

# Elaborazione dei segnali biologici

## 1. SEGNALI

I segnali sono classificati in base alla loro natura fisica:

- Bioelettrici (EEG, ECG, EMG...)
- Biomagnetici (MEG, MCG...)
- Biomeccanici (forze, pressioni, flussi...)
- Bioacustici (fonocardiogramma)
- Termici (temperatura corporea...)
- Biochimici (concentrazioni, pressioni parziali, pH...)
- Espressione genica

## SEGNALI BIOELETTRICI

Sono tra i segnali più utilizzati perché facili da misurare grazie al tipo di propagazione.

### Potenziale d'azione

Potenziale elettrico che si propaga lungo la membrana dell'assone. La sua registrazione richiede di isolare una cellula e di utilizzare un microelettrodo (dimensioni dell'ordine di alcuni micron) per stimolare la cellula e registrare la sua risposta. Interessa cellule nervose e muscolare. È la componente fondamentale di tutti i segnali bioelettrici. È diverso a seconda dei tipi di cellula ma è caratterizzato da un potenziale di riposo (circa 70mV) ampiezza nell'ordine dei mV e durata di circa 0.5s.

### Elettroencefalogramma (EEG)

Potenziale elettrico relativo all'attività cerebrale registrato sullo scalpo in posizioni standard (la posizione influenza la misura). Si utilizza per individuare e quantificare deficit dell'attività dell'encefalo, diagnosi di epilessia, monitoraggio durante anestesia, studio delle fasi del sonno effetto farmaci e droghe.

Banda: 0.1-50Hz

Aampiezza: decine di mV

Stazionario a tratti

Non periodico ma con ritmo prevalente

Ci sono vari standard per il posizionamento degli elettrodi: sistema 10/20 (20 elettrodi), sistema 10/10 (fino a 81 elettrodi) e 10/5 (oltre 300

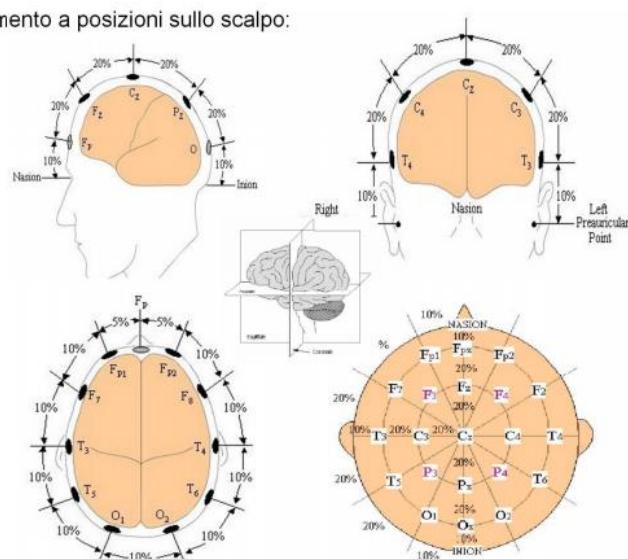
elettrodi). Un maggior numero di elettrodi assicura una maggiore risoluzione spaziale

Le lettere usate fanno riferimento a posizioni sullo scalpo:

- occipitale (O)
- parietale (P)
- centrale (C)
- frontale (F)
- temporale (T)

