

Implementazione di un algoritmo per analizzare dei segnali EEG con Matlab

Manuel Delucchi, Jordan Germano, Dante Baroncini

1 Introduzione EEG

1.1 Che cosa è l'elettroencefalogramma (EEG) e a che cosa serve?

L'elettroencefalografia (**EEG**) è una **metodica neurofisiologica** che permette la registrazione dell'attività elettrica generata dalla corteccia cerebrale, che a sua volta è la somma dell'attività elettrica di ogni singolo neurone. Generalmente si utilizzano **elettrodi di superficie** che vengono poggiati sul cuoio capelluto appositamente preparato. Siccome il voltaggio di questa attività elettrica è molto piccolo, il segnale deve essere **amplificato** un milione di volte per essere misurato e registrato, attraverso l'ausilio di una appropriata strumentazione, e visualizzato sotto forma di traccia (**elettroencefalogramma**).



Figure 1: "Cuffia" per EEG. E' possibile vedere gli elettrodi di superficie.

L'EEG rappresenta un valido strumento d'indagine nelle patologie in grado di modificare ed alterare l'attività elettrica cerebrale, quali malattie neurodegenerative (epilessia, parkinson, alzheimer), disturbi del sonno, traumi cranici. Viene anche usato per osservare le risposte corticali a stimoli sensoriali.

1.2 Onde EEG

La lettura dell'EEG si basa sull'osservazione dalle variazioni di potenziali registrate dagli elettrodi, chiamate **Onde**. Queste si classificano in base alla frequenza dominante in:

1. **Onde alfa**: 8-13 Hz, si osservano quando il soggetto è in una situazione di riposo, con gli occhi chiusi (ma non addormentato);
2. **Onde beta**: 14-30 Hz, si hanno in veglia e solitamente sono associate ad un livello di allerta, per esempio *stress*, *ansia*, *frustrazione* ed altri stati d'animo indesiderati;
3. **Onde delta**: 2-4 Hz, sono associate al sonno *NREM* (sonno profondo, importante per la ristorazione del sistema immunitario e della salute);
4. **Onde theta**: 5-7 Hz, sono associate al soggetto addormentamento.

2 Implementazione del Programma

2.1 Obiettivi esercitazione

Utilizzando Matlab, si vuole implementare un algoritmo per distinguere tra distribuzioni di potenza patologiche (Alzheimer - AD) e fisiologiche (soggetti sani - C) nelle registrazioni EEG. I segnali EEG sono stati acquisiti con una frequenza di campionamento di 100Hz e sono stati filtrati con un filtro passa banda (1-80Hz), per eliminare le oscillazioni lente e le alte frequenze ed un filtro stop-banda (50Hz) per eliminare il rumore di linea.

1. Calcolare la densità spettrale di potenza (PSD) per ciascun dato EEG e tracciare la PSD media di ciascun gruppo (cioè AD e C) in un'unica figura includendo anche gli intervalli di confidenza o la varianza;
2. Calcola la Potenza Assoluta (AP) e la Potenza Relativa (RP) per ogni gruppo per banda di frequenza delta, theta, alfa e beta e visualizzare media e varianza di AP e RP;
3. Applicare il test T-student per valutare se la differenza tra attività cerebrale patologica e fisiologica è significativa per PSD, AP, RP

2.2 Procedimento

E' stata implementata una funzione `my importdata()` che andrà ad utilizzare la funzione `dir()` per leggere i vari dati contenuti in una cartella (chiamata `data`) e la funzione `load()` per caricare ciascuno di essi (è stato utilizzato il comando `strcat()` che permette di ottenere il percorso di ogni singolo file concatenando il percorso della cartella con il nome di ciascun file). Per facilitare l'accesso alle singole matrici di dati è stato necessario utilizzare un cell array.

