Serie Temporali

April 25, 2021
Edoardo Teesco

1 Definizione della procedura

Ho un dataset di eventi $X = \{x_i\}_{i=1,\dots,n}$ dove ogni x_i è l'istante temporale in cui si verifica l'evento.

1.1 Gli eventi hanno statistiche in linea col modello nullo?

Condizione **necessaria** affinché gli eventi siano **indipendenti** è che essi **seguano un Processo di Poisson**. *Devo quindi verificare se le Δx_i si configurano secondo una distribuzione esponenziale.*

1.1.1 Anderson-Darling test

Ho scelto come stimatore l'Anderson-Darling test. Assunta l'ipotesi che la mia distribuzione X segua un comportamento esponenziale, selezionata una confidenza a piacere ($c = 1\%$), tale stimatore ti restituisce un valore a e un valore critico $b(c, X)$, tali per cui, se $a < b$, la distribuzione testata si comporta come l'ipotesi. *Il punto di forza di questo stimatore è che non ha bisogno di essere calibrato, dal momento che restituisce già un valore critico sulla base della tua distribuzione!*

1.2 Cluster o repulsione?

I miei eventi possono tendere ad un cluster oppure a repellersi. Per quantificare l'interazione utilizzo lo Hopkin Statistic.

1.2.1 Hopkis-Statistic

Lo Hopkins Statistic è un modo per misurare la tendenza di un cluster di un dataset, dove in tale caso restituisce un valore vicino a 1.

2 Calibrazione degli stimatori

2.1 Quanto mi posso fidare della stima di indipendenza, al variare del numero di eventi N?

In questo momento so discernere un comportamento in linea con il modello nullo con una confidenza del 1%. Ma quanto posso fidarmi del mio risultato? dove è chiaro che tale risultato dipenda dal numero di eventi N.

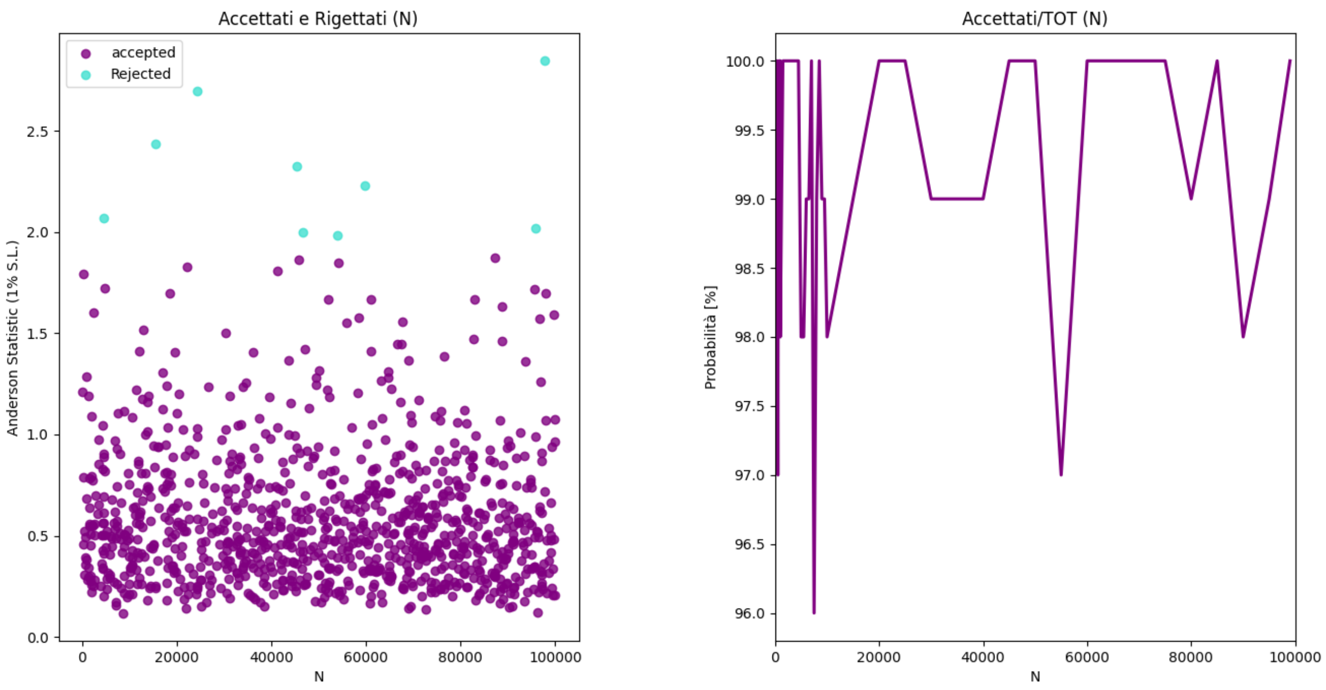

Figure 1: **Anderson Statistic**. Applico lo stimatore ad una distribuzione esponenziale (clusterizzata) e verifico quante volte mi restituisce un valore di verità al variare del numero di eventi N per una C.L pari a 1%. Praticamente sempre!