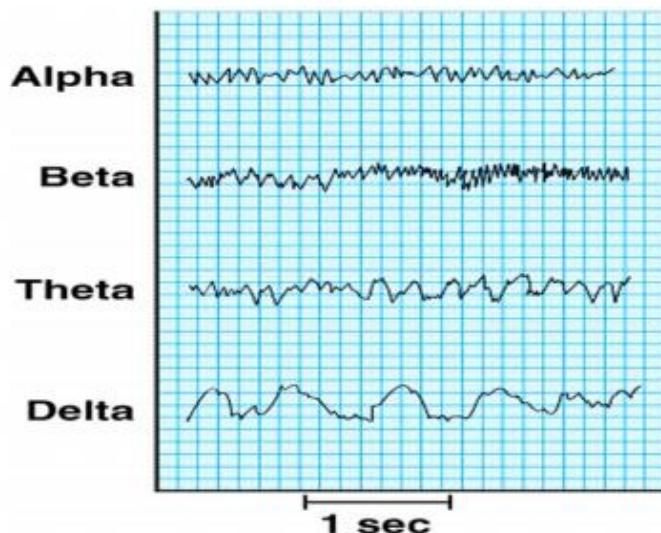
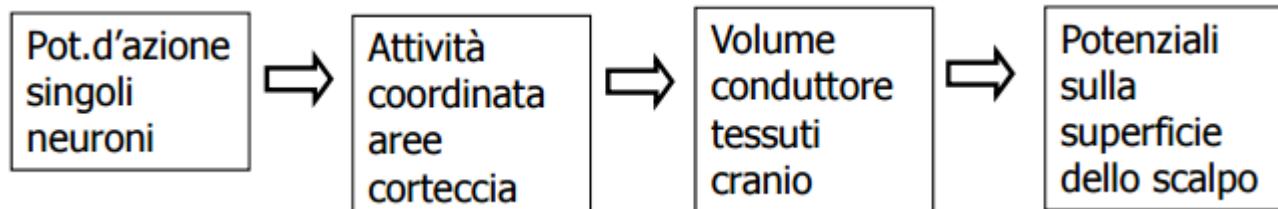
Gli indici numerici identificano il lato del cervello:

- pari, lato destro
  - dispari, lato sinistro
  - zeta, linea mediana
- e il grado di spostamento dalla linea mediana.



I neuroni corticali sono organizzati in modo da formare ammassi colonnari ad orientamento perpendicolare alla superficie della corteccia cerebrale, di cui costituiscono le unità funzionali elementari. L'EEG è l'espressione dei processi sinaptici (potenziali pre- e post-sinaptici), di potenziali dendritici e probabilmente anche di potenziali della neuroglia (cellule di sostegno). I potenziali rilevabili tramite EEG sono quelli associati a correnti all'interno dell'encefalo che fluiscono perpendicolarmente rispetto allo scalpo. Non si misura l'attività di singoli neuroni ma l'attività coordinata di aree di corteccia cerebrale. Una tecnica complementare all'EEG è la magnetoencefalografia (MEG), che permette di misurare le correnti che fluiscono parallelamente allo scalpo.

I vari neuroni sulla corteccia eseguono funzioni specifiche (vista, parola, mani, piedi...)



#### Range di frequenza e ampiezza

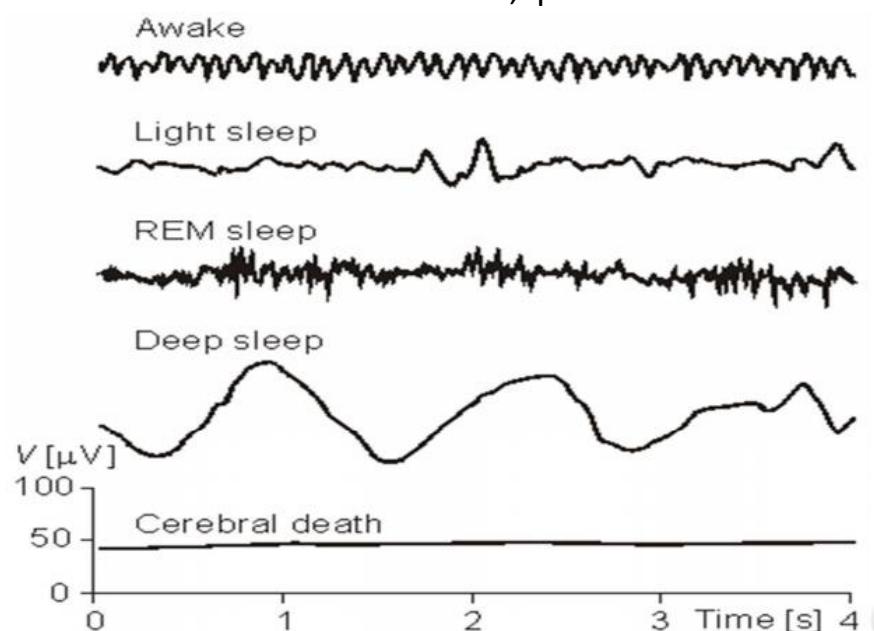
<b>Alpha:</b>	8-13Hz	50-60 $\mu$ V
<b>Beta:</b>	14-30Hz	2-20 $\mu$ V
<b>Theta:</b>	5-7Hz	
<b>Delta:</b>	.5-3.5Hz	20-100 $\mu$ V
	Gamma	30-50Hz 2-20 $\mu$ V