2.2.1 PSD

Per calcolare lo spettro di densità di potenza (PSD) per ogni dato EEG, si è implementata una funzione `my pwelch()`, ovvero un'implementazione del **Periodogramma di Welch**.

Metodo del Periodogramma: consiste nel suddividere il segnale in M segmenti sovrapposti di lunghezza L ottenendo così una serie di segnali di durata limitata e ad energia finita ognuno dei quali deve essere trasformato secondo Fourier (FFT) e, calcolando il modulo al quadrato della trasformata, si ottiene la sua densità spettrale di energia per ciascuno di essi. Infine effettuando la media si ricava la densità spettrale di potenza (G_{xx}).

$$\widehat{G_{xx}}(k) = \frac{1}{M} \sum_{i=1}^M \frac{2}{L} [FFT_i]^2$$

In questo programma è stato scelto di implementare la versione semplificata del *Periodogramma di Welch* (ovvero senza l'uso di finestre diverse dal gradino unitario come per esempio la *Hamming* o il *Coseno Rialzato*).

2.2.2 PSD medio e Intervallo di Confidenza

Per poter visualizzare il PSD medio (plot semilogaritmico) è stato necessario creare due matrici contenenti gli spettri dei pazienti sani (C) e di quelli patologici (AD) in modo tale da poterli confrontare; per far ciò è opportuno normalizzare gli spettri, dividendo ciascuno di essi per l'area sottesa alla curva, utilizzando la funzione di Matlab `trapz`.

Si è, inoltre, utilizzata la funzione di Matlab `pwelch()` per verificare la correttezza delle PSD medie ottenute con la funzione `mypwelch()` creata:

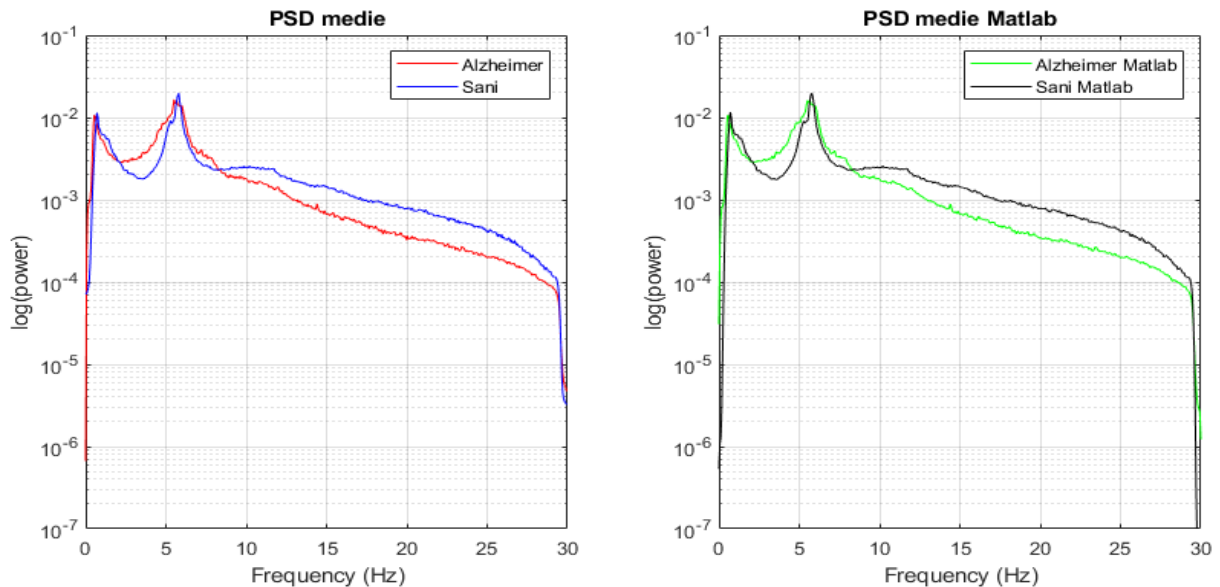


Figure 2: Confronto fra i grafici semilogaritmici delle PSD medie di soggetti sani e malati ottenute attraverso la funzione `mypwelch()` e tra le PSD medie ottenute attraverso la funzione `pwelch()` di Matlab. Come si può notare, le curve ottenute con la `mypwelch()` risultano essere molto simili a quelle ottenute con la `pwelch()` di Matlab.

