b_j con $i \neq j$.

Per provare i risultati sopra menzionati, nelle prossime sezioni andremo a calcolare le covarianze tra le variabili aleatorie definite nelle sezioni precedenti. Essendo le variabili $R_G^B(t)$ e $n^B(t)$ definite tramite somme di funzioni caratteristiche, per semplificare la trattazione, risulta conveniente definire per ogni $t \in \mathbb{R}$, $e \in E$ e $i \in [n]$ le seguenti variabili aleatorie:

$$\Gamma_{i,t} : [n]^n \rightarrow \{0, 1\} \quad \Gamma_{i,t}(x) := \mathbb{1}_{\pi_i(x) \leq t},$$

$$\Theta_{e,t} : [n]^n \rightarrow \{0, 1\} \quad \Theta_{e,t}(x) := \mathbb{1}_{g_i(t)(x) \neq g_j(t)(x)} \text{ dove } e = \{i, j\}.$$

Grazie a questa introduzione possiamo scrivere

$$R_G^B(t) = \sum_{e \in E} \Theta_{e,t}, \tag{6.0.1}$$

$$n^B(t) = \sum_{I=1}^n \Gamma_{i,t}. \tag{6.0.2}$$

Prima di procedere al calcolo presentiamo una stima relativa alla somma dei quadrati dei gradi di un grafo

Corollario 6.0.1 (Corollario al Teorema C.4). *Sia $(G_n)_{n \in \mathbb{N}}$ con $G_n = ([n], E_n)$ una famiglia di grafi. Se la Condizione 1 (cfr. Sezione) è verificata allora, definitivamente in n , vale*

$$\sum_{i \in [n]} d_{G_n}(i)^2 \leq 2n^{\alpha+1}.$$

Dimostrazione. Dal Teorema C.4 e ricordando la Condizione 1 otteniamo

$$\sum_{i \in n} d_G(i)^2 \leq |E| \left(\frac{2|E|}{n-1} + n \right) \leq n^{\alpha+1} \left(\frac{2n^\alpha}{n(n-1)} + 1 \right).$$

Ricordando che $0 < \alpha < 2$ otteniamo che $\frac{2n^\alpha}{n(n-1)} \rightarrow 0$ e dunque, per $n \gg 1$, vale $\frac{2n^\alpha}{n(n-1)} \leq 1$. □

Corollario 6.0.2. *Sia $0 < u < 1$ e poniamo $t = \lfloor nu \rfloor$. Sia $(G_n)_{n \in \mathbb{N}}$ con $G_n = ([n], E_n)$ una famiglia di grafi. Se la Condizione 1 è verificata allora, definitivamente in n , vale*

$$\text{Var}_B (R_{G_n}^B(t)) \leq Cn^{\alpha+1}$$

dove $C = C(u) > 0$ è indipendente da n .

Dimostrazione. Da (5.1.21) otteniamo

$$\begin{aligned} \text{Var}_B (R_{G_n}^B(t)) &= \frac{4t^2(n-t)^2}{n^4} |E| + \left(\sum_{i \in [n]} d_G^2(i) \right) \frac{t(n-t)(n-2t)^2}{n^4} \\ &\leq \frac{4n^2 u^2 (n - nu + 1)^2}{n^4} |E| \\ &\quad + \left(\sum_{i \in [n]} d_G(i)^2 \right) \frac{nu(n - nu + 1)(n - 2nu + 2)^2}{n^4} \\ &= 4u^2 \left(1 - u + \frac{1}{n} \right)^2 |E| \\ &\quad + \left(\sum_{i \in [n]} d_G^2(i) \right) u \left(1 - u + \frac{1}{n} \right) \left(1 - 2u + \frac{2}{n} \right) \\ &\leq u^2 (2 - u)^2 |E| + \frac{1}{2} \left(\sum_{i \in [n]} d_G^2(i) \right) u (2 - u) (1 - u)^2 \end{aligned}$$

dove la prima disuguaglianza segue dal fatto che $x - 1 \leq \lfloor x \rfloor \leq x$ mentre per l'ultima abbiamo usato che se $n \gg 1$ allora vale $\frac{1}{n} \leq \frac{u}{2}$. Utilizzando il corollario precedente, la Condizione 1 e poiché $0 < u < 1$ otteniamo

$$\begin{aligned} \text{Var}_B (R_{G_n}^B(t)) &\leq 2n^\alpha + 2n^{\alpha+1}(1-u)^2 = 2n^{\alpha+1} \left(\frac{1}{n} + (1-u)^2 \right) \\ &\leq 2n^{\alpha+1} \left(u^2 + 1 - \frac{3}{2}u \right) \end{aligned}$$

dove l'ultima disuguaglianza segue dal fatto che per $n \gg 1$ vale $\frac{1}{n} \leq \frac{u}{2}$. Ponendo $C = 2(u^2 + 1 - \frac{3}{2}u) > 0$ otteniamo la tesi. \square

6.1 | $Z_G^B(t)$ e $Z_G^B(s)$

Lemma 6.1.1. *Siano $t < s \in [n]$ e siano $e, l \in E$. Allora vale*

$$\text{Cov}_B(\Theta_{e,t}, \Theta_{l,s}) = \begin{cases} 0 & \text{se } e \cap l = \emptyset \\ \frac{t(n-s)(n-2s)(n-2t)}{n^4} & \text{se } |e \cap l| = 1. \\ \frac{2t(n-s)(n^2-2ns+2st)}{n^4} & \text{se } e = l \end{cases}$$

Dimostrazione. Sia $e = \{i, j\}$ e sia $l = \{a, b\}$. Andiamo a distinguere i possibili casi.

- Supponiamo che $e \cap l = \emptyset$. Notiamo che $\Theta_{e,t}(x)$ dipende solo da $\pi_i(x)$ e $\pi_j(x)$ mentre $\Theta_{l,s}(x)$ dipende solo da $\pi_a(x)$ e $\pi_b(x)$. Poiché siamo sotto la distribuzione bootstrap (cfr. Sezione 5.1) allora le due variabili aleatorie risultano indipendenti.
- Se $|e \cap l| = 1$ allora possiamo supporre, senza perdita di generalità, che $a = i$ e $b \neq i, j$. Per calcolare la covarianza, calcoliamo prima $\mathbb{E}_B[\Theta_{e,t}\Theta_{l,s}]$. Notiamo che

$$\mathbb{E}_B[\Theta_{e,t}\Theta_{l,s}] = \mathbb{P}_B(A) \tag{6.1.1}$$

dove

$$A = \{x \in [n]^n \mid g_i(t)(x) \neq g_j(t)(x) \text{ e } g_i(s)(x) \neq g_b(s)(x)\}.$$

Poiché $t < s$, ricordando la definizione di g_i , si ha per $x \in [n]^n$

$$g_i(s)(x) = 1 \quad \Rightarrow \quad g_i(t)(x) = 1, \tag{6.1.2a}$$

$$g_i(t)(x) = 0 \quad \Rightarrow \quad g_i(s)(x) = 0. \tag{6.1.2b}$$

Da cui $A = B \cup C \cup D$ dove

$$\begin{aligned} B &:= \{x \in [n]^n \mid g_i(s)(x) = 1 \text{ e } g_j(t), g_b(s)(x) = 0\} \\ &= \{x \in [n]^n \mid \pi_i(x) > s, \pi_j(x) \leq t \text{ e } \pi_b(x) \leq s\}, \end{aligned}$$

$$\begin{aligned} C &:= \{x \in [n]^n \mid g_i(t)(x) = 0 \text{ e } g_j(t), g_b(s)(x) = 1\} \\ &= \{x \in [n]^n \mid \pi_i(x) \leq t, \pi_j(x) > t \text{ e } \pi_b(x) > s\}, \end{aligned}$$

$$\begin{aligned} D &:= \{x \in [n]^n \mid g_i(t)(x) = 1, g_i(s)(x) = 0, g_j(t)(x) = 1 \text{ e } g_b(s)(x) = 1\} \\ &= \{x \in [n]^n \mid t < \pi_i(x) \leq s, \pi_j(x) \leq t \text{ e } \pi_b(x) > s\}. \end{aligned}$$

Ricordando che siamo sotto la distribuzione bootstrap otteniamo

$$\begin{aligned}\mathbb{P}_B(B) &= \frac{ts(n-s)}{n^3}, \\ \mathbb{P}_B(C) &= \frac{t(n-t)(n-s)}{n^3}, \\ \mathbb{P}_B(D) &= \frac{t(s-t)(n-s)}{n^3}.\end{aligned}\tag{6.1.3}$$

Da (6.1.1) e (6.1.3) e notando che gli eventi B , C e D sono a due a due disgiunti otteniamo

$$\begin{aligned}\mathbb{E}_B[\Theta_{e,t}\Theta_{l,s}] &= \frac{ts(n-s)}{n^3} + \frac{t(n-t)(n-s)}{n^3} + \frac{(s-t)t(n-s)}{n^3} \\ &= \frac{t(n-s)(n-2t+2s)}{n^3}\end{aligned}$$

e dunque

$$\begin{aligned}\text{Cov}_B(\Theta_{e,t}, \Theta_{l,s}) &= \mathbb{E}_B[\Theta_{e,t}\Theta_{l,s}] - \mathbb{E}_B[\Theta_{e,t}]\mathbb{E}_B[\Theta_{l,s}] \\ &= \frac{t(n-s)(n-2t+2s)}{n^3} - \frac{4ts(n-t)(n-s)}{n^4} \\ &= \frac{t(n-s)(n-2s)(n-2t)}{n^4}\end{aligned}$$

dove per la penultima uguaglianza abbiamo utilizzato (5.1.13).

- Supponiamo ora $e = l$ e calcoliamo $\mathbb{E}_B[\Theta_{e,t}\Theta_{e,s}]$. Notiamo che

$$\mathbb{E}_B[\Theta_{e,t}\Theta_{e,s}] = \mathbb{P}_B(A).\tag{6.1.4}$$

dove

$$A = \{x \in [n]^n \mid g_i(t)(x) \neq g_j(t)(x) \text{ e } g_i(s)(x) \neq g_j(s)(x)\}$$

Ricordando (6.1.2a) e (6.1.2b) otteniamo $A = B \cup C$ dove

$$\begin{aligned}B &:= \{x \in [n]^n \mid g_i(t)(x) = 0 \text{ e } g_j(s)(x) = 1\} \\ &= \{x \in [n]^n \mid \pi_i(x) \leq t \text{ e } \pi_j(x) > s\}, \\ C &:= \{x \in [n]^n \mid g_i(s)(x) = 1 \text{ e } g_j(t)(x) = 0\} \\ &= \{x \in [n]^n \mid \pi_i(x) > s \text{ e } \pi_j(x) \leq t\}.\end{aligned}$$

Dalla definizione di distribuzione bootstrap otteniamo

$$\mathbb{P}_B(B) = \mathbb{P}_B(C) = \frac{t(n-s)}{n^2}.\tag{6.1.5}$$

Da (6.1.4) e (6.1.5), notando che $B \cap C = \emptyset$, otteniamo

$$\mathbb{E}_B [\Theta_{e,t} \Theta_{e,s}] = \frac{2t(n-s)}{n^2}$$

e dunque

$$\begin{aligned} \text{Cov}_B (\Theta_{e,t}, \Theta_{e,s}) &= \mathbb{E}_B [\Theta_{e,t} \Theta_{e,s}] - \mathbb{E}_B [\Theta_{e,t}] \mathbb{E}_B [\Theta_{e,s}] = \\ &= \frac{2t(n-s)}{n^2} - \frac{4ts(n-t)(n-s)}{n^4} \\ &= \frac{2t(n-s)(n^2 - 2ns + 2st)}{n^4} \end{aligned}$$

dove per la penultima uguaglianza abbiamo utilizzato (5.1.13). □

Teorema 6.1.2. *Sia $t < s \in [n]$ allora*

$$\begin{aligned} \text{Cov}_B (R_G^B(t), R_G^B(s)) &= \frac{4t^2(n-s)^2}{n^4} |E| \\ &\quad + \frac{t(n-s)(n-2s)(n-2t)}{n^4} \sum_{i \in [n]} d_G(i)^2. \end{aligned}$$