Onde alfa: veglia rilassata, occhi chiusi. Max ampiezza nelle regioni occipitali

Onde Beta: veglia attiva- max ampiezza nelle regioni frontali-precentrali

Onde delta: sonno profondo

Onde theta: bambini da 1 a 6 anni, ipnosi o meditazione



L'attività elettrica osservabile dall'esterno mediante elettrodi di superficie posti sullo scalpo in modo standard è la sovrapposizione di attività spontanea (EEG) + attività legata alla ricezione di stimoli esterni sensoriali (potenziali evocati). Mentre l'EEG riflette l'attività cerebrale di base, i potenziali evocati riflettono la risposta a stimoli sensoriali. Tale risposta consente in clinica di verificare ad es. l'integrità delle varie vie di conduzione nervosa (periferiche o centrali).

I potenziali evocati si classificano in base alla natura dello stimolo ( $\rightarrow$  sistema sensoriale esplorato):

- PE uditivi
- PE somato-sensoriali
- PE visivi
- PE cognitivi

I potenziali evocati presentano una successione di picchi e valli. I parametri di maggior uso clinico sono le latenze, ovvero i tempi che intercorrono tra la presentazione dello stimolo e i picchi di attività (massimi, a volte marcati con P, o minimi, a volte marcati con N) registrati.

### Stimolazione Magnetica Transcranica (TMS)

La TMS è una metodica non invasiva introdotta di recente per stimolare la corteccia cerebrale in maniera indolore. Uno stimolo magnetico applicato sullo scalpo induce un movimento nei muscoli contralaterali allo stimolo. Lo scopo è studiare il funzionamento dei circuiti e delle connessioni neuronali all'interno del cervello provocando uno squilibrio ridotto e transitorio. Il funzionamento si basa sulla legge di Faraday (una corrente elettrica in uno stimolatore produce un campo magnetico, e la variazione del campo magnetico nel tempo induce un flusso di corrente nei conduttori vicini, inclusi i tessuti umani.).

È possibile combinare EEG e TMS, ovvero effettuare la registrazione dell'EEG durante la TMS per studiare come lo stimolo può modulare l'attività oscillatoria corticale.

Si possono combinare anche potenziali evocati motori (PEM) e TMS per studiare i periodi di latenza (La latenza è di circa 20ms per un muscolo distale dell'arto superiore e di 30ms per un muscolo della gamba, compatibile con la conduzione di fibre corticospinali di grosso calibro).

### Elettromiogramma (EMG)

È la registrazione del potenziale extracellulare delle fibre muscolari, strettamente dipendente dal potenziale intracellulare.

Banda: 5Hz-10KHz

Aampiezza: 100microV-qualche mV

Si misura con elettromiografi utilizzando elettrodi di superficie o elettrodi ad aghi. Si utilizza per la diagnosi di patologie a carico del sistema nervoso e muscolare, per l'analisi del movimento e il controllo di protesi.