Nella visualizzazione delle PSD medie sono stati inclusi anche gli intervalli di confidenza dei pazienti sani e di quelli malati.

Def: L'intervallo di confidenza è l'intervallo di valori entro i quali si stima che cada, con un livello di probabilità prescelto, il valore vero di una variabile analizzata relativa all'intera popolazione.

Per calcolare l'intervallo di confidenza attorno alla media (*mean*) è necessario (in questa implementazione) calcolare la deviazione standard (*std*) per ciascun punto dello spettro. Quindi tale intervallo sarà:

$$I_c = [mean + std, mean - std]$$

Per visualizzare graficamente tale intervallo è stata utilizzata la funzione `fillbetween()` (https://github.com/drbenvincent/fill_between).

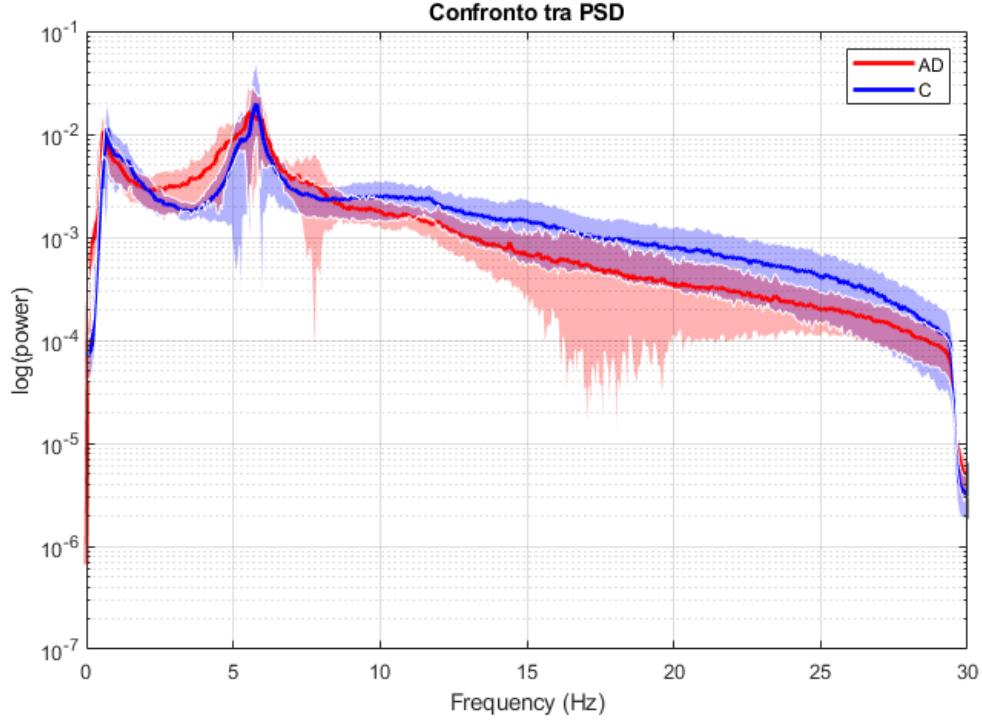


Figure 3: Confronto tra i grafici semilogaritmici delle PSD medie dei soggetti sani (blu) e dei soggetti malati (rosso) con intervalli di confidenza

2.2.3 AP e RP

È stata implementata la funzione `computeabsreldpower()` per poter calcolare i valori di Potenza Assoluta (AP) e Relativa (RP) per ogni gruppo in specifiche bande di frequenza: onde delta (2-4 Hz), onde theta (5-7 Hz), onde alpha (8-13 Hz) e onde beta (14-30 Hz).