Dimostrazione. Utilizzando (6.0.1) e la bilinearità della covarianza otteniamo

$$\begin{aligned} \text{Cov}_B (R_G^B(t), R_G^B(s)) &= \sum_{e \in E} \sum_{l \in E} \text{Cov}_B (\Theta_{e,t}, \Theta_{l,s}) \\ &= \sum_{e \in E} \text{Cov}_B (\Theta_{e,t}, \Theta_{e,t}) + \sum_{e \in E} \sum_{\substack{l \in E \\ |l \cap e|=1}} \text{Cov}_B (\Theta_{e,t}, \Theta_{l,s}) \\ &\quad + \sum_{e \in E} \sum_{\substack{l \in E \\ l \cap e = \emptyset}} \text{Cov}_B (\Theta_{e,t}, \Theta_{l,s}) \end{aligned}$$

Dal lemma precedente si ha

$$\text{Cov}_B (R_G^B(t), R_G^B(s))$$

$$\begin{aligned}
&= \sum_{e \in E} \frac{2t(n-s)(n^2 - 2ns + 2st)}{n^4} + \sum_{e \in E} \sum_{\substack{l \in E \\ |l \cap e|=1}} \frac{t(n-s)(n-2t)(n-2s)}{n^4} \\
&= \sum_{\{i,j\} \in E} \frac{2t(n-s)(n^2 - 2ns + 2st)}{n^4} \\
&\quad + \sum_{\{i,j\} \in E} \frac{t(n-s)(n-2t)(n-2s)}{n^4} (d_G(i) + d_G(j) - 2) \\
&= \frac{t(n-s)}{n^4} 2|E| [n^2 - 2ns + 2st - (n-2t)(n-2s)] \\
&\quad + \frac{t(n-s)}{n^4} \sum_{i \in [n]} d_G(i)^2 (n-2t)(n-2s) \\
&= \frac{t(n-s)}{n^4} \left\{ 4|E|t(n-s) + \sum_{i \in [n]} d_G(i)^2 (n-2t)(n-2s) \right\}
\end{aligned}$$

dove la seconda uguaglianza segue dal Lemma C.2 e la terza dal Lemma C.3. \square

Corollario 6.1.3. *Sia $(G_n)_{n \in \mathbb{N}}$ con $G_n = ([n], E_n)$ una famiglia di grafi. Fissati $0 < u < w < 1$ e posti $t_n = \lfloor nu \rfloor$ e $s_n = \lfloor nw \rfloor$. Se la Condizione 1 è verificata allora, definitivamente in n , vale*

$$\text{Cov}_B (R_{G_n}^B(t_n), R_{G_n}^B(s_n)) \geq E \begin{cases} n^\alpha & \text{se } \alpha > 1 \wedge (n - 2s_n > 0 \vee n - 2t_n > 0) \\ n^{2\alpha-1} & \text{se } \alpha \leq 1 \wedge (n - 2s_n > 0 \vee n - 2t_n > 0) \\ n^\alpha & \text{se } n - 2s_n = 0 \vee n - 2t_n = 0 \\ -n^{\alpha+1} & \text{se } n - 2t_n > 0 \wedge n - 2s_n < 0 \end{cases}$$

dove $E = E(u, w) > 0$ è una costante indipendente da n .

Dimostrazione. Andiamo a stimare la covarianza distinguendo i vari casi.

- Supponiamo $n - 2s_n > 0$. Poiché essendo $t_n < s_n$ anche $n - 2t_n > 0$ ed in

particolare $u, w < \frac{1}{2}$. Allora

$$\begin{aligned} \text{Cov}_B (R_G^B(t_n), R_G^B(s_n)) &= \frac{4t_n^2(n-s_n)^2}{n^4} |E| + \\ &\quad + \frac{t_n(n-s_n)(n-2t_n)(n-2s_n)}{n^4} \sum_{i \in [n]} d_G(i)^2 \\ &\geq 4 \left(u - \frac{1}{n}\right)^2 (1-w)^2 |E| \\ &\quad + \left(u - \frac{1}{n}\right) (1-w)(1-2u)(1-2w) \sum_{i \in [n]} d_G(i)^2 \end{aligned}$$

Ora se $n \gg 1$ allora $\frac{1}{n} \leq \frac{u}{2}$. Dunque

$$\begin{aligned} \text{Cov}_B (R_G^B(t_n), R_G^B(s_n)) &\geq u^2(1-w)^2 |E| \\ &\quad + \frac{u(1-w)(1-2u)(1-2w)}{2} \sum_{i \in [n]} d_G(i)^2 \\ &\geq u^2(1-w)^2 C n^\alpha \\ &\quad + 2u(1-w)(1-2u)(1-2w) C^2 n^{2\alpha-1} \\ &= A n^\alpha + B n^{2\alpha-1} \end{aligned}$$

dove l'ultima disuguaglianza segue dalla Condizione 1 e dal Teorema C.4 e l'ultima uguaglianza segue ponendo

$$A = A(u, w) := C u^2 (1-w)^2 \text{ e } B = B(u, w) := 2 C^2 u (1-w) (1-2u) (1-2w).$$

Andiamo a distinguere tre distinti casi.

– Se $\alpha > 1$ allora

$$\text{Cov}_B (R_G^B(t_n), R_G^B(s_n)) \geq B n^{2\alpha-1} \left(\frac{A}{B} n^{1-\alpha} + 1 \right).$$

Ora $\frac{A}{B} n^{1-\alpha} + 1 \rightarrow 1$. Fissato $0 < \varepsilon < 1$, per $n \gg 1$, si ha

$$\frac{A}{B} n^{1-\alpha} + 1 \leq 1 - \varepsilon.$$

Abbiamo, dunque,

$$\text{Cov}_B (R_G^B(t_n), R_G^B(s_n)) \geq B n^{\alpha-1} (1 - \varepsilon) = E n^{2\alpha-1}$$

dove $E := B(1 - \varepsilon) > 0$.

– Se $\alpha < 1$ allora

$$\text{Cov}_B (R_G^B(t_n), R_G^B(s_n)) \geq An^\alpha \left(1 + \frac{B}{A} n^{\alpha-1} \right).$$

Ora $1 + \frac{B}{A} n^{\alpha-1} \rightarrow 1$. Fissato $0 < \varepsilon < 1$, per $n \gg 1$, si ha

$$\frac{B}{A} n^{\alpha-1} + 1 \leq 1 - \varepsilon.$$

Abbiamo, dunque,

$$\text{Cov}_B (R_G^B(t_n), R_G^B(s_n)) \geq An^\alpha(1 - \varepsilon) = En^\alpha$$

dove $E := A(1 - \varepsilon) > 0$

– Se $\alpha = 1$ allora

$$\text{Cov}_B (R_G^B(t), R_G^B(s)) = n(A + B) = En$$

dove $E := A + B > 0$.

• Supponiamo $n - 2s_n = 0$ oppure $n - 2t_n = 0$. Allora vale

$$\begin{aligned} \text{Cov}_B (R_G^B(t_n), R_G^B(s_n)) &= \frac{t_n(n - s_n)}{n^4} 4|E|t(n - s) \\ &\geq \frac{(nu - 1)^2(n - nw)^2}{n^4} 4Cn^\alpha \\ &= u^2(1 - w)^2 Cn^\alpha = En^\alpha \end{aligned}$$

dove abbiamo usato che se $n \gg 1$ allora $\frac{n}{2} \leq \frac{u}{2}$ e $E := u^2(1 - w)^2 C > 0$.

• Supponiamo $n - 2t_n < 0$ allora anche $n - 2s_n < 0$. Dunque $(n - 2t_n)(n - 2s_n) > 0$. Procedendo come nel primo caso otteniamo la tesi.

• Supponiamo $n - 2s_n < 0$ e $n - 2t_n > 0$. Allora con disuguaglianze analoghe al primo caso possiamo ottenere

$$\begin{aligned} \text{Cov}_B (R_G^B(t_n), R_G^B(s_n)) &\geq 2 \frac{t_n(n - s_n)}{n^2} Cu(1 - w)n^\alpha \\ &\quad + 2 \frac{t_n(n - s_n)}{n^2} (1 - 2w) \left(1 - 2u + \frac{1}{n} \right) n^{\alpha+1} \\ &= t_n(n - s_n)n^{\alpha+1} \\ &\quad \cdot \left\{ 2Cu(1 - w)\frac{1}{n} + 2(1 - 2w) \left(1 - 2u + \frac{1}{n} \right) \right\}. \end{aligned}$$

Ma

$$2Cu(1-w)\frac{1}{n} + 2(1-2w) \left(1 - 2u + \frac{1}{n}\right) \rightarrow 2(1-2w)(1-2u) := D < 0.$$

Fissato $\varepsilon > 0$, per $n \gg 1$, si ha

$$2Cu(1-w)\frac{1}{n} + 2(1-2w) \left(1 - 2u + \frac{1}{n}\right) \geq D - \varepsilon.$$

quindi

$$\begin{aligned} \text{Cov}_B (R_G^B(t_n), R_G^B(s_n)) &\geq t_n(n-s_n)n^{\alpha+1}(D-\varepsilon) \\ &\geq nu(n-nw+1)n^{\alpha-1}(D-\varepsilon) \\ &= u \left(1-w + \frac{1}{n}\right) n^{\alpha+1}(D-\varepsilon). \end{aligned}$$

Da $1-w + \frac{1}{n} \rightarrow 1-w$, se $n \gg 1$ allora $1-w + \frac{1}{n} \geq 1-w - \varepsilon$ per un qualche fissato $\varepsilon < 1-w$. Si conclude notando che

$$\text{Cov}_B (R_G^B(t_n), R_G^B(s_n)) \geq u(1-w-\varepsilon)n^{\alpha+1}(D-\varepsilon) = En^{\alpha+1}$$

dove $E := u(1-w-\varepsilon)n^{\alpha+1}(D-\varepsilon) < 0$.