2.2 Calibrazione dello Hopkins

2.2.1 Verifico la consistenza dello Hopkins al variare di N

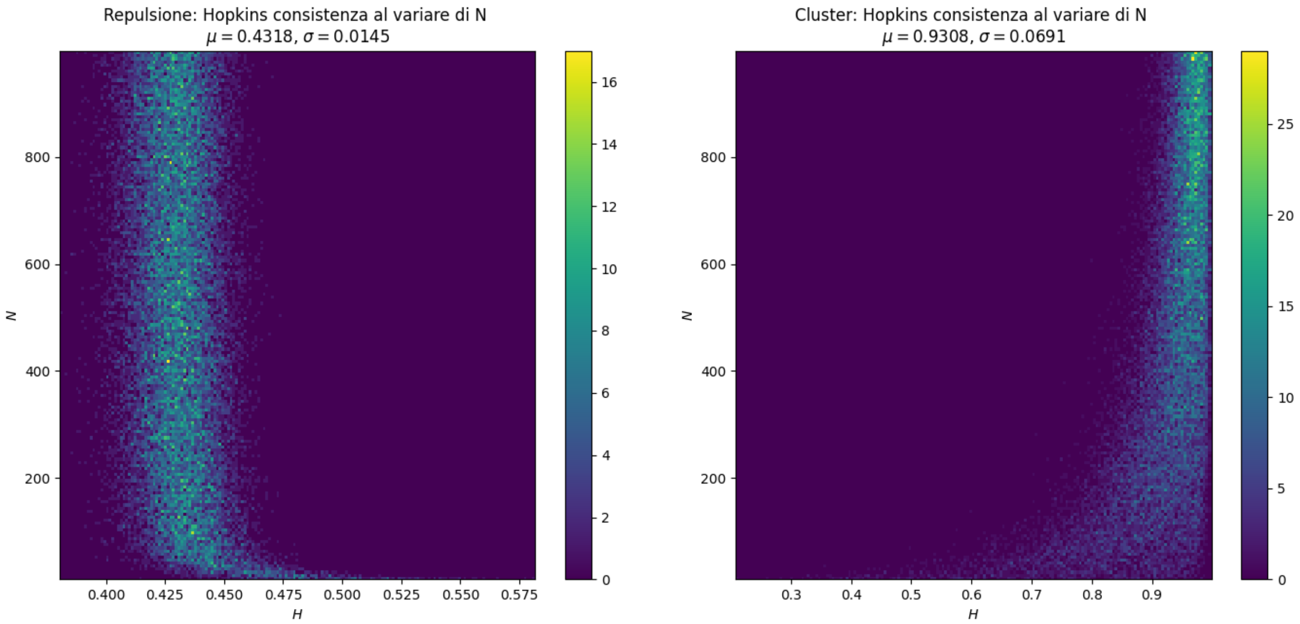

Figure 2: **Hopkins Statistic al variare di N**. Osservo che al crescere di N lo stimatore è consistente, fuorché per bassi valori

2.2.2 Verifico la consistenza dello Hopkins al variare di σ

Se gli eventi non sono uniformemente distribuiti allora essi potrebbero formare clusters oppure repellersi. Costruisco un modello in cui al variare di σ ho un diverso tipo di distribuzione.

- Per $\sigma \to 0$ ho punti equamente distanziati. Per $\sigma \to \infty$ ho punti distribuiti uniformemente.
- Per $\sigma \to 0$ ho un cluster. Per $\sigma \to \infty$ ho punti distribuiti uniformemente.