Per ciascuna banda vengono plottati i grafici a violino, attraverso la funzione `violin()`, dei soggetti sani e malati per poterli confrontare ed avere una visualizzazione migliore della distribuzione dei dati intorno alla media.

(<https://it.mathworks.com/matlabcentral/fileexchange/45134-violin-plot>)

Def: La **potenza assoluta** è definita come lo spettro di potenza medio in una specifica banda di frequenze normalizzato rispetto alla banda di frequenze considerate:

$$AP(b) = \frac{1}{|b|} \sum_{f \in B} P(f)$$

Dove $P(f)$ è lo spettro medio di potenza in una specifica banda di frequenza (delta, theta, alpha, beta), B corrisponde all'intero spettro di frequenza (1-30 Hz) e b corrisponde alle singole bande.

Def: La **potenza relativa** invece è definita come il rapporto tra la potenza assoluta in uno specifico intervallo di frequenze b e la potenza assoluta calcolata su tutto lo spettro considerato B :

$$RP(b) = \frac{AP(b)}{AP(B)}$$

Dalla visualizzazione dei grafici si notano delle differenze tra le medie e le relative distribuzioni dei dati intorno a esse. Per valutare se queste differenze sono significative o meno è necessario applicare un test d'ipotesi. Nel caso in esame si è deciso di svolgere il test t di Student.

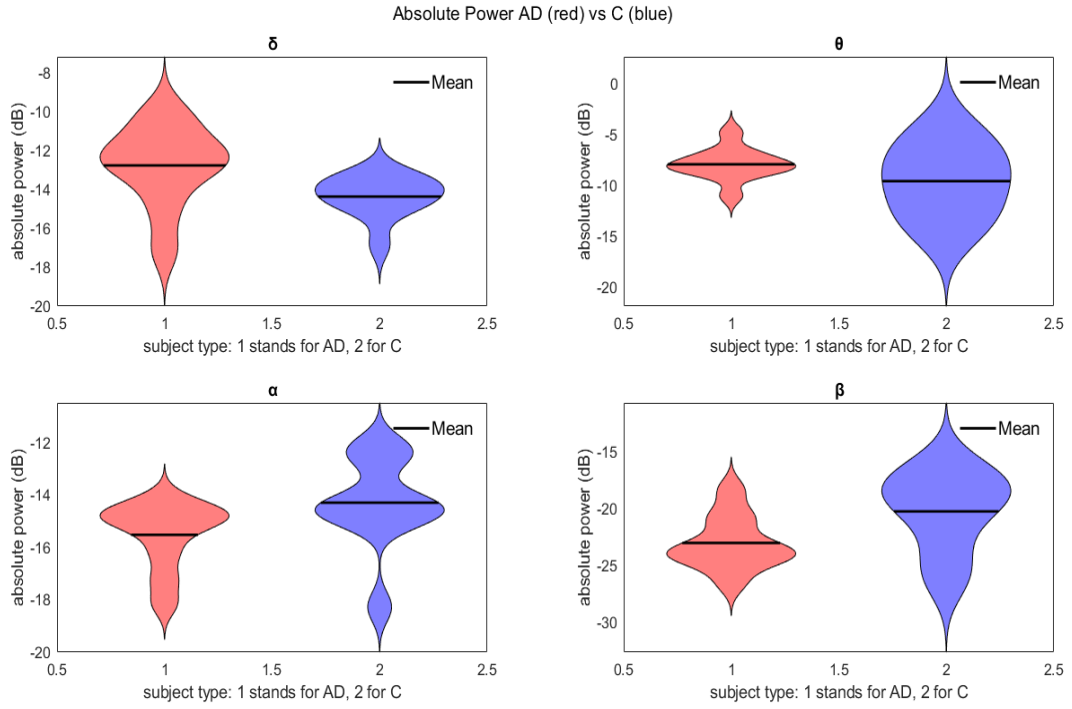


Figure 4: *Grafici a violino della potenza assoluta dei soggetti malati (AD, rosso) e di quelli sani (C, blu)*