I muscoli sono formati da unità motorie (insieme di fibre innervate da un neurone) di actina e miosina che si attivano grazie all'acetilcolina a sua volta liberata grazie al segnale che arriva da un assone.

Il potenziale d'azione raggiunge i rami terminali dell'assone che innerva le fibre muscolari  $\rightarrow$

Si libera un mediatore chimico (acetilcolina)  $\rightarrow$

Si genera il potenziale d'azione muscolare (l'acetilcolina altera la permeabilità della membrana)  $\rightarrow$

Il pda muscolare stimola la liberazione di ioni CA<sup>2+</sup> che permettono l'aggancio delle teste globulari di miosina al filamento di actina, facendo scorrere le fibre le une sulle altre  $\rightarrow$

Il muscolo si contrae. Con l'aumentare dello sforzo muscolare aumenta il numero di unità motorie reclutate.

SMU (Single Motor Unit) è il segnale generato da una singola unità motoria (insieme di fibre muscolari innervate da un singolo neurone). Viene misurata durante un leggero sforzo muscolare. Se aumenta lo sforzo:

- sommazione temporale (aumenta la frequenza di scarica di una SMU)
- sommazione spaziale (vengono reclutate più unità motorie).

Si arriva così al tracciato di interferenza.

## Elettrocardiogramma (ECG)

È la registrazione sul torace dell'attività elettrica del miocardio.

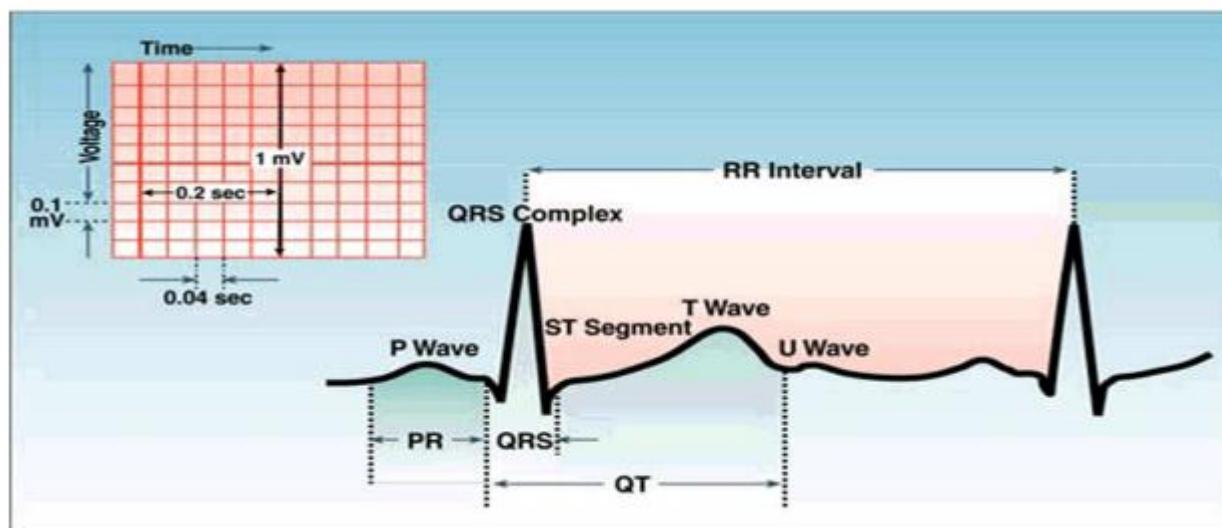
Banda: 0.1-100Hz

Aampiezza: decine di mV

Pseudoperiodico

Si utilizza per la diagnosi di alterazioni/patologie come aritmie, alterazioni della conduzione, ischemia, infarto, ipertrofia ventricolare.

Il potenziale si origina nel nodo seno-atriale grazie alle cellule pace maker, particolari cellule con un PdA che si autostimola non avendo una fase costante. Si propaga prima alle cellule degli atrii, poi attraverso il fascio di His a quelle dei ventricoli. La somma dei PdA determina la contrazione (ritmica e con la giusta temporizzazione) delle cellule delle 4 camere del cuore. Attraverso i tessuti nel torace, si propaga alla superficie esterna.