□

Proposizione 6.1.4. *Sia $(G_n)_{n \in \mathbb{N}}$ con $G_n = ([n], E_n)$ una famiglia di grafi. Fissati $0 < u < w < 1$ e posti $t_n = \lfloor nu \rfloor$ e $s_n = \lfloor nw \rfloor$. Se la Condizione 1 è verificata allora per ogni scelta di $(a, b) \in \mathbb{R}^2 \setminus \{(0, 0)\}$ esiste $\tilde{C} = \tilde{C}(a, b, u, w) \geq 0$ indipendente da n tale che, definitivamente in n , valga*

$$\text{Var}_B (aZ_{G_n}^B(t_n) + bZ_{G_n}^B(s_n)) \geq \tilde{C}.$$

Dimostrazione. Dalla linearità della varianza e dalla definizione di $Z_{G_n}^B$ e $Z_{G_n}^B(s)$ otteniamo

$$\text{Var}_B (aZ_{G_n}^B(t_n) + bZ_{G_n}^B(s_n)) = a^2 + b^2 + 2ab \frac{\text{Cov}_B (R_{G_n}^B(t_n), R_{G_n}^B(s_n))}{\sqrt{\text{Var}_B (R_{G_n}^B(s_n)) \cdot \text{Var}_B (R_{G_n}^B(t_n))}}.$$

Poniamo $x = 2ab \text{Cov}_B (R_{G_n}^B(t_n), R_{G_n}^B(s_n))$ e andiamo a distinguere 4 differenti casi

- Se $(n - 2t_n \geq 0 \vee n - 2s_n \geq 0) \wedge ab > 0$ allora dal corollario precedente si ha $x > 0$. Dunque

$$\text{Var}_B (aZ_{G_n}^B(t_n) + bZ_{G_n}^B(s_n)) \geq a^2 + b^2$$

- Supponiamo $(n - 2t_n \geq 0 \vee n - 2s_n \geq 0) \wedge ab < 0$. Dal corollario precedente si ha $\text{Cov}_B(R_{G_n}^B(t_n), R_{G_n}^B(s_n)) \geq 0$. Ricordando che per ogni coppia di variabili aleatorie (X, Y) vale

$$|\text{Cov}(X, Y)| \leq \sqrt{\text{Var}(X) \text{Var}(Y)}$$

otteniamo

$$\text{Var}_B(aZ_{G_n}^B(t_n) + bZ_{G_n}^B(s_n)) \geq a^2 + b^2 + 2ab = \tilde{C}.$$

- Se $n - 2t_n > 0 \wedge n - 2s_n < 0 \wedge ab < 0$ allora dal corollario precedente si ha $x > 0$. Dunque

$$\text{Var}_B(aZ_{G_n}^B(t_n) + bZ_{G_n}^B(s_n)) \geq a^2 + b^2 = \tilde{C} > 0$$

- Se $n - 2t_n > 0 \wedge n - 2s_n < 0 \wedge ab > 0$ allora utilizzando il Corollario 6.0.2 e il corollario precedente otteniamo

$$\text{Cov}_B(aZ_{G_n}^B(t_n), bZ_{G_n}^B(s_n)) \geq -2ab \frac{En^{\alpha+1}}{Cn^{\alpha+1}}$$

dunque

$$\text{Var}_B(aZ_{G_n}^B(t_n) + bZ_{G_n}^B(s_n)) \geq a^2 + b^2 - 2ab \frac{E}{C} = \tilde{C}$$

dove $\tilde{C} = a^2 + b^2 - 2ab \frac{E}{C}$.

- Se $n - 2t_n > 0 \wedge n - 2s_n < 0 \wedge ab < 0$ allora da (??) otteniamo

$$-1 \leq \frac{\text{Cov}_B(R_{G_n}^B(t_n), R_{G_n}^B(s_n))}{\sqrt{\text{Var}_B(R_{G_n}^B(s_n)) \cdot \text{Var}_B(R_{G_n}^B(t_n))}} \leq 1$$

e dunque

$$-ab \geq \frac{\text{Cov}_B(R_{G_n}^B(t_n), R_{G_n}^B(s_n))}{\sqrt{\text{Var}_B(R_{G_n}^B(s_n)) \cdot \text{Var}_B(R_{G_n}^B(t_n))}} \geq ab.$$

Otteniamo, quindi,

$$\text{Var}_B(aZ_{G_n}^B(t_n) + bZ_{G_n}^B(s_n)) \geq a^2 + b^2 + 2ab \geq 0.$$

□

6.2 | $X^B(t)$ e $X^B(s)$

Lemma 6.2.1. *Siano $t < s \in [n]$ e $i, j \in [n]$. Allora vale*

$$\text{Cov}_B(\Gamma_{i,t}, \Gamma_{j,s}) = \begin{cases} 0 & \text{se } i \neq j \\ \frac{t(n-s)}{n^2} & \text{se } i = j \end{cases}.$$

Dimostrazione. Andiamo a distinguere i possibili casi.

- Supponiamo che $i \neq j$. Notiamo che $\Gamma_{i,t}(x)$ dipende solo da $\pi_i(x)$ mentre $\Gamma_{j,s}(x)$ dipende solo da $\pi_j(x)$. Poiché siamo sotto la distribuzione bootstrap, le due variabili aleatorie sono indipendenti.
- Supponiamo ora che $i = j$. Per calcolare $\text{Cov}_B(\Gamma_{i,t}, \Gamma_{i,s})$ calcoliamo prima $\mathbb{E}_B[\Gamma_{i,t}\Gamma_{i,s}]$.

$$\mathbb{E}_B[\Gamma_{i,t}\Gamma_{i,s}] = \mathbb{P}_B(\pi_i(x) \leq t \text{ e } \pi_i(x) \leq s) = \mathbb{P}_B(\pi_i(x) \leq t) = \frac{t}{n}$$

dove per l'ultima uguaglianza abbiamo utilizzato (5.2.6). Dunque otteniamo

$$\begin{aligned} \text{Cov}_B(\Gamma_{i,t}, \Gamma_{i,s}) &= \mathbb{E}_B[\Gamma_{i,t}\Gamma_{i,s}] - \mathbb{E}_B[\Gamma_{i,t}]\mathbb{E}_B[\Gamma_{i,s}] \\ &= \frac{t}{n} - \frac{t}{n} \cdot \frac{s}{n} = \frac{t(n-s)}{n^2}. \end{aligned}$$

dove la penultima uguaglianza segue da (5.2.6).

□

Teorema 6.2.2. *Sia $t < s \in [n]$ allora*

$$\text{Cov}_B(n^B(t), n^B(s)) = \frac{t(n-s)}{n}.$$

Dimostrazione. Utilizzando (6.0.2) e la bilinearità della covarianza otteniamo

$$\text{Cov}_B(n^B(t), n^B(s)) = \sum_{i=1}^n \sum_{j=1}^n \text{Cov}_B(\Gamma_{i,t}, \Gamma_{j,t}).$$

Dal lemma precedente si ha

$$\text{Cov}_B(n^B(t), n^B(s)) = \sum_{i=1}^n \frac{t(n-s)}{n^2} = \frac{t(n-s)}{n}.$$

Proposizione 6.2.3. Fissati $0 < u < w < 1$ e posti $t_n = \lfloor nu \rfloor$ e $s_n = \lfloor nw \rfloor$. Allora per ogni scelta di $(a, b) \in \mathbb{R}^2 \setminus \{(0, 0)\}$ esiste $\tilde{C} = \tilde{C}(a, b, u, w) \geq 0$ indipendente da n tale che, definitivamente in n , valga

$$\text{Var}_B (aX^B(t_n) + bX^B(s_n)) \geq C.$$

Dimostrazione. Dalla definizione di $X^B(t_n)$ e $X^B(s_n)$ otteniamo

$$\begin{aligned} \text{Var}_B (aX^B(t_n) + bX^B(s_n)) &= a^2 \text{Var}_B (X^B(t_n)) + b^2 \text{Var}_B (X^B(s_n)) \\ &\quad + 2ab \text{Cov}_B (X^B(t_n), X^B(s_n)) \\ &= a^2 + b^2 + 2ab \frac{\text{Cov}_B(n^B(t_n), n^B(s_n))}{\sqrt{\text{Var}_B(n^B(t_n)) \cdot \text{Var}_B(n^B(s_n))}} \\ &= a^2 + b^2 + 2ab \frac{\frac{t_n(n-s_n)}{n}}{\sqrt{\frac{t_n(n-t_n)}{n} \cdot \frac{s_n(n-s_n)}{n}}} \\ &= a^2 + b^2 + 2ab \sqrt{\frac{t_n(n-s_n)}{s_n(n-t_n)}} \end{aligned}$$

- Sia $ab < 0$. Dal teorema precedente abbiamo

$$\text{Cov}_B(n^B(t_n), n^B(s_n)) \geq 0.$$

Ricordando che per ogni coppia di variabili aleatorie (X, Y) vale

$$|\text{Cov}(X, Y)| \leq \sqrt{\text{Var}(X) \text{Var}(Y)}$$

si ha

$$\text{Var}_B (aX^B(t_n) + bX^B(s_n)) \geq a^2 + b^2 - 2ab = C.$$

- Se $ab > 0$ allora vale

$$\begin{aligned} \text{Var}_B (aX^B(t_n) + bX^B(s_n)) &\geq a^2 + b^2 + 2ab \sqrt{\frac{t_n(n-s_n)}{s_n(n-t_n)}} \\ &\geq a^2 + b^2 + 2ab \sqrt{\frac{(nu-1)(n-nw)}{(nw-1)(n-nu)}}. \end{aligned}$$

Ora

$$\sqrt{\frac{(nu-1)(n-nw)}{(nw-1)(n-nu)}} \rightarrow \sqrt{\frac{u(1-w)}{w(1-u)}} := D > 0$$

e dunque per $n \gg 1$ abbiamo

$$\sqrt{\frac{(nu-1)(n-nw)}{(nw-1)(n-nu)}} \geq D - \varepsilon > 0.$$

Mettendo insieme quanto provato otteniamo

$$\text{Var}_B (aX^B(t_n) + bX^B(s_n)) \geq a^2 + b^2 + 2ab(D - \varepsilon) := C > 0.$$

□

6.3 | $X^B(t)$ e $Z_G^B(t)$

Lemma 6.3.1. *Siano $t, s \in \mathbb{R}$ e $e \in E$. Allora vale*

$$\text{Cov}_B (\Gamma_{i,t}, \Theta_{e,s}) = \begin{cases} 0 & \text{se } i \notin e \\ \frac{t(n-t)(n-2t)}{n^3} & \text{se } i \in e \text{ e } s = t \\ \frac{t(n-s)(n-2s)}{n^3} & \text{se } i \in e \text{ e } t < s \\ \frac{s(n-t)(n-2s)}{n^3} & \text{se } i \in e \text{ e } t > s \end{cases}.$$

Dimostrazione. Supponiamo $e = \{a, b\}$ e andiamo a distinguere i possibili casi

- Supponiamo che $i \notin e$. Notiamo che $\Gamma_{i,t}(x)$ dipende da $\pi_i(x)$ mentre $\Theta_{e,t}(x)$ dipende solo da $\pi_a(x)$ e $\pi_b(x)$. Poiché siamo sotto la distribuzione bootstrap le due variabili aleatorie sono indipendenti.
- Supponiamo che $s = t, a = i$ e $b \neq i$. Per calcolare $\text{Cov}_B (\Gamma_{i,t}, \Theta_{e,t})$ calcoliamo prima $\mathbb{E}_B [\Gamma_{i,t} \Theta_{e,t}]$.

$$\begin{aligned} \mathbb{E}_B [\Gamma_{i,t} \Theta_{e,t}] &= \mathbb{P}_B (\{x \in [n]^n \mid \pi_i(x) \leq t \text{ e } g_i(t)(x) \neq g_b(t)(x)\}) \\ &= \mathbb{P}_B (\{x \in [n]^n \mid \pi_i(x) \leq t \text{ e } \pi_b(x) > t\}) = \frac{t(n-t)}{n^2} \end{aligned}$$

dove la penultima disuguaglianza segue dalla definizione di $g_i(t)$ e l'ultima dal fatto che siamo sotto la distribuzione bootstrap. Tornando al calcolo della covarianza otteniamo

$$\begin{aligned} \text{Cov}_B (\Gamma_{i,t}, \Theta_{e,t}) &= \mathbb{E}_B [\Gamma_{i,t} \Theta_{e,t}] - \mathbb{E}_B [\Gamma_{i,t}] \mathbb{E}_B [\Theta_{e,t}] \\ &= \frac{t(n-t)}{n^2} - \frac{t}{n} \cdot \frac{2t(n-t)}{n^2} \\ &= \frac{t(n-t)(n-2t)}{n^3} \end{aligned}$$

dove la penultima uguaglianza segue da (5.2.6) e da (5.1.13).