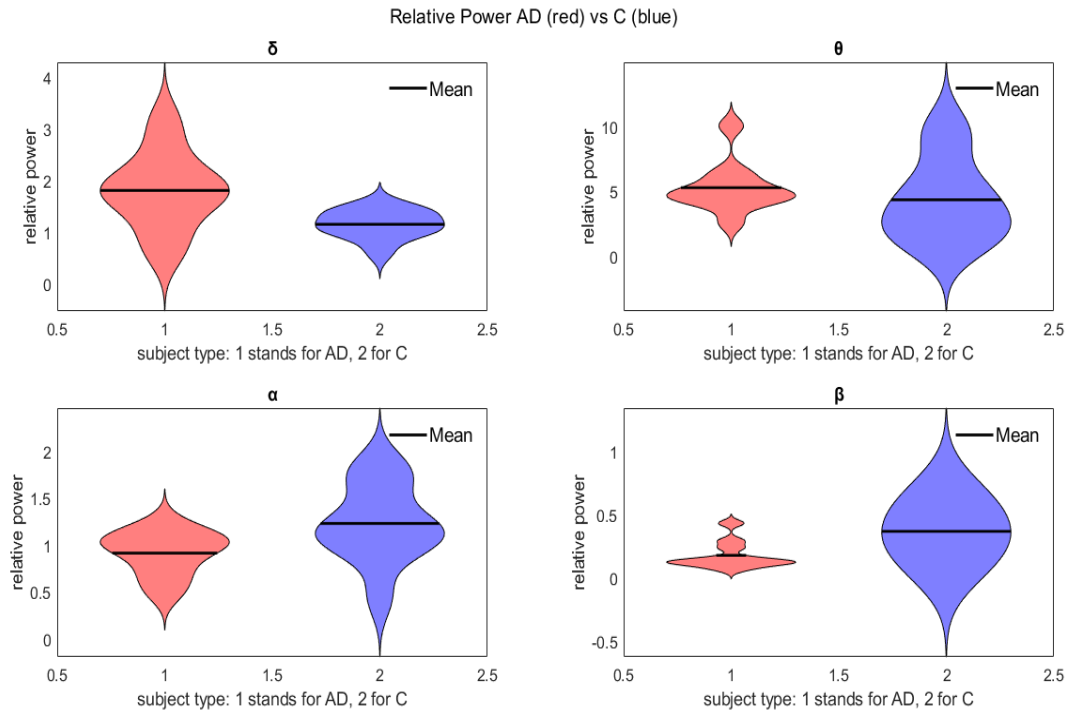


Figure 5: *Grafici a violino della potenza relativa dei soggetti malati (AD, rosso) e di quelli sani (C, blu)*

2.2.4 Test t di Student

Def: Il **t-test** (noto anche come **test t di Student**) è uno strumento usato per valutare le medie di una o due popolazioni tramite verifica d'ipotesi. Il t-test può essere usato per determinare se un singolo gruppo differisce da un valore conosciuto (*t-test a un campione*), se due gruppi differiscono l'uno dall'altro (*t-test a due campioni indipendenti*), o se c'è una differenza significativa nelle misure appaiate (*t-test a campioni dipendenti, o appaiati*). In questo caso si utilizza il t-test a due campioni indipendenti.