ONDA P: depolarizzazione degli atrii a partire dal nodo seno atriale

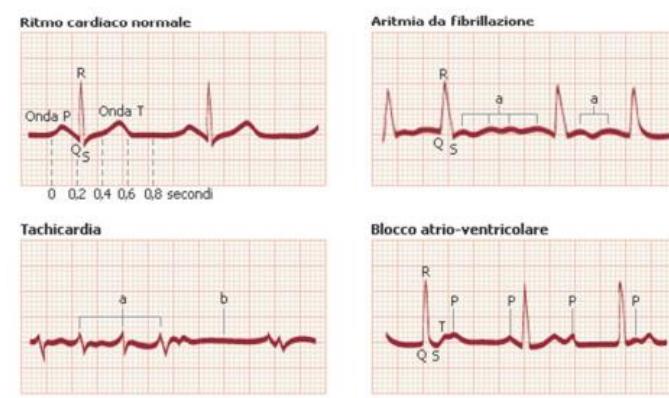
TRATTO PQ: pausa nodo AV

COMPLESSO QRS: depolarizzazione ventricoli, contrazione del cuore

TRATTO ST: cuore contratto

ONDA T: ripolarizzazione dei ventricoli

In condizioni patologiche, l'andamento del tracciato risulta alterato e la sequenza delle onde non viene rispettata. In particolare, nella tachicardia vi è un aumento del numero delle pulsazioni, che diventano più rapide e deboli (tratto a) e quindi si arrestano (tratto b); nell'aritmia da fibrillazione, le onde ORS non sono equidistanti, come avviene in condizioni normali, e le onde P sono sostituite dal tratto a, che esprime la fibrillazione degli atrii (contrazione non coordinata delle fibre muscolari cardiache); nel blocco atrio-ventricolare si osserva una distribuzione irregolare di onde QRS e di onde P.



Dall'ECG si possono derivare altri segnali. Per esempio la Heart Rate Variability che è una tecnica per misurare ed analizzare la variabilità della frequenza cardiaca da cui si possono dedurre molte informazioni come ad esempio il rischio di aritmie cardiache ed infarto.

### Elettroneurogramma (ENG)

L'obiettivo è la misura della velocità di conduzione di un nervo (sensoriale o motorio). Ad es. si stimola il nervo sensoriale mediante elettrodi ad anello applicati al dito indice, e si registra in due punti a distanza nota D lungo il percorso del nervo. La velocità di conduzione dei nervi motori è normalmente di circa 50 m/s. È ridotta nelle polineuropatie e neuropatie infiammatorie oppure nelle neuropatie locali da compressione meccanica, come la sindrome del solco del nervo ulnare o la sindrome del tunnel carpale

Aampiezza: decine di microV

### Elettrooculogramma (EOG)

È la registrazione del potenziale tra cornea e retina. Viene utilizzato per monitorare i movimenti dell'occhio, per scopi soprattutto di ricerca (studio delle fasi del sonno). È misurato con coppie di elettrodi posizionati ai lati dell'occhio.

Aampiezza: 10 microV - 5 mV

### Elettroretinogramma (ERG)

E' la registrazione del potenziale generato dalla retina in risposta ad uno stimolo luminoso. Viene usato in clinica, e per ricerca in oftalmologia. E' misurato con elettrodi inseriti su delle lenti a contatto.

Aampiezza: 0.5 microV-1 mV.

## SEGNALI BIOMAGNETICI

Vari organi, quali il cuore, i polmoni e il cervello, producono un campo magnetico, che può essere registrato sulla superficie esterna. Si parla quindi di magnetocardiogramma (MCG), magnetopneumogramma (MPG) e di magnetoencefalogramma (MEG). L'esiguità dei segnali (10-15T) richiede strumentazione complessa e costosa pertanto gli studi sono ancora confinati ad un ambito di ricerca. Il MEG è però più informativo rispetto all'EEG perché permette di individuare le sorgenti tangenziali da quelle radiali.