- Supponiamo che $t < s, a = i$ e $b \neq i$. Per calcolare $\text{Cov}_B (\Gamma_{i,t}, \Theta_{e,s})$ calcoliamo prima $\mathbb{E}_B [\Gamma_{i,t} \Theta_{e,s}]$.

$$\mathbb{E}_B [\Gamma_{i,t} \Theta_{e,s}] = \mathbb{P}_B (\{x \in [n]^n \mid \pi_i(x) \leq t \text{ e } g_i(s)(x) \neq g_b(s)(x)\}).$$

Essendo $t < s$, se $\pi_i(x) \leq t$ allora $\pi_i(x) \leq s$ e dunque $g_i(s)(x) = 0$. Dunque l'identità precedente può essere riscritta come

$$\mathbb{E}_B [\Gamma_{i,t} \Theta_{e,s}] = \mathbb{P}_B (\{x \in [n]^n \mid \pi_i(x) \leq t \text{ e } \pi_b(x) > s\}) = \frac{t(n-s)}{n^2}.$$

Possiamo passare al calcolo della covarianza:

$$\begin{aligned} \text{Cov}_B (\Gamma_{i,t}, \Theta_{e,s}) &= \mathbb{E}_B [\Gamma_{i,t} \Theta_{e,s}] - \mathbb{E}_B [\Gamma_{i,t}] \mathbb{E}_B [\Theta_{e,s}] \\ &= \frac{t(n-s)}{n^2} - \frac{t}{n} \cdot \frac{2s(n-s)}{n^2} \\ &= \frac{t(n-s)(n-2s)}{n^3} \end{aligned}$$

dove la penultima uguaglianza segue da (5.2.6) e da (5.1.13).

- Supponiamo che $t > s$, $a = i$ e $b \neq i$. Per calcolare $\text{Cov}_B (\Gamma_{i,t}, \Theta_{e,s})$, calcoliamo prima $\mathbb{E}_B [\Gamma_{i,t} \Theta_{e,s}]$.

$$\mathbb{E}_B [\Gamma_{i,t} \Theta_{e,s}] = \mathbb{P}_B (A) \tag{6.3.1}$$

dove

$$A := \{x \in [n]^n \mid \pi_i(x) \leq t \text{ e } g_i(s)(x) \neq g_b(s)(x)\}.$$

Usando la definizione di g_i , si ha $A = B \cup C$ dove

$$\begin{aligned} B &:= \{x \in [n]^n \mid \pi_i(x) \leq t, \pi_i(x) > s \text{ e } \pi_b(x) \leq s\} \\ &= \{x \in [n]^n \mid s < \pi_i(x) \leq t \text{ e } \pi_b(x) \leq s\}, \\ C &:= \{x \in [n]^n \mid \pi_i(x) \leq t, \pi_i(x) \leq s \text{ e } \pi_b(x) > s\} \\ &= \{x \in [n]^n \mid \pi_i(x) \leq s \text{ e } \pi_b(x) > s\}. \end{aligned}$$

Dalla definizione di distribuzione bootstrap otteniamo

$$\mathbb{P}_B (B) = \frac{s(t-s)}{n^2}, \quad \mathbb{P}_B (C) = \frac{s(n-s)}{n^2}. \tag{6.3.2}$$

Da (6.3.1) e (6.3.2), notando che $B \cap C = \emptyset$, si ha

$$\mathbb{E}_B [\Gamma_{i,t} \Theta_{e,s}] = \frac{s(n-2s+t)}{n^2}.$$

Possiamo passare al calcolo della covarianza:

$$\begin{aligned} \text{Cov}_B (\Gamma_{i,t}, \Theta_{e,s}) &= \mathbb{E}_B [\Gamma_{i,t} \Theta_{e,s}] - \mathbb{E}_B [\Gamma_{i,t}] \mathbb{E}_B [\Theta_{e,s}] \\ &= \frac{s(n-2s+t)}{n^2} - \frac{t}{n} \cdot \frac{2s(n-s)}{n^2} \\ &= \frac{s(n^2 - 2sn - tn + 2ts)}{n^3} = \frac{s(n-t)(n-2s)}{n^3} \end{aligned}$$

dove la penultima uguaglianza segue da (5.2.6) e da (5.1.13).

□

Teorema 6.3.2. Sia $G = ([n], E)$ un grafo e $t \in \mathbb{R}$ allora

$$\text{Cov}_B(n^B(t), R_G^B(t)) = 2|E| \frac{t(n-t)(n-2t)}{n^3}.$$

Dimostrazione. Da (6.0.1), da (6.0.2) e dalla bilinearità della covarianza otteniamo

$$\text{Cov}_B(n^B(t), R_G^B(t)) = \sum_{i=1}^n \sum_{e \in E} \text{Cov}_B(\Gamma_{i,t}, \Theta_{e,t}) = \sum_{i=1}^n \sum_{\substack{e \in E \\ i \in E}} \frac{t(n-t)(n-2t)}{n^3}$$

dove l'ultima identità segue dal lemma precedente. Poiché il numero di archi incidenti al vertice i è $d_G(i)$, l'identità precedente può essere riscritta come

$$\text{Cov}_B(n^B(t), R_G^B(t)) = \sum_{i=1}^n d_G(i) \frac{t(n-t)(n-2t)}{n^3} = 2|E| \frac{t(n-t)(n-2t)}{n^3}$$

dove l'ultima uguaglianza segue dal Lemma C.1.

□

Corollario 6.3.3. Sia $(G_n)_{n \in \mathbb{N}}$ dove $G_n = ([n], E_n)$ una successione di grafi. Fissato $0 < u < 1$ si ponga $t_n = \lfloor nu \rfloor$. Se la Condizione 1 è verificata allora, definitivamente in n , vale

$$\text{Cov}_B(n^B(t_n), R_{G_n}^B(t_n)) \geq \text{sgn}(n - 2t_n) \tilde{C} n^\alpha$$

dove $\tilde{C} = \tilde{C}(u) \geq 0$ è indipendente da n .

Dimostrazione. Andiamo a distinguere due casi.

- Se $n - 2t_n \geq 0$ allora $1 - 2u \geq 0$ ed in particolare

$$\begin{aligned} \text{Cov}_B(n^B(t_n), R_{G_n}^B(t_n)) &\geq 2Cn^\alpha \frac{(nu-1)(n-nu)(n-2nu)}{n^3} \\ &= n^\alpha Cu(1-u)(1-2u). \end{aligned}$$

dove abbiamo usato che se $n \gg 1$ allora vale $\frac{1}{n} \leq u/2$. Ponendo $\tilde{C} := Cu(1-u)(1-2u)$ otteniamo la tesi.

- Se $n - 2t_n > 0$ allora $1 - 2u < 0$ ed in particolare

$$\begin{aligned} \text{Cov}_B(n^B(t_n), R_{G_n}^B(t_n)) &\geq 2Cn^\alpha \frac{(nu)(n-nu+1)(n-2nu+2)}{n^3} \\ &= n^\alpha Cu \left(1 - u + \frac{1}{n}\right) \left(1 - 2u + \frac{2}{n}\right). \end{aligned}$$

Ora $Cu \left(1 - u + \frac{1}{n}\right) \left(1 - 2u + \frac{2}{n}\right) \rightarrow u(1 - u)(1 - 2u) := E < 0$ dunque, fissato $\varepsilon > 0$, per $n \gg 1$ si ha

$$Cu \left(1 - u + \frac{1}{n}\right) \left(1 - 2u + \frac{2}{n}\right) \geq E - \varepsilon := \tilde{C}.$$

□

Proposizione 6.3.4. Sia $(G_n)_{n \in \mathbb{N}}$ dove $G_n = ([n], E_n)$ una successione di grafi. Fissato $0 < u < 1$ si ponga $t_n = \lfloor nu \rfloor$. Se la Condizione 1 è verificata allora per ogni scelta di $(a, b) \in \mathbb{R}^2 \setminus \{(0, 0)\}$ esiste $C = C(a, b, u) > 0$ indipendente da n tale che, definitivamente in n , valga

$$\text{Var}_B (aZ_{G_n}^B(t_n) + bX^B(t_n)) \geq C.$$

Dimostrazione. Dalla definizione di $X^B(t_n)$ e $Z_G^B(t_n)$ otteniamo

$$\begin{aligned} \text{Var}_B (aX^B(t_n) + bZ_G^B(t_n)) &= a^2 \text{Var}_B (X^B(t_n)) + b^2 \text{Var}_B (Z_G^B(t_n)) \\ &\quad + 2ab \text{Cov}_B (X^B(t_n), Z_G^B(t_n)) \\ &= a^2 + b^2 + \frac{2ab \text{Cov}_B (R_G^B(t_n), n^B(t_n))}{\sqrt{\text{Var}_B (R_G^B(t_n)) \cdot \text{Var}_B (n^B(t_n))}} \end{aligned}$$

Posto $x = 2ab \text{Cov}_B (R_G^B(t_n), n^B(t_n))$ andiamo a distinguere i quattro possibili casi.

- Sia $n - 2t_n > 0 \wedge ab > 0$. Utilizzando il corollario precedente, (5.2.3) e il Corollario 6.0.2 otteniamo

$$\begin{aligned} \text{Var}_B (aX^B(t_n) + bZ_G^B(t_n)) &\geq a^2 + b^2 + 2ab \frac{\tilde{C}n^\alpha}{\sqrt{Cn^{\alpha+1} \cdot \frac{t_n(n-t_n)}{n}}} \\ &\geq a^2 + b^2 + 2ab \frac{\tilde{C}n^\alpha}{\sqrt{Cn^\alpha \cdot t_n(n-t_n)}} \\ &\geq a^2 + b^2 + 2ab \frac{\tilde{C}n^\alpha}{\sqrt{Cn^{\alpha+2} \cdot u(1-u + \frac{1}{n})}}. \end{aligned}$$

Ora $Cu \left(1 - u + \frac{1}{n}\right) \rightarrow Cu(1 - u) := E$ dunque esiste $\varepsilon > 0$ tale che

$$Cu \left(1 - u + \frac{1}{n}\right) \leq E + \varepsilon.$$

Combinando i due risultati abbiamo che

$$\text{Var}_B (aX^B(t_n) + bZ_G^B(t_n)) \geq a^2 + b^2 + 2ab \frac{\tilde{C}}{\sqrt{E + \varepsilon}} n^{\frac{\alpha}{2}-1}.$$

Essendo $\alpha < 2$ (cfr. Condizione 1) allora

$$a^2 + b^2 + 2ab \frac{\tilde{C}}{\sqrt{E + \varepsilon}} n^{\frac{\alpha}{2}-1} \rightarrow a^2 + b^2.$$

Esiste, dunque, $\varepsilon < a^2 + b^2$ con

$$a^2 + b^2 + 2ab \frac{\tilde{C}}{\sqrt{E + \varepsilon}} n^{\frac{\alpha}{2}-1} \geq a^2 + b^2 - \varepsilon := \tilde{C}.$$

- Sia $n - 2t_n \geq 0 \wedge ab < 0$. Utilizzando il corollario precedente otteniamo

$$\text{Cov}_B (R_G^B(t_n), n^B(t_n)) \geq 0.$$

Poiché per ogni coppia di variabili aleatorie (X, Y) si ha

$$|\text{Cov}(X, Y)| \leq \sqrt{\text{Var}(X) \cdot \text{Var}(Y)}$$

abbiamo

$$\text{Var}_B (aX^B(t_n) + bZ_G^B(t_n)) \geq a^2 + b^2 + 2ab = \tilde{C}.$$

- Sia $n - 2t_n < 0 \wedge ab > 0$. Utilizzando il corollario precedente otteniamo

$$\text{Cov}_B (R_G^B(t_n), n^B(t_n)) \leq 0.$$

Con argomentazioni analoghe al caso precedente si ha

$$\text{Var}_B (aX^B(t_n) + bZ_G^B(t_n)) \geq a^2 + b^2 - 2ab = \tilde{C}.$$

- Sia $n - 2t_n < 0 \wedge ab < 0$. Utilizzando il Corollario 6.0.2 otteniamo

$$\begin{aligned} \text{Var}_B (aX^B(t_n) + bZ_G^B(t_n)) &\geq a^2 + b^2 + 2|ab| \frac{\tilde{C}n^\alpha}{\sqrt{Cn^{\alpha+1} \cdot \frac{t_n(n-t_n)}{n}}} \\ &\geq a^2 + b^2 + 2|ab| \frac{\tilde{C}n^\alpha}{\sqrt{Cn^\alpha \cdot t_n(n-t_n)}} \\ &\geq a^2 + b^2 + 2|ab| \frac{\tilde{C}n^\alpha}{\sqrt{Cn^{\alpha+2} \cdot u(1-u + \frac{1}{n})}}. \end{aligned}$$

Si conclude con le stesse argomentazioni del primo punto.

