

Metodo di Stein e applicazioni al problema del change point detection

Discussione Tesi di laurea Magistrale



SAPIENZA
UNIVERSITÀ DI ROMA

Simmaco Di Lillo

La Sapienza - Università di Roma

20 settembre 2023

Struttura tesi

- 1 Il metodo di Stein per l'approssimazione gaussiana
- 2 Change-point detection

Il Teorema del Limite Centrale (TLC)

Teorema del Limite Centrale

Sia $(X_n)_{n \in \mathbb{N}}$ una sequenza di variabili aleatorie i.i.d con $\mathbb{E}[X_1^2] < \infty$. 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 $\sigma^2 = \text{Var}(X_1)$.

Versiona quantitativa del TLC

Teorema di Berry-Esseen per v.a. 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 di

$\frac{S_n}{\sigma\sqrt{n}} = \frac{\sum_{i=1}^n X_i}{\sigma\sqrt{n}}$ e Φ quella della normale standard allora vale

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

Metodo di Stein

Il metodo di Stein per l'approssimazione gaussiana permette di stimare la distanza tra una variabile aleatoria S e $\mathcal{N}(0, 1)$ data da

$$d_{\mathcal{H}}(S, Z) = \sup_{h \in \mathcal{H}} |\mathbb{E}[h(S)] - \mathbb{E}[h(Z)]|$$

con \mathcal{H} una classe separante di funzioni boreliane e $Z \sim \mathcal{N}(0, 1)$.

Separante

\forall v.a. W , $\{\mathbb{E}[h(W)] : h \in \mathcal{H}\}$ determina la legge di W .

Alcune classi \mathcal{H}

- La classe della **variazione totale**

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

- La classe di **Kolmogorov**

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

- La classe di **Wasserstein**

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

Il lemma di Stein

Lemma di Stein

Data una variabile aleatoria Z , $Z \sim \mathcal{N}(0, 1)$ se e solo se

$$\mathbb{E}[f'(Z)] = \mathbb{E}[Zf(Z)]$$

per ogni funzione $f \in AC_{loc}(\mathbb{R})$.

L'equazione di Stein

Sia $Z \sim \mathcal{N}(0, 1)$ e $h : \mathbb{R} \rightarrow \mathbb{R}$ funzione di Borel con $\mathbb{E}[|h(Z)|] < \infty$.
L'equazione di Stein associata ad h è data da

$$f'(x) - xf(x) = h(x) - \mathbb{E}[h(Z)]. \quad (1)$$

Proposizione: Esistenza e Unicità

$\exists!$ soluzione f_h di (1) tale che $\lim_{x \rightarrow +\infty} f_h(x)e^{-\frac{x^2}{2}} = 0$

L'equazione di Stein

Ricordiamo che

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

Corollario

Se S è una v.a. allora

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

Il bound uniforme per la distanza di Kolmogorov

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

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

Per la classe di Kolmogorov, $\mathcal{H}_K = \{h_z : z \in \mathbb{R}\}$, otteniamo

$$d_{\mathcal{H}_K}(S, Z) \leq \sup_{\substack{f: \mathbb{R} \rightarrow \mathbb{R} \text{ q.c. derivabile} \\ \|f\|_\infty \leq \frac{\sqrt{2\pi}}{4} \\ \|f'\|_\infty \leq 1}} |\mathbb{E}[f'(S) - Sf(S)]|$$

Limitare $\mathbb{E}[f'(S) - Sf(S)]$

In letteratura esistono varie tecniche che usando le proprietà di f e di S limitano $\mathbb{E}[f'(S) - Sf(S)]$. Nella tesi sono state studiate le seguenti tecniche:

- leave one out (con applicazioni al change-point detection),
- exchangeable pairs,
- trasformata zero-bias,
- trasformata size-bias

Change-point detection: alcune applicazioni

Il change-point detection trova applicazioni in svariati ambiti:

- Monitoraggio delle condizioni mediche
- Rilevamento del cambiamento climatico
- Riconoscimento vocale
- Analisi delle immagini
- Analisi delle attività umane
- ...

Change-point detection: formulazione statistica

Sia $(y_i)_{i=1}^n$ una sequenza di osservazioni. 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:

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

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

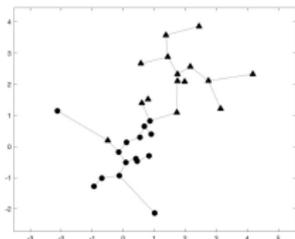
Grafi associati alle quantità osservate $(y_i)_{i=1}^n$

Sia $V = \{y_i \mid i = 1, \dots, n\} \subseteq \mathcal{X}$ e sia $d : \mathcal{X}^2 \rightarrow \mathbb{R}$ una funzione di dissimilarità. Possiamo associare a V alcuni grafi aventi V come insieme dei vertici

- **Esempio 1:** l'albero di copertura di dissimilarità minima

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

dove \mathcal{T} è l'insieme degli alberi di copertura di V .



- cerchi estratti da $\mathcal{N}((0,0), I_2)$
- triangoli estratti da $\mathcal{N}((2,2), I_2)$

Grafi associati alle quantità osservate $(y_i)_{i=1}^n$

Sia $V = \{y_i \mid i = 1, \dots, n\}$ e sia $d : V^2 \rightarrow \mathbb{R}$ una funzione di dissimilarità. Possiamo associare a V alcuni grafi aventi V come insieme dei vertici

- **Esempio 1:** l'albero di copertura di dissimilarità minima
- **Esempio 2:** il grafo di minimo accoppiamento
- **Esempio 3:** il grafo del vicino più vicino

Due contatori sul grafo

Supponiamo di aver osservato i valori (y_1, \dots, y_n) . Se $\pi \in \mathcal{S}_n$ allora, sotto \mathcal{H}_0 le stringhe (y_1, \dots, y_n) e $(y_{\pi(1)}, \dots, y_{\pi(n)})$ hanno la stessa probabilità di essere estratte.