## SEGNALI BIOACUSTICI

### Fonocardiogramma (PCG)

E' la registrazione dei suoni generati dalle vibrazioni del sistema cardiovascolare (apertura e chiusura delle valvole, suoni prodotti dal flusso di sangue nei grossi vasi). Viene registrato con microfoni o trasduttori di pressione posti sul torace. Dà informazioni sul ritmo, la contrattilità cardiaca e sull'eventuale presenza di patologie o anomalie a carico del sistema cardiovascolare. Solitamente si effettua una registrazione simultanea di ECG, PCG e pressione carotidea.

### Suoni respiratori

Si possono effettuare anche registrazioni dei suoni respiratori ovvero dell'aria che entra negli alveoli polmonari e poi passa nei bronchi e nella trachea. Si riconoscono 4 diversi tipi di segnali: alveolare (VBS), bronchiale (BBS) bronco-alveolare (BVBS) e tracheale (TBS), misurati con microfoni in posizioni diverse del torace. Sono comunque segnali pseudoperiodici, con un range di frequenza che va da 20Hz a 2 KHz.

### Voce

Il segnale che si registra è interessante nell'ottica di trasmettere e sintetizzare il parlato. Può essere anche utilizzato a scopi diagnostici, in quanto riflette la presenza di anomalie e patologie dell'apparato vocale. E' prodotto dall'espulsione di aria dai polmoni attraverso la trachea fino al tratto vocale. La forma del tratto vocale varia per produrre tipi diversi di suoni elementari (fonemi) che, concatenati tra loro, formano le

parole. In pratica, il tratto vocale si comporta come un filtro che modula il segnale che riceve in ingresso. E' evidente quindi che il segnale vocale è fortemente nonstazionario.

## SEGNALI BIOMECCANICI

Segnali di natura meccanica come pressioni, flussi, movimenti, variabili cinematiche...

## SEGNALI BIOCHIMICI

Si tratta della registrazione delle fluttuazioni temporali di grandezze chimiche organiche e nonorganiche, quali concentrazioni di substrati, ormoni, enzimi, ioni, pressioni parziali di gas, PH, ecc. Le misure sono generalmente eseguite su campioni di sangue, oppure nell'aria espirata e vengono eseguite in laboratori clinici o di ricerca.

Gli sviluppi recenti dell'elaborazione dei segnali stanno nell'integrazione tra segnali prelevati da posizioni diverse o segnali di natura diversa, nell'integrazione segnali-immagini e nell'analisi di segnali high-throughput a livello molecolare.

## 2. Filtri

Matematicamente un segnale si rappresenta come una funzione di una o più variabili indipendenti di cui una solitamente il tempo. Un segnale è una funzione del tipo:

$x(t)$  segnale a tempo continuo

$x(n)$  segnale a tempo discreto

per i segnali biologici si possono applicare descrizioni e metodi di analisi molto diversi tra loro in quanto i segnali stessi hanno caratteristiche e forme molto diverse tra loro.

Esistono grandezze deterministiche (note con precisione, rappresentabili in modo univoco e una sola osservazione è rappresentativa del fenomeno.) e grandezza aleatorie (non assumono valori univoci, una singola osservazione non è rappresentativa del fenomeno). Utilizziamo l'approccio deterministico per studiare fenomeni semplici con osservazioni singole (età di una persona) e l'approccio aleatorio per studiare fenomeni complessi e che coinvolgono un alto numero di realizzazioni.

### SEGNALI A TEMPO DISCRETO

Sono spesso ottenuti per il campionamento di segnali continui e sono del tipo  $x(n)$  definito per valori interi di  $n$ . Passando da segnale campionato a segnale discreto si perde la nozione di tempo di campionamento.

$x(t)$  è il segnale continuo

$x^*(t) = \sum_{n=-\infty}^{\infty} x(nT) \cdot \delta(t - nT)$

è il segnale continuo  
campionato (diverso da zero  
solo negli istanti di campionamento)

$x(n) = x(nT)$

è il segnale discreto,  
definito per i soli valori  
interi di  $n$

Per la rappresentazione in frequenza di segnali a tempo discreto si utilizza la trasformata di Fourier continua:

$X(\omega) = FT[x(n)] = \sum_{n=-\infty}^{\infty} x(n)e^{-j\omega n}$  ottenendo una funzione complessa in  $w$  periodica di periodo  $2\pi$ . Se  $x(n)$  è reale, allora  $X(w)$  ha modulo pari e fase dispari pertanto è sufficiente rappresentarla nell'intervallo  $[0:\pi]$ . Le principali proprietà della FT sono:

Linearità:  $FT[ax_1(n)+bx_2(n)] = aX_1(\omega) + bX_2(\omega)$

Traslazione:  $FT[x(n+k)] = X(\omega) e^{jk\omega}$

Convoluzione:  $y(n) = \text{somma di conv tra } x(n) \text{ e } h(n)$   
 $Y(\omega) = X(\omega) H(\omega)$

Per i segnali  $x(n)$  di durata finita ( $n=0 \dots N-1$ ) è possibile definire anche la Trasformata di Fourier Discreta:

$$X(k) = DFT[x(n)] = \sum_{n=0}^{N-1} x(n) e^{-j\frac{2\pi}{N} nk}$$

$K=0 \dots N-1$

è un segnale complesso ad ascissa discreta. Di fatto  $X(k)$  è il campionamento di  $X(w)$  in un periodo di  $2\pi$  con passo  $2\pi/N$ .

Partendo dal segnale campionato  $X^*(\Omega)$ , con un cambio di variabile  $w=\Omega T$ , otteniamo  $X(w)$ .

Quindi la pulsazione e la frequenza sono correlate, all'intervallo  $[0:\pi]$  corrisponde  $[0:1/2Fc]$ .

FT del segnale campionato

$$X^*(\Omega) = FT[x^*(t)] = FT\left[\sum_{n=-\infty}^{\infty} x(nT) \delta(t - nT)\right] = \sum_{n=-\infty}^{\infty} x(nT) e^{-j\Omega nT}$$

FT della sequenza

$$X(\omega) = FT[x(n)] = \sum_{n=-\infty}^{\infty} x(n) e^{-j\omega n}$$

Essendo  $x(n) = x(nT)$ , si ha  $X(\omega) = X^*(\Omega)$  con  $\omega = \Omega T$   
quindi rappresentiamo la FT della sequenza:

in termini di pulsazione $\omega$	da 0 a $\pi$
in termini di frequenza $f$	da 0 a 0.5

che corrispondono, nel continuo, ai seguenti intervalli

in termini di pulsazione $\Omega$	da 0 a $0.5\Omega_c = \pi Fc$
in termini di frequenza $F$	da 0 a $0.5Fc$
MATLAB	da 0 a 1

$\omega$  viene indicata come pulsazione normalizzata  $\omega = \Omega T = \Omega / Fc$   
 $f$  come frequenza normalizzata  $f = F / Fc$

## TRASFORMATA ZETA

$$X(z) = \text{ZT}[x(n)] = \sum_{n=-\infty}^{\infty} x(n)z^{-n}$$

Dato un segnale  $x(n)$ , la sua trasformata zeta è definita da:  
di variabile complessa  $z$ . È utile per rappresentare segnali e sistemi a td.

La regione in cui la ZT converge si chiama regione di convergenza. Si ha convergenza se

. Le regioni di convergenza sono dei settori circolari del tipo  $R^- < |z| < R^+$ . La ZT gode delle seguenti proprietà:

Linearità:  $\text{ZT}[ax_1(