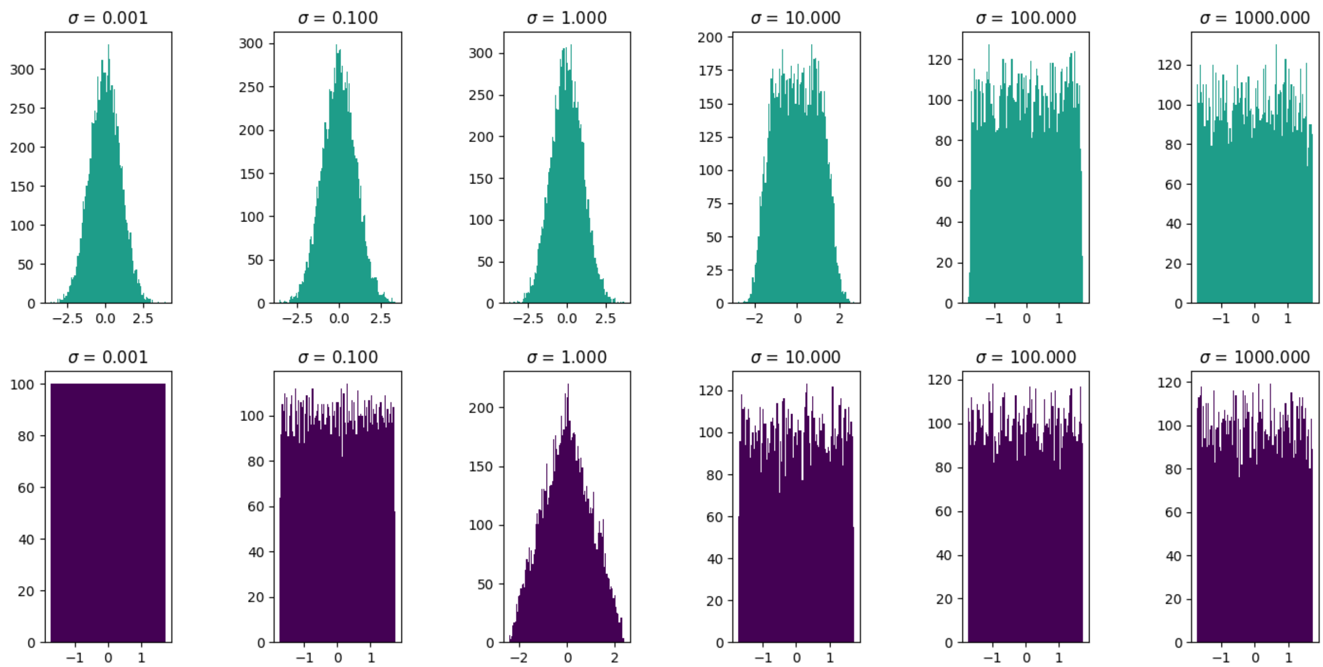

Figure 3: **Modello di interazione**. Al variare di σ il comportamento degli eventi passa da cluster/eventi equispaziati a uniformemente distribuiti.