6.4 | $X^B(t)$ e $Z_G^B(s)$

In maniera analoga al Teorema 6.3.2 e alla Proposizione 6.3.4 si prova che

Teorema 6.4.1. *Sia $t < s$ allora*

$$\text{Cov}_B(n^B(t), R_G^B(s)) = 2|E| \frac{t(n-s)(n-2s)}{n^3}.$$

Sia $t > s$ allora

$$\text{Cov}_B(n^B(t), R_G^B(s)) = 2|E| \frac{s(n-t)(n-2s)}{n^3}.$$

Proposizione 6.4.2. *Sia $(G_n)_{n \in \mathbb{N}}$ dove $G_n = ([n], E_n)$ una successione di grafi. Fissato $0 < u < w < 1$ si ponga $t_n = \lfloor nu \rfloor$ e $s_n = \lfloor nw \rfloor$. Se la Condizione 1 è verificata allora per ogni scelta di $(a, b) \in \mathbb{R}^2 \setminus \{(0, 0)\}$ esiste $C = C(a, b, u, w) > 0$ indipendente da n tale che, definitivamente in n , valga*

$$\text{Var}_B(aZ_{G_n}^B(t_n) + bX^B(s_n)) \geq C.$$

Dimostrazione. La dimostrazione segue usando il corollario precedente e ripercorrendo la dimostrazione della Proposizione 6.3.4 sostituendo opportunamente t_n con s_n .

A | Richiami di analisi funzionale

A.1 | Assoluta continuità

Definizione A.1. Diremo che due intervalli (cioè due insiemi connessi di \mathbb{R}) $A, B \subseteq \mathbb{R}$ sono **sovrapponibili** se

$$A^\circ \cap B^\circ \neq \emptyset$$

dove con A° e B° denotiamo la parte interna rispettivamente di A e B .

Definizione A.2. Sia $I \subseteq \mathbb{R}$ un intervallo. Diremo che una funzione $f : I \rightarrow \mathbb{R}$ è **assolutamente continua** se $\forall \varepsilon > 0$ esiste un numero positivo δ_ε tale che per ogni sequenza di sotto-intervalli di I $([x_k, y_k])_{k=1}^n$ a due a due non sovrapponibili vale

$$\sum_{k=1}^n (y_k - x_k) \leq \delta_\varepsilon \quad \Rightarrow \quad \sum_{k=1}^n |f(x_k) - f(y_k)| \leq \varepsilon.$$

In tal caso scriveremo $f \in \text{AC}(I)$.

Osservazione 10. Se $f \in \text{AC}(I)$ allora è continua.

Lemma A.1. Sia $f : \mathbb{R} \rightarrow \mathbb{R}$ una funzione. Se f è lipschitziana allora f è assolutamente continua.

Dimostrazione. Sia C la migliore costante di Lipschitz associata alla funzione f . Fissato $\varepsilon > 0$ e fissata $([x_k, y_k])_{k=1}^n$ una sequenza di intervalli a due a due non sovrapponibili otteniamo

$$\sum_{k=1}^n |f(y_k) - f(x_k)| \leq \sum_{k=1}^n C (y_k - x_k).$$

Ponendo $\delta_\varepsilon := \frac{\varepsilon}{C}$ otteniamo la tesi.

□

Definizione A.3. Diremo che una funzione $f : \mathbb{R} \rightarrow \mathbb{R}$ è **localmente assolutamente continua** se la sua restrizione ad ogni intervallo compatto è assolutamente continua. Scriveremo, in questo caso, $f \in \text{AC}_{loc}(\mathbb{R})$.

Osservazione 11. Se $f \in AC(\mathbb{R})$ allora $f \in AC_{loc}(\mathbb{R})$.

Il seguente teorema, la cui dimostrazione si può trovare in [35, Teorema 7.20, p.148], mostra una proprietà importante delle funzioni assolutamente continue definite su un intervallo compatto.

Teorema A.2. Sia $f \in AC([a, b])$, allora f è differenziabile quasi ovunque su $[a, b]$. In particolare, $f' \in L^1([a, b])$ e vale

$$f(x) - f(a) = \int_a^x f'(t) dt, \quad \forall x \in [a, b].$$

Dal suddetto teorema si ottiene in modo immediato:

Corollario A.3. Sia $f \in AC(\mathbb{R})$, allora f è differenziabile quasi ovunque su \mathbb{R} con $f' \in L^1_{loc}(\mathbb{R})$.

Il Teorema A.4, che segue, fornisce un'importante caratterizzazione delle funzioni assolutamente continue non decrescenti definite su un intervallo compatto. La sua dimostrazione può essere trovata in [35, Teorema 7.18, p.146].

Teorema A.4. Sia $f : [a, b] \rightarrow \mathbb{R}$ una funzione continua e non decrescente. Allora i seguenti fatti sono tra loro equivalenti

(i) $f \in AC([a, b])$;

(ii) f è differenziabile quasi ovunque su $[a, b]$, $f' \in L^1([a, b])$ e

$$f(x) - f(a) = \int_a^x f'(s) ds \quad \forall x \in [a, b].$$

Il Teorema A.5, presentato sotto, prova l'assoluta continuità per una classe di funzioni definite mediante integrali e fornisce un'espressione della loro derivata quasi certa. La dimostrazione del teorema si può trovare in [34, Teorem 8.17, p.165]

Teorema A.5. Sia $f \in L^1(\mathbb{R})$. Allora la funzione definita da

$$g(x) = \int_{-\infty}^x f(s) ds$$

è una funzione ben definita e assolutamente continua. Inoltre, $g' = f$ quasi ovunque.

Andiamo ora a dimostrare che il prodotto di funzioni assolutamente continue definite su un intervallo compatto è ancora una funzione assolutamente continua. Mostriamo anche, attraverso un controesempio, che tale proprietà non vale per funzioni assolutamente continue definite su \mathbb{R} . Tale proprietà, però, continua a valere se ci restringiamo a funzioni localmente assolutamente continue. Mostriamo, anche, che la somma di funzioni assolutamente continue definite su un intervallo è ancora una funzione assolutamente continua.

Proposizione A.6. Siano $f, g \in AC([a, b])$. Allora $h := f \cdot g \in AC([a, b])$.

Dimostrazione. Grazie all'Osservazione 10, f e g sono funzioni continue. Dal teorema di Weirstrassesistono finite $\|f\|_\infty$ e $\|g\|_\infty$; poniamo $M := \max \{\|f\|_\infty, \|g\|_\infty\}$. Sia $\varepsilon > 0$ un numero fissato e $([x_k, y_k])_{k=1}^n$ una sequenza di sotto-intervalli di $[a, b]$ a due a due non sovrapponibili. Essendo f e g assolutamente continue, esistono $\delta_\varepsilon^1, \delta_\varepsilon^2 > 0$ tali che

$$\begin{aligned} \sum_{k=1}^n (y_k - x_k) \leq \delta_\varepsilon^1 &\Rightarrow \sum_{k=1}^n |f(x_k) - f(y_k)| \leq \frac{\varepsilon}{2M}, \\ \sum_{k=1}^n (y_k - x_k) \leq \delta_\varepsilon^2 &\Rightarrow \sum_{k=1}^n |g(x_k) - g(y_k)| \leq \frac{\varepsilon}{2M}. \end{aligned} \tag{A.1.1}$$

Posto $\delta_\varepsilon := \max \{\delta_\varepsilon^1, \delta_\varepsilon^2\}$, da (A.1.1) otteniamo

$$\sum_{k=1}^n (y_k - x_k) \leq \delta_\varepsilon \Rightarrow \begin{cases} \sum_{k=1}^n |f(x_k) - f(y_k)| \leq \frac{\varepsilon}{2M}, \\ \sum_{k=1}^n |g(x_k) - g(y_k)| \leq \frac{\varepsilon}{2M}. \end{cases} \tag{A.1.2}$$

Dunque

$$\begin{aligned} \sum_{k=1}^n |h(x_k) - h(y_k)| &= \sum_{k=1}^n |f(x_k)g(x_k) - f(y_k)g(y_k)| \\ &\leq \sum_{k=1}^n |f(x_k)g(x_k) - f(x_k)g(y_k) + f(x_k)g(y_k) - f(y_k)g(y_k)| \\ &= \sum_{k=1}^n |f(x_k)(g(x_k) - g(y_k)) + g(y_k)(f(x_k) - f(y_k))| \\ &\leq \sum_{k=1}^n |f(x_k)| |g(x_k) - g(y_k)| + \sum_{k=1}^n |g(y_k)| |f(x_k) - f(y_k)| \\ &\leq M \sum_{k=1}^n |g(x_k) - g(y_k)| + M \sum_{k=1}^n |f(x_k) - f(y_k)| \end{aligned} \tag{A.1.3}$$

Se $\sum_{k=1}^n (y_k - x_k) \leq \delta_\varepsilon$, combinando opportunamente (A.1.2) e (A.1.3) otteniamo

$$\sum_{k=1}^n |h(x_k) - h(y_k)| \leq M \left(\frac{\varepsilon}{2M} + \frac{\varepsilon}{2M} \right) = \varepsilon.$$

Per arbitrarietà di ε e di $([x_k, y_k])_{k=1}^n$ otteniamo che $h \in AC([a, b])$. □

Il risultato precedente non vale se consideriamo funzioni assolutamente continue definite su tutto \mathbb{R} . Infatti, vale il seguente risultato:

Lemma A.7. *Esistono $f, g \in AC(\mathbb{R})$ tali che $fg \notin AC(\mathbb{R})$.*

Dimostrazione. È banale verificare $f(x) := x$ è in $AC(\mathbb{R})$. Se proviamo che $g(x) := x^2$ non è in $AC(\mathbb{R})$ otteniamo il controesempio voluto. Supponiamo per assurdo che g sia assolutamente continua. Fissato $\varepsilon > 0$ sia $\delta_\varepsilon > 0$ come nella definizione di assoluta continuità e sia $n \in \mathbb{N}$ con $n \geq \frac{\varepsilon}{\delta}$. Allora abbiamo

$$\left| g(n) - g\left(n + \frac{\delta}{2}\right) \right| = \left| n^2 - \left(n + \frac{\delta}{2}\right)^2 \right| = \left| \frac{\delta^2}{4} + n\delta \right| > n\delta \geq \varepsilon.$$

Poiché $\left[n, n + \frac{\delta}{2} \right]$ ha ampiezza minore di δ otteniamo un assurdo. □

Corollario A.8. *Siano $f, g \in AC_{loc}(\mathbb{R})$ allora $h := f \cdot g \in AC_{loc}(\mathbb{R})$.*