Per applicare il t-test è necessario avere campioni provenienti da una distribuzione originale Gaussiana: bisogna quindi

o avere la popolazione originale dalla quale si estraggono i campioni con distribuzione Gaussiana, o avere un numero di campioni superiore a 30 (rientrando così nelle condizioni pratiche del Teorema del Limite Centrale). Non avendo informazioni sulla popolazione originale e non avendo motivi per supporla Gaussiana, essendo il numero dei campioni limitato a 10 per ogni gruppo da confrontare segue che applicare il t-test è discutibile ed i risultati ottenuti non avranno un significato statistico ma puramente didattico.

Per svolgere il test si è usata la funzione di Matlab *ttest2()* : in questo caso l'ipotesi che si vuole verificare, detta ipotesi nulla H_0 , è che i campioni delle due popolazioni che confrontiamo provengano da distribuzioni normali con media e varianza uguali (quest'ultima è però sconosciuta). L'ipotesi alternativa (ipotesi H_1) sarà quindi che i campioni provengono da popolazioni con medie diverse. Il risultato del test sarà 1 se l'ipotesi H_0 viene rifiutata con livello di significatività del 5%, 0 altrimenti.

La funzione restituisce anche il **P-value**, uno scalare compreso nell'intervallo [0,1], che indica il valore tale che se il livello di significatività è maggiore di P, il test viene rifiutato. Cambiando quindi il valore del livello di significatività alpha cambia l'esito del test. Nel caso in esame quindi l'ipotesi H_0 sarà che le differenze riscontrate nelle medie sono dovute al caso e non sono quindi significative. Accettare invece l'ipotesi H_1 significherà avere differenze significative.

Riportiamo di seguito i risultati ottenuti:

	Esito del Test		P-Value
$\delta - AP$	1	Rifiutata	0.0431
$\theta - AP$	0	Accettata	0.1926
$\alpha - AP$	0	Accettata	0.0927
$\beta - AP$	0	Accettata	0.0509
$\delta - RP$	1	Rifiutata	0.0186
$\theta - RP$	0	Accettata	0.4340
$\alpha - RP$	0	Accettata	0.0522
$\beta - RP$	1	Rifiutata	0.0272
$\delta - PSD$	1	Rifiutata	1.6662*e-20
$\theta - PSD$	0	Accettata	0.1667
$\alpha - PSD$	1	Rifiutata	5.9242*e-26
$\beta - PSD$	1	Rifiutata	2.5939*e-32

2.2.5 Considerazioni Finali

Dai risultati dei vari test si trova conferma del fatto che, differenze nella media che possiamo riscontrare graficamente (attraverso i grafici a violino) non sempre sono statisticamente significative ma, solo in alcuni casi.

Nel caso delle PSD si nota che nelle bande theta, alpha e beta l'ipotesi H_0 viene rifiutata, mentre in delta è accettata.

Si nota anche che i P-value per queste 3 bande sono estremamente piccoli: per accettare l'ipotesi H_0 bisognerebbe fare un test con un valore di alfa estremamente basso, ergo con una probabilità di compiere errori tendenzialmente nulla. Ciò, comporterebbe un enorme errore di seconda specie, un tipo di errore collegato all'errore di prima specie (ergo il valore di significatività alpha) che da informazioni sulla probabilità con la quale si commette un errore nel caso si accettasse l'ipotesi nulla qualora essa fosse falsa.

Per quanto riguarda i confronti tra le potenze relative e tra quelle assolute si riscontrano dei P-value più elevati, e quindi l'ipotesi nulla è accettata ben 3 volte per le potenze assolute e 2 per quelle relative.

2.2.6 Conclusione

In conclusione, si può quindi vedere che questo algoritmo ci consente di ottenere dei buoni risultati ma la semplice ispezione visiva non è sufficiente perciò è necessario usare dei test di ipotesi per essere sicuri della significatività o meno di tali risultati.

3 Fonti e Bibliografia

- Prof L. Avanzino, *Slides del corso di Fisiologia AA 2021-2022*
- Prof G. Arnulfo, *Use-case-EEG AA 2021-2022*
- Prof M. Fato, *Slides sul Test t di Student AA 2021-2022*