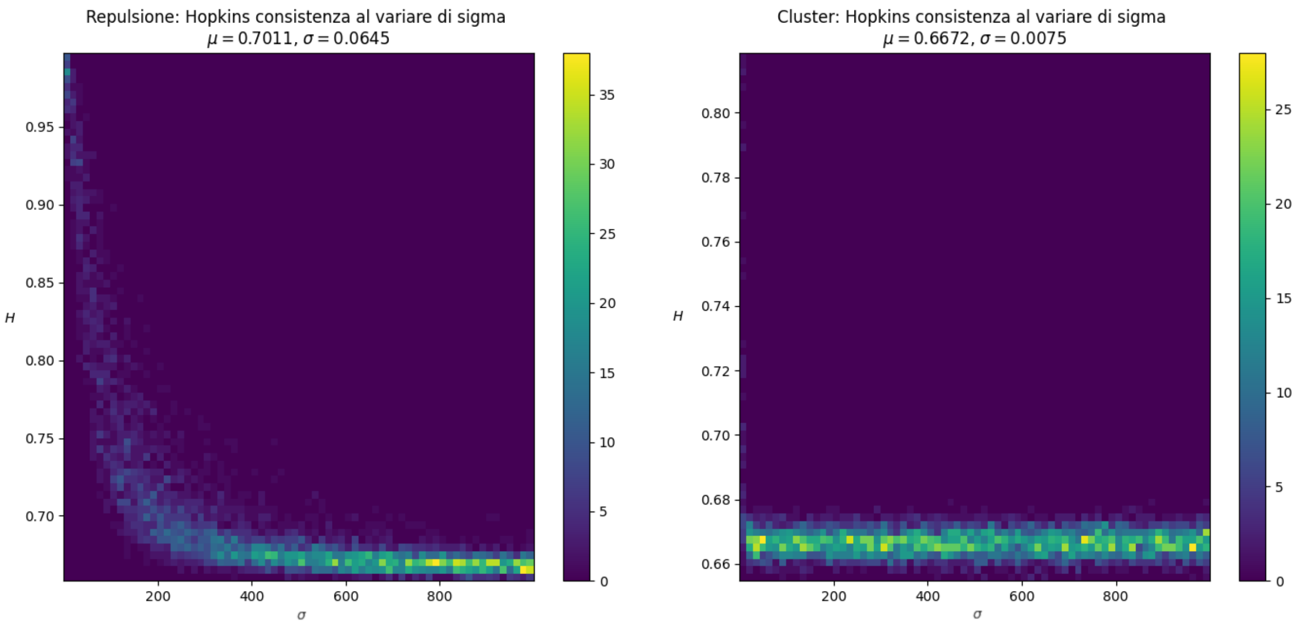

Figure 4: **Hopkins Statistic al variare di σ.**

3 Applicazione della procedura ad eventi reali

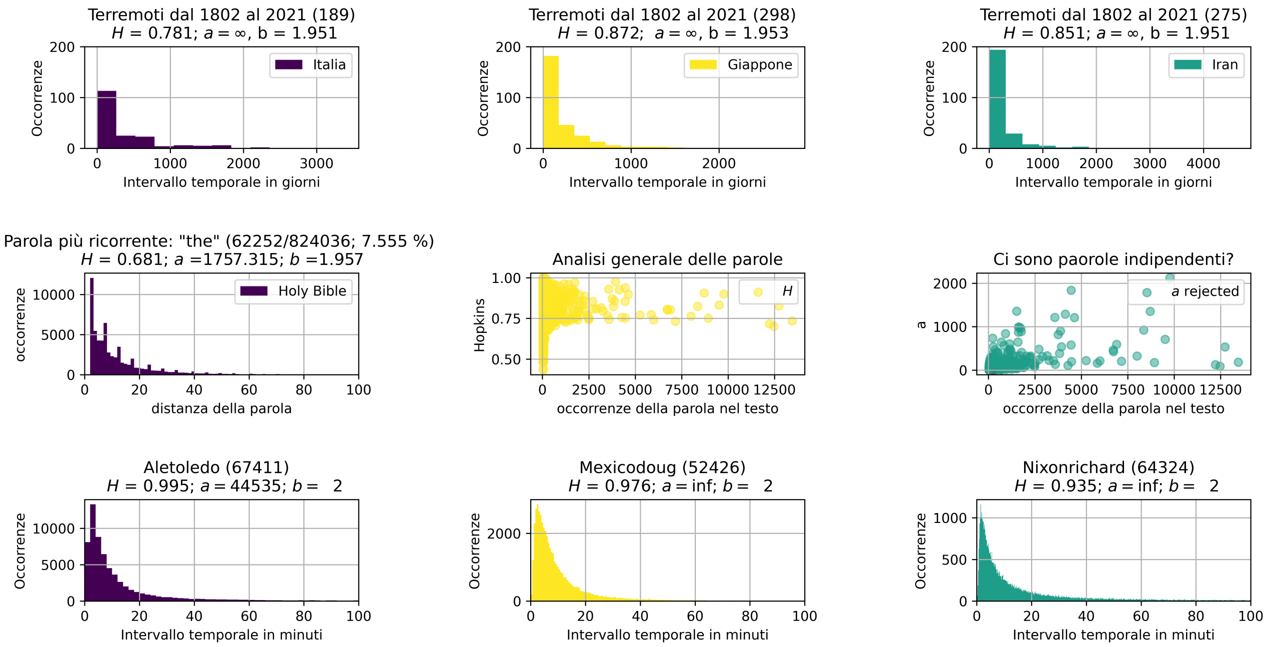

Figure 5: **Analisi di dataset reali**. Sono stati analizzati Terremoti, frequenza delle parole in un testo, interazioni di utenti in un social network.

3.1 Osservazioni

- **Terremoti**. Le analisi hanno rivelato che i dataset non sono in linea con il modello nullo. Dall'H rilevato, se confrontato per il numero di occorrenze in figura 2, gli eventi sono in una zona di clusterizzazione. Posso interpretare il fenomeno per le scosse di riassestamento.
- **Parole**. Le analisi hanno rivelato che i dataset non sono in linea con il modello nullo. (Osservando la parole *the* si nota anche infatti un pattern ripetitivo di frequenza). L'analisi generale invece rivela sia una tendenza a formare clusters, piuttosto che a repellersi.
- **Reddit**. Le analisi hanno rivelato che i dataset non sono in linea con il modello nullo. In suggerimento al valore di H, gli eventi tendono a clusterizzarsi.