Dimostrazione. Poiché $f, g \in AC_{loc}(\mathbb{R})$ allora per ogni coppia di numeri reali (a, b) con $a < b$ vale $f|_{[a,b]}, g|_{[a,b]} \in AC([a, b])$. Dalla Proposizione A.6 otteniamo che $f|_{[a,b]}g|_{[a,b]} \in AC([a, b])$. La tesi segue notando che $f|_{[a,b]}g|_{[a,b]} \equiv h|_{[a,b]}$. □

Proposizione A.9. *Sia $I \subseteq \mathbb{R}$ un intervallo. Se $f, g \in AC(I)$, allora $h := f + g \in AC(I)$.*

Dimostrazione. Sia $\varepsilon > 0$ un numero fissato e $([x_k, y_k])_{k=1}^n$ una sequenza di sotto-intervalli di I a due a due non contigui. Essendo f e g assolutamente continue, esistono $\delta_\varepsilon^1, \delta_\varepsilon^2 > 0$ tali per che

$$\begin{aligned} \sum_{k=1}^n y_k - x_k \leq \delta_\varepsilon^1 &\Rightarrow \sum_{k=1}^n |f(x_k) - f(y_k)| \leq \frac{\varepsilon}{2}, \\ \sum_{k=1}^n y_k - x_k \leq \delta_\varepsilon^2 &\Rightarrow \sum_{k=1}^n |g(x_k) - g(y_k)| \leq \frac{\varepsilon}{2}. \end{aligned} \tag{A.1.4}$$

Posto $\delta_\varepsilon = \max \{ \delta_\varepsilon^1, \delta_\varepsilon^2 \}$, da (A.1.4) otteniamo

$$\sum_{k=1}^n y_k - x_k \leq \delta_\varepsilon \Rightarrow \sum_{k=1}^n |f(x_k) - f(y_k)|, \sum_{k=1}^n |g(x_k) - g(y_k)| \leq \frac{\varepsilon}{2}. \tag{A.1.5}$$

Dunque

$$\begin{aligned} \sum_{k=1}^n |h(x_k) - h(y_k)| &= \sum_{k=1}^n |f(x_k) + f(y_k) - f(y_k) - f(y_k)| \leq \\ &\leq \sum_{k=1}^n |f(x_k) - f(y_k)| + \sum_{k=1}^n |g(x_k) - g(y_k)|. \end{aligned} \quad (\text{A.1.6})$$

Se $\sum_{k=1}^n y_k - x_k \leq \delta_\varepsilon$, combinando (A.1.5) e (A.1.6) otteniamo

$$\sum_{k=1}^n |h(x_k) - h(y_k)| \leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

Per arbitrarietà di ε e di $([x_k, y_k])_{k=1}^n$ otteniamo la tesi. □

A.2 | Funzioni a tratti

Definizione A.4. Scriveremo $f \in C_{tratti}^1(\mathbb{R})$ se $f : \mathbb{R} \rightarrow \mathbb{R}$ soddisfa le seguenti proprietà:

1. f è continua;
2. esistono $-\infty = x_0 < x_1 < \dots < x_n < x_{n+1} = +\infty$ per cui, per ogni $i \in \{0, \dots, n\}$, vale:

(a) $f \in C^1([x_i, x_{i+1}])$;

(b) esistono finiti $\lim_{x \rightarrow x_i^+} f'(x)$ e $\lim_{x \rightarrow x_{i+1}^-} f'(x)$ con le convenzioni $-\infty^+ = -\infty$
e $+\infty^- = +\infty$.

Osservazione 12. Se $f \in C_{tratti}^1(\mathbb{R})$ allora f' è definita a meno di un insieme numerabile: f è derivabile quasi ovunque.

A.3 | Funzioni lipschitziane

Definizione A.5. Diremo che una funzione $f : \mathbb{R} \rightarrow \mathbb{R}$ è K -lipschitziana se

$$|f(x) - f(y)| \leq K |x - y|.$$

Denoteremo con $Lip_K(\mathbb{R})$ lo spazio delle funzioni K -lipschitziane.

Osservazione 13. Sia $f \in Lip_K(\mathbb{R})$. Se f è derivabile allora $\|f'\|_\infty \leq K$.

B | Richiami di probabilità

B.1 | Probabilità generale

Definizione B.1. Con $\mathcal{B}(\mathbb{R})$ denotiamo la σ -algebra dei Boreliani di \mathbb{R} ovvero la più piccola σ -algebra che contiene gli aperti di \mathbb{R} .

Il seguente lemma mostra che i boreliani sono anche la più piccola σ -algebra a contenere le semirette chiuse a destra. La dimostrazione viene omessa essendo un classico risultato di teoria della misura.

Lemma B.1. Sia

$$\mathcal{S} = \{(-\infty, x] \mid x \in \mathbb{R}\} \cup \{\emptyset\}$$

Se $\sigma(\mathcal{S})$ è la più piccola σ -algebra che contiene \mathcal{S} allora vale $\sigma(\mathcal{S}) = \mathcal{B}(\mathbb{R})$

Sia X è una variabile aleatoria a valori reali. Denoteremo con \mathbb{P}_X la sua legge cioè la misura di probabilità su $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ definita da

$$\mathbb{P}_X(A) := \mathbb{P}(\{\omega \in \Omega \mid X(\omega) \in A\})$$

Siano X, Y due variabili aleatorie. Scriveremo $X \sim Y$ per indicare che X e Y hanno la stessa legge cioè $\mathbb{P}_X = \mathbb{P}_Y$

Definizione B.2. Siano X, Y due variabili aleatorie reali definite sullo stesso spazio di probabilità. Se X, Y, XY sono integrabili si definisce la **covarianza** di X e Y la quantità

$$\text{Cov}(X, Y) := \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])].$$

Se X è quadrato integrabile si definisce la **varianza** di X come

$$\text{Var}(X) := \text{Cov}(X, X)$$

Osservazione 14. Dalle definizioni con un semplice calcolo si prova che

$$\text{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y],$$

$$\text{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2.$$

Le prossime due proposizioni forniscono delle importanti proprietà sulla covarianza e varianza. Per le dimostrazioni si può ricorrere ad un qualsiasi testo di probabilità di base; si veda, ad esempio, rispettivamente [7, Proposizione 3.68, p. 144] e [7, Proposizione 3.69, p.145]

Proposizione B.2. *La covarianza è un operatore simmetrico e bilineare, cioè per ogni X, Y, Z integrabili con XZ, YZ integrabili e per ogni $\alpha, \beta \in \mathbb{R}$ vale*

$$\text{Cov}(X, Z) = \text{Cov}(Z, X),$$

$$\text{Cov}(\alpha X + \beta Y, Z) = \alpha \text{Cov}(X, Z) + \beta \text{Cov}(Y, Z).$$

In particolare, se Z è quasi certamente costante allora $\text{Cov}(X, Z) = 0$.

Proposizione B.3. *Sia X variabile aleatoria quadrato integrabile e siano $a, b \in \mathbb{R}$ allora*

$$\text{Var}(aX + b) = a^2 \text{Var}(X).$$

Se $(X_i)_{i=1}^n$ è una sequenza di v.a. a quadrato integrabile allora

$$\text{Var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \text{Var}(X_i) + \sum_{\substack{1 \leq i, j \leq n \\ i \neq j}} \text{Cov}(X_i, X_j)$$

Lemma B.4. *Sia \mathcal{I} un insieme finito di indici e $X_1, \dots, X_{|\mathcal{I}|}$ variabili aleatorie con $\text{Var}(X_i) = 1$ e siano $(a_i)_{i \in \mathcal{I}}$ una \mathcal{I} -sequenza di numeri reali. Allora*

$$\text{Var}\left(\sum_{i \in \mathcal{I}} a_i X_i\right) \leq (a |\mathcal{I}|)^2$$

dove $a = \max_{i \in \mathcal{I}} |a_i|$.

Dimostrazione. Dalle proprietà della varianza otteniamo

$$\begin{aligned} \text{Var}\left(\sum_{i \in \mathcal{I}} a_i X_i\right) &= \sum_{i \in \mathcal{I}} a_i^2 \text{Var}(X_i) + \sum_{\substack{i, j \in \mathcal{I} \\ i \neq j}} a_i a_j \text{Cov}(X_i, X_j) \\ &\leq \sum_{i \in \mathcal{I}} a_i^2 + \sum_{\substack{i, j \in \mathcal{I} \\ i \neq j}} |a_i| |a_j| \leq a^2 |\mathcal{I}| + a^2 |\mathcal{I}| (|\mathcal{I}| - 1) = a^2 |\mathcal{I}|^2. \end{aligned}$$

dove la penultima disuguaglianza segue da $|\text{Cov}(X, Y)| \leq \sqrt{\text{Var}(X) \text{Var}(Y)}$. □

Il seguente teorema, che segue dal teorema delle classi monotone e la cui dimostrazione si può trovare in [28, Corollario 6.1, p.36], fornisce un risultato molto importante per dimostrare quando due misure di probabilità coincidono.

Teorema B.5. Sia (Ω, \mathcal{A}) uno spazio di misura e siano P, Q due misure di probabilità definite su \mathcal{A} . Supponiamo che $P \equiv Q$ su una classe $\mathcal{C} \subseteq \mathcal{A}$ chiusa per intersezioni finite. Allora $P \equiv Q$ su $\sigma(\mathcal{C})$, dove $\sigma(\mathcal{C})$ è la più piccola σ -algebra che contiene \mathcal{C} .

B.2 | Convergenza in legge

Definizione B.3. Sia $(X_n)_{n \in \mathbb{N}}$ una successione di vettori aleatori a valori in \mathbb{R}^k . Diremo che la successione (X_n) **converge in legge** al vettore X e scriveremo $X_n \Rightarrow X$ se per ogni $A \subseteq \mathbb{R}^k$ con $\mathbb{P}(X \in \partial A) = 0$ vale

$$\lim_{n \rightarrow +\infty} \mathbb{P}(X_n \in A) = \mathbb{P}(X \in A).$$

Il teorema seguente mostra un'importante caratterizzazione della convergenza in legge. Questo fatto, insieme ad un'altra serie caratterizzazioni, prende il nome di Teorema di Portmanteau. Si veda [30, Teorema 13.16, p.254] per una dimostrazione.

Teorema B.6. Sia $(X_n)_{n \in \mathbb{N}}$ una successione di vettori aleatori a valori in \mathbb{R}^k e X un vettore aleatorio in \mathbb{R}^k . Allora

$$X_n \Rightarrow X \quad \Leftrightarrow \quad \mathbb{E}[h(X_n)] \rightarrow \mathbb{E}[h(X)] \quad \forall h \in \text{Lip}(\mathbb{R}) \text{ limitata.}$$

Per il seguente teorema si veda [30, Teorema 15.56, p.329]

Teorema B.7 (Teorema di Cramér-Wold). Siano $(Y_n)_{n \in \mathbb{N}}$ e Y vettori aleatori a valori in \mathbb{R}^d . Condizione necessaria e sufficiente per la convergenza di Y_n a Y in legge è che $\theta \cdot Y_n$ converge in legge a $\theta \cdot Y$ per ogni $\theta \in \mathbb{R}^d$.

B.3 | Valore atteso condizionato

Fissiamo $(\Omega, \mathcal{A}, \mathbb{P})$ spazio di probabilità

Definizione B.4. Sia X una variabile aleatoria con $\mathbb{E}[|X|] < \infty$ e $\mathcal{F} \subseteq \mathcal{A}$ una σ -algebra. L'**aspettazione condizionata** di X data \mathcal{F} è una qualsiasi variabile aleatoria Y tale che

- Y è \mathcal{F} -misurabile,
- per ogni $A \in \mathcal{F}$ vale

$$\int_A Y \, d\mathbb{P} = \int_A X \, d\mathbb{P}.$$

Una qualsiasi funzione Y che soddisfa le due proprietà si dice una versione dell'aspettazione condizionata e si indica con $\mathbb{E}[X | \mathcal{F}]$. In [43, Teorema 9.2, p. 84] viene provata l'esistenza e l'unicità quasi certa dell'aspettazione condizionata. Siano X, Y due variabili aleatorie con $\mathbb{E}[|X|] < \infty$, allora definiamo

$$\mathbb{E}[X | Y] := \mathbb{E}[X | \sigma(Y)]$$

dove $\sigma(Y)$ denota la σ -algebra generata da Y .

Similmente, se X, Y e Z sono tre variabili aleatorie con $\mathbb{E}[|X|] < \infty$, allora definiamo

$$\mathbb{E}[X | Y, Z] := \mathbb{E}[X | \sigma(Y, Z)]$$

Andiamo ad elencare alcune proprietà utili riguardanti l'aspettazione condizionata. Per la dimostrazione di questi risultati si vedano le proprietà (a), (b), (j) e (k) in [43, Sezione 9.7].

Proposizione B.8. *Sia X una variabile aleatoria con $\mathbb{E}[|X|] < \infty$ e $\mathcal{F} \subseteq \mathcal{A}$ una σ -algebra. Allora valgono i seguenti risultati*

(i)

$$\mathbb{E}[\mathbb{E}[X | \mathcal{F}]] = \mathbb{E}[X]. \quad (\text{B.3.1})$$

(ii) *Se X è \mathcal{F} -misurabile allora*

$$\mathbb{E}[X | \mathcal{F}] = X. \quad (\text{B.3.2})$$

(iii) *Se X è \mathcal{F} -misurabile e $\mathbb{E}[|Y|], \mathbb{E}[|XY|] < \infty$ allora*

$$\mathbb{E}[XY | \mathcal{F}] = X\mathbb{E}[Y | \mathcal{F}]. \quad (\text{B.3.3})$$

(iv) *Se X è indipendente da \mathcal{F} allora*

$$\mathbb{E}[X | \mathcal{F}] = \mathbb{E}[X]. \quad (\text{B.3.4})$$

Teorema B.9 (Legge dell'aspettazione totale). *Siano X, Y due variabili aleatorie definite sullo stesso spazio di probabilità $(\Omega, \mathcal{A}, \mathbb{P})$ per cui $\mathbb{E}[|X|] < \infty$. Se Y assume valori su un insieme discreto S allora*

$$\mathbb{E}[X] = \sum_{y \in S} \mathbb{E}[X | Y = y] \mathbb{P}(Y = y).$$

B.4 | Processi stocastici e vettori multivariati

Definizione B.5. Sia T un insieme non vuoto, $(\Omega, \mathcal{A}, \mathbb{P})$ uno spazio di probabilità e (E, \mathcal{E}) uno spazio misurabile. Un **processo stocastico** X definito su $(\Omega, \mathcal{A}, \mathbb{P})$ e a valori in (E, \mathcal{E}) con T come insieme degli indici è una funzione $X : \Omega \times T \rightarrow E$ tale che per ogni $t \in T$ la funzione

$$X_t : \Omega \rightarrow E \quad X_t(\omega) := X(\omega, t)$$

è misurabile. Nel seguito, denoteremo il processo come $(X_t)_{t \in T}$.

Definizione B.6. Un vettore aleatorio $Z = (Z_1, \dots, Z_k) : \Omega \rightarrow \mathbb{R}^k$ si dice **gaussiano multivariato** se per ogni $\alpha = (\alpha_1, \dots, \alpha_k) \in \mathbb{R}^k$ la variabile aleatoria $\sum_{i=1}^k \alpha_i Z_i$ è normale.

Per la dimostrazione del seguente risultato si veda [17, Proposizione 3.13, p.116].

Teorema B.10. Sia $X = (X_i)_{i \in \mathcal{I}}$ un vettore aleatorio e sia $\{\mathcal{J}, \mathcal{K}\}$ una partizione di \mathcal{I} tale che il vettore $(X_{\mathcal{J}}, X_{\mathcal{K}})$ è normale multivariato con media (μ_1, μ_2) e matrice di covarianza

$$\text{Cov}(X) = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12} & \Sigma_{22} \end{pmatrix}$$

allora

$$X_{\mathcal{J}} \sim_{X_{\mathcal{K}}=x_2} \mathcal{N}(\mu_1 + \Sigma_{12}\Sigma_{22}^-(x_2 - \mu_2), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^-\Sigma_{12}^T)$$

dove Σ_{22}^- è la pseudoinversa di Σ_{22} , Σ_{12}^T è la trasposta di Σ_{12} e $Y \sim_{Z=z} Y'$ denota che la legge condizionata di Y data $Z = z$ è uguale a quella di Y'

Definizione B.7. Un processo stocastico $(X_t)_{t \in T}$ si dice **processo gaussiano** se per ogni scelta finita t_1, \dots, t_k di indici in T il vettore $(X_{t_1}, \dots, X_{t_k})$ è un vettore aleatorio gaussiano multivariato.

C | Alcuni risultati sui grafi

Il collegamento tra numero di archi e gradi dei vertici è dato dal seguente classico lemma. La dimostrazione può essere reperita in un qualsiasi libro di teoria dei grafi. Si veda ad esempio [4, Teorema 1.1., p 7].

Lemma C.1 (Lemma delle strette di mano). *Sia $G = (V, E)$ un grafo. Allora*

$$\sum_{x \in V} d_G(x) = 2|E|.$$

Lemma C.2. *Sia $G = (V, E)$ un grafo. Se $e = \{x, y\} \in E$ allora il numero di archi a lui incidenti è $d_G(x) + d_G(y) - 1$.*

Dimostrazione. Sia l un arco incidente ad e con $l \neq e$. Allora

- $l \cap e = \{x\}$. In questo caso $l = \{x, h\}$ con $h \in \Gamma_G(x) \setminus \{y\}$. Il vertice h può essere scelto in $d_G(x) - 1$ modi distinti.
- $l \cap e = \{y\}$. In questo caso $l = \{y, h\}$ con $h \in \Gamma_G(y) \setminus \{x\}$. Il vertice h può essere scelto in $d_G(y) - 1$ modi distinti.

Poiché l'arco e è incidente a se stesso, il numero di archi incidenti ad e è

$$d_G(x) - 1 + d_G(y) - 1 + 1 = d_G(x) + d_G(y) - 1.$$

□

Lemma C.3. *Sia $G = (V, E)$ allora*

$$\sum_{\{i,j\} \in E} \{d_G(i) + d_G(j)\} = \sum_{i \in V} d_G(i)^2.$$

Dimostrazione. Poniamo

$$\Gamma_G^1(i) = \{(i, j) \in V \times V \mid j \in \Gamma_G(i)\}$$

e

$$\Gamma_G^2(i) = \{(j, i) \in V \times V \mid j \in \Gamma_G(i)\}.$$

Osserviamo che vale

$$\sum_{\{i,j\} \in E} \{d_G(i) + d_G(j)\} = \frac{1}{2} \sum_{\substack{(i,j) \in V \times V \\ \{i,j\} \in E}} \{d_G(i) + d_G(j)\}$$

infatti ogni arco da origine a due coppie. Notando che

$$(i, j) \in V \times V \wedge \{i, j\} \in E \Leftrightarrow i \in V \wedge j \in \Gamma_G^1(i) \Leftrightarrow j \in V \wedge i \in \Gamma_G^2(j)$$

otteniamo

$$\begin{aligned} \sum_{\substack{(i,j) \in V \times V \\ \{i,j\} \in E}} \{d_G(i) + d_G(j)\} &= \sum_{i \in V} \sum_{j \in \Gamma_G^1(i)} d_G(i) + \sum_{j \in V} \sum_{i \in \Gamma_G^2(j)} d_G(j) \\ &= \sum_{i \in V} \{d_G(i) |\Gamma_G^1(i)|\} + \sum_{j \in V} \{d_G(j) |\Gamma_G^2(j)|\}. \end{aligned}$$

Poiché $|\Gamma_G^1(s)| = |\Gamma_G^2(s)| = |\Gamma_G(s)|$ per ogni $s \in V$ otteniamo

$$\sum_{\substack{(i,j) \in V \times V \\ \{i,j\} \in E}} \{d_G(i) + d_G(j)\} = \sum_{i \in V} d_G(i)^2 + \sum_{j \in V} d_G(j)^2 = 2 \sum_{i \in V} d_G(i)^2$$

da cui la tesi. □

Il seguente risultato, la cui dimostrazione può essere trovata in [13, Teorema 1], fornisce una limitazione dall'alto e dal basso alla somma dei quadrati dei gradi in un grafo.

Teorema C.4. *Sia $G = (V, E)$ se $n := |V| \geq 2$ allora vale*

$$\frac{4|E|^2}{n} \leq \sum_{i \in V} d_G(i)^2 \leq |E| \left(\frac{2|E|}{n-1} + n \right)$$

Lemma C.5. *Sia $G = (V, E)$ un grafo con almeno un arco. Se $i \in V$ allora esiste un arco e tale che $d_G(i) \leq |A_e|$.*

Dimostrazione. Andiamo a distinguere due differenti casi.

- Se i è isolato allora $\Gamma_G(i) = \emptyset$ dunque $d_G(i) := |\Gamma_G(i)| = 0$. La tesi è banalmente vera.
- Se i non è isolato deve esistere $e \in E$ con $i \in e$. Tutti gli archi uscenti da i , che sono $d_G(i)$, stanno in A_e . Abbiamo, quindi, $d_G(i) \leq |A_e|$.

□

Lemma C.6. Sia $G = (V, E)$ un grafo con almeno un arco. Se $i \in V$ allora esiste un arco e tale che $|\Gamma_G^2(i)| \leq |B_e|$.

Dimostrazione. Andiamo a distinguere due differenti casi.

- Se i è isolato allora $\Gamma_G(i) = \emptyset$ e dunque anche $\Gamma_G^2(i)$ è vuoto. La tesi è banalmente vera.
- Se i non è isolato deve esistere $e \in E$ con $i \in e$. Sia $l \in \Gamma_G^2(i)$ allora, dalla definizione (4.2.4) di $\Gamma_G^2(i)$, si ha $l = \{j, x\}$ dove $x \in V$ e $j \in \Gamma_G(i)$. Distinguiamo due casi.
 - Se $x = i$ allora $i \in l \cap e \neq \emptyset$. In tal caso $l \in A_e \subseteq B_e$.
 - Se $x \neq i$ allora $p = \{j, i\} \in E$ infatti $j \in \Gamma_G(i)$. Allora $j \in p \cap l \neq \emptyset$ e quindi $l \in A_p$. Poiché $i \in p \cap e \neq \emptyset$ allora $p \in A_e$. Ricordando la definizione (4.2.2) di B_e otteniamo $l \in B_e$.

In entrambi i casi $l \in B_e$. Dunque, dall'arbitrarietà di $l \in \Gamma_G^2(i)$, otteniamo $\Gamma_G^2(i) \subseteq B_e$.

□

Osservazione 15. Rileggendo le dimostrazioni dei due lemmi possiamo dimostrare che per ogni $i \in V$ esiste un arco e tale che

$$d_G(i) \leq |A_e|, \quad |\Gamma_G^2(i)| \leq |B_e|.$$

Ovvero, le due tesi valgono con il medesimo arco.

Lemma C.7. Sia $G = (V, E)$ un grafo. Se $e \in E$ allora vale $|C_e| \leq |A_e| + 1$.

Dimostrazione. Sia $e = \{a, b\}$. Definiamo la mappa

$$\psi : A_e \setminus \{e\} \rightarrow C_e \setminus \{a, b\}$$

che manda l'arco l nell'unico vertice in l diverso da a o b . Banalmente tale mappa è suriettiva quindi vale $|C_e \setminus \{a, b\}| \leq |A_e \setminus \{e\}|$. La tesi segue osservando che $e \in A_e$ e $a, b \in C_e$ infatti $a, b \in e$ con $e \in A_e$.

□

C.1 | Alcuni algoritmi sui grafi

Nelle prossime sezioni andremo a presentare gli algoritmi utilizzati per la realizzazione della Figura 4.2. Per farlo siamo partiti da un grafo pesato completo $G = (V, E, w)$ dove V è l'insieme delle osservazioni y_1, \dots, y_n e la funzione peso è data dalla funzione di dissimilarità.

C.1.1 Algoritmo di Kruskal

Definizione C.1. Un **grafo pesato** è una terna $G = (V, E, w)$ dove (V, E) è un grafo e $w : \mathbb{E} \rightarrow \mathbb{R}$ è una funzione detta **funzione peso**.

Dato un grafo pesato $G = (V, E, w)$ definiamo il suo peso (o **costo**) come

$$\text{cost}(G) = \sum_{e \in E} w(e).$$

Dato un grafo pesato $G = (V, E, w)$ connesso per costruire un suo albero di copertura di costo minimo si può utilizzare l'algoritmo di Kruskal di cui forniamo uno pseudocodice. Il Teorema C.8 garantirà che l'algoritmo presentato fornisce realmente un albero di copertura di costo minimo.

Algoritmo 1 Kruskal

Input: $G = (V, E, w)$ grafo connesso e pesato

Output: T albero di copertura di costo minimo

ordina l'insieme E in base al costo in modo non decrescente

$T \leftarrow (V, \emptyset)$

for $\{x, y\} \in E$ secondo l'ordine **do**

if x e y non sono connessi in T **then**

 aggiungi al grafo T l'arco $\{x, y\}$

return T

Teorema C.8. L'Algoritmo 1 calcola correttamente un minimo albero ricoprente. Questo teorema corrisponde a [14, Teorema 12.2, p.305].

C.1.2 Accoppiamento di costo minimo

Definizione C.2. Dato un grafo $G = (V, E)$ un sottoinsieme $S \subseteq E$ si dice **accoppiamento** (in inglese *matching*) se S non contiene archi a due a due adiacenti. In altre parole $S \subseteq E$ è un accoppiamento se $\forall e \neq l \in S$ vale $e \cap l = \emptyset$.

Definizione C.3. Dato un grafo $G = (V, E)$ un accoppiamento S si dice **perfetto** se ogni vertice appartiene ad esattamente un arco di S .

Per produrre il grafo di minimo accoppiamento della Sezione 4.3 abbiamo implementato in MATLAB l'algoritmo presente in [24] che permette di ottenere un accoppiamento perfetto di costo massimo. Per ottenere quello di costo minimo abbiamo applicato l'algoritmo dando in input un grafo con gli stessi vertici e archi di quello originale ma con pesi opposti.

Bibliografia

- [1] Andrew D. Barbour and Louis H. Y. Chen, editors. *An introduction to Stein's method*, volume 4 of *Lecture Notes Series. Institute for Mathematical Sciences. National University of Singapore*. Singapore University Press, Singapore; World Scientific Publishing Co. Pte. Ltd., Hackensack, NJ, 2005. Lectures from the Meeting on Stein's Method and Applications: a Program in Honor of Charles Stein held at the National University of Singapore, Singapore, July 28–August 31, 2003.
- [2] John J. Benedetto and Wojciech Czaja. *Integration and modern analysis*. Birkhäuser Advanced Texts: Basler Lehrbücher. [Birkhäuser Advanced Texts: Basel Textbooks]. Birkhäuser Boston, Ltd., Boston, MA, 2009.
- [3] Andrew C. Berry. The accuracy of the Gaussian approximation to the sum of independent variates. *Trans. Amer. Math. Soc.*, 49:122–136, 1941.
- [4] John A. Bondy and Uppaluri S. R. Murty. *Graph theory*, volume 244 of *Graduate Texts in Mathematics*. Springer, New York, 2008.
- [5] Marcel Bosc, Fabrice Heitz, Jean-Paul Armspach, Izzie Namer, Daniel Gounot, and Lucien Rumbach. Automatic change detection in multimodal serial mri: application to multiple sclerosis lesion evolution. *NeuroImage*, 20(2):643–656, 2003.
- [6] Lorenzo Bruzzone and Diego Fernandez-Prieto. An adaptive semiparametric and context-based approach to unsupervised change detection in multi-temporal remote-sensing images. *IEEE Transactions on image processing*, 11(4):452–466, 2002.
- [7] Francesco Caravenna and Paolo Dai Pra. *Probabilità: Un'introduzione attraverso modelli e applicazioni*, volume 67. Springer Science & Business Media, 2013.
- [8] Hao Chen and Nancy Zhang. Graph-based change-point detection. *Ann. Statist.*, 43(1):139–176, 2015.

- [9] Louis H. Y. Chen, Larry Goldstein, and Qi-Man Shao. *Normal approximation by Stein's method*. Probability and its Applications (New York). Springer, Heidelberg, 2011.
- [10] Md Foezur Rahman Chowdhury, S-A Selouani, and D O'Shaughnessy. Bayesian on-line spectral change point detection: a soft computing approach for on-line asr. *International Journal of Speech Technology*, 15:5–23, 2012.
- [11] George W. Cobb. The problem of the Nile: conditional solution to a changepoint problem. *Biometrika*, 65(2):243–251, 1978.
- [12] Robert T. Collins, Alan J Lipton, and Takeo Kanade. Introduction to the special section on video surveillance. *IEEE Transactions on pattern analysis and machine intelligence*, 22(8):745–746, 2000.
- [13] D. de Caen. An upper bound on the sum of squares of degrees in a graph. *Discrete Math.*, 185(1-3):245–248, 1998.
- [14] Camil Demetrescu, Irene Finocchi, and Giuseppe F Italiano. *Algoritmi e strutture dati*. McGraw-Hill, 2004.
- [15] Jean-François Ducre-Robitaille, Lucie A Vincent, and Gilles Boulet. Comparison of techniques for detection of discontinuities in temperature series. *International Journal of Climatology: A Journal of the Royal Meteorological Society*, 23(9):1087–1101, 2003.
- [16] Rick Durrett. *Probability—theory and examples*, volume 49 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge, fifth edition, 2019.
- [17] Morris L. Eaton. *Multivariate statistics*, volume 53 of *Institute of Mathematical Statistics Lecture Notes—Monograph Series*. Institute of Mathematical Statistics, Beachwood, OH, 2007. A vector space approach, Reprint of the 1983 original [MR0716321].
- [18] Duane R Edgington, Karen A Salamy, Michael Risi, RE Sherlock, Dirk Walther, and Christof Koch. Automated event detection in underwater video. In *Oceans 2003. Celebrating the Past... Teaming Toward the Future (IEEE Cat. No. 03CH37492)*, volume 5, pages P2749–P2753. IEEE, 2003.
- [19] Carl-Gustav Esseen. On the Liapounoff limit of error in the theory of probability. *Ark. Mat. Astr. Fys.*, 28A,(9):19, 1942.
- [20] Lawrence C. Evans and Ronald F. Gariepy. *Measure theory and fine properties of functions*. Studies in Advanced Mathematics. CRC Press, Boca Raton, FL, 1992.

- [21] Chiung-Yao Fang, Sei-Wang Chen, and Chiou-Shann Fuh. Automatic change detection of driving environments in a vision-based driver assistance system. *IEEE Transactions on Neural Networks*, 14(3):646–657, 2003.
- [22] William Feller. *An introduction to probability theory and its applications. Vol. II*. John Wiley & Sons, Inc., New York-London-Sydney, 1966.
- [23] Hans Fischer. *A history of the central limit theorem*. Sources and Studies in the History of Mathematics and Physical Sciences. Springer, New York, 2011. From classical to modern probability theory.
- [24] Zvi Galil. Efficient algorithms for finding maximum matching in graphs. *Comput. Surveys*, 18(1):23–38, 1986.
- [25] Larry Goldstein and Gesine Reinert. Stein’s method and the zero bias transformation with application to simple random sampling. *Ann. Appl. Probab.*, 7(4):935–952, 1997.
- [26] Larry Goldstein and Yosef Rinott. Multivariate normal approximations by Stein’s method and size bias couplings. *J. Appl. Probab.*, 33(1):1–17, 1996.
- [27] Frank Harary. *Graph theory*. Addison-Wesley Publishing Co., Reading, Mass.-Menlo Park, Calif.-London, 1969.
- [28] Jean Jacod and Philip Protter. *Probability essentials*. Universitext. Springer-Verlag, Berlin, second edition, 2003.
- [29] Eamonn Keogh and Jessica Lin. Clustering of time-series subsequences is meaningless: implications for previous and future research. *Knowledge and information systems*, 8:154–177, 2005.
- [30] Achim Klenke. *Probability theory*. Universitext. Springer, London, second edition, 2014. A comprehensive course.
- [31] Rakesh Malladi, Giridhar P Kalamangalam, and Behnaam Aazhang. Online bayesian change point detection algorithms for segmentation of epileptic activity. In *2013 Asilomar conference on signals, systems and computers*, pages 1833–1837. IEEE, 2013.
- [32] Jaxk Reeves, Jien Chen, Xiaolan L Wang, Robert Lund, and Qi Qi Lu. A review and comparison of changepoint detection techniques for climate data. *Journal of applied meteorology and climatology*, 46(6):900–915, 2007.
- [33] Nathan Ross. Fundamentals of Stein’s method. *Probab. Surv.*, 8:210–293, 2011.

- [34] Walter Rudin. *Real and Complex Analysis*. McGraw-Hill series in higher mathematics. McGraw-Hill, 1966.
- [35] Walter Rudin. *Real and complex analysis*. McGraw-Hill Book Co., New York, third edition, 1987.
- [36] David Rybach, Christian Gollan, Ralf Schluter, and Hermann Ney. Audio segmentation for speech recognition using segment features. In *2009 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 4197–4200. IEEE, 2009.
- [37] David Siegmund and Benjamin Yakir. *The statistics of gene mapping*. Statistics for Biology and Health. Springer, New York, 2007.
- [38] M Staudacher, S Telser, A Amann, H Hinterhuber, and M Ritsch-Martel. A new method for change-point detection developed for on-line analysis of the heart beat variability during sleep. *Physica A: Statistical Mechanics and its Applications*, 349(3-4):582–596, 2005.
- [39] Charles Stein. A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability (Univ. California, Berkeley, Calif., 1970/1971), Vol. II: Probability theory*, pages 583–602. Univ. California Press, Berkeley, CA, 1972.
- [40] Charles Stein. *Approximate computation of expectations*, volume 7 of *Institute of Mathematical Statistics Lecture Notes—Monograph Series*. Institute of Mathematical Statistics, Hayward, CA, 1986.
- [41] Dang-Hoan Tran. Automated change detection and reactive clustering in multivariate streaming data. In *2019 IEEE-RIVF International Conference on Computing and Communication Technologies (RIVF)*, pages 1–6. IEEE, 2019.
- [42] Cédric Villani. *Optimal transport*, volume 338 of *Grundlehren der mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 2009. Old and new.
- [43] David Williams. *Probability with martingales*. Cambridge Mathematical Textbooks. Cambridge University Press, Cambridge, 1991.
- [44] Ping Yang, Guy Dumont, and John Mark Ansermino. Adaptive change detection in heart rate trend monitoring in anesthetized children. *IEEE transactions on biomedical engineering*, 53(11):2211–2219, 2006.