Fissata la probabilità uniforme su \mathcal{S}_n definiamo su \mathcal{S}_n la v.a. $R_G(t)$ dove $R_G(t)(\pi)$ conta il numero di lati $\{i, j\}$ per cui $\pi(i) \leq t < \pi(j)$ o $\pi(j) \leq t < \pi(i)$.

Definiamo, inoltre,

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

Le due variabili e il change-point

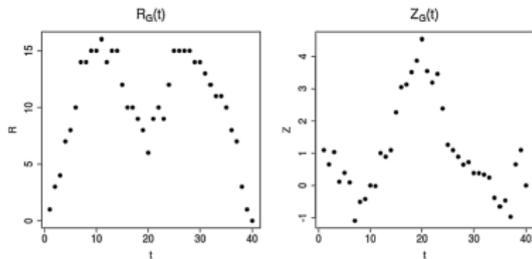


Grafico di $R_G(\cdot)$ e $Z_G(\cdot)$ su 40 punti in \mathbb{R}^2 .
20 campionati da $\mathcal{N}(0, I_2)$ e 20 da
 $\mathcal{N}((2, 2), I_2)$.

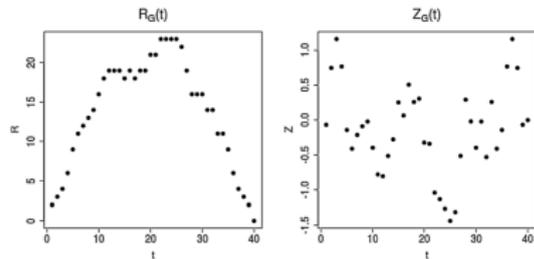


Grafico di $R_G(\cdot)$ e $Z_G(\cdot)$ su 40 punti in \mathbb{R}^2
campionati da $\mathcal{N}(0, I_2)$.

Il test statistico di Chen-Zhang (2015)

Per ottenere un test con livello di significatività α , rigettiamo l'ipotesi nulla se

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

dove b è scelto in modo che

$$\mathbb{P} \left(\max_{n_0 \leq t \leq n_1} Z_G(t) > b \right) \leq \alpha$$

Chen-Zhang danno un'espressione approssimata per $b = b(\alpha, n)$.

Il teorema di approssimazione

Teorema [Chen-Zhang (2015)]

Sia $G = ([n], E)$ un grafo parametrizzato da $n \in \mathbb{N}$ (quindi $G = G(n)$ e $E = E(n)$). Sotto opportune condizioni sul grafo G 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 di un processo gaussiano centrato $(Z_G^*(u))_{u \in (0,1)}$ con funzione di covarianza calcolabile.

Alcuni commenti al teorema

- Il teorema si basa sull'applicazione della tecnica leave one out per il metodo di Stein.
- La dimostrazione fa un'assunzione (ansatz) non banale e implicito sulle varianze di opportune v.a. Nella tesi abbiamo dimostrato la validità di tale ansatz in diversi casi.

Bibliografia essenziale

- [1] Hao Chen and Nancy Zhang. Graph-based change-point detection. *Ann. Statist.*, 43(1):139–176, 2015.
- [2] 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.
- [3] Barry James, Kang Ling James, and David Siegmund. Tests for a change-point. *Biometrika*, 74(1):71–83, 1987.
- [4] Nathan Ross. Fundamentals of Stein's method. *Probab. Surv.*, 8:210–293, 2011.
- [5] 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.

Leave one out - Una stima "tipo" Berry-Essen

Teorema

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

Exchangeable pairs

Definizione

(S, S') formano un'exchangeable pair se se $(S, S') \sim (S', S)$

Definizione

(S, S') formano una λ -Stein pair se se sono un'exchangeable pairs e

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

Zero-bias

Definizione

Data S con $\mathbb{E}[S] = 0$ e $\text{Var}(S) = \sigma^2$ diremo che una variabile aleatoria S^* ha la distribuzione **zero-bias** associata ad S se per ogni $f \in \text{AC}_{loc}(\mathbb{R})$ con $\mathbb{E}[|Sf(S)|], \mathbb{E}[|f'(S^*)|] < \infty$ vale

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

Zero-bias per variabili di Bernoulli

Proposizione

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

Size-bias

Definizione

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

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

Size-bias per variabili di Poisson

Proposizione

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

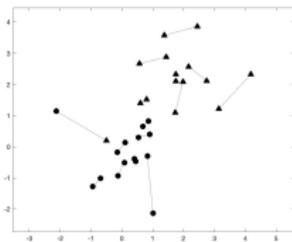
Il grafo di minimo accoppiamento

Il grafo di minimo accoppiamento: (V, \bar{P}) dove

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

dove

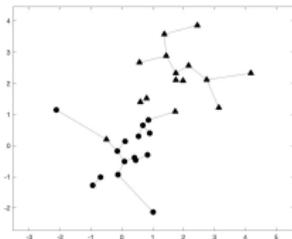
$$\mathcal{P} = \left\{ P \subseteq \binom{V}{2} \mid P \text{ partizione di } V \right\}.$$



Il grafo del vicino più vicino

Il grafo del vicino più vicino: (V, E) dove $E = \{\{x, n(x)\} \mid x \in V\}$ con

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



L'ansatz

Ansatz

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

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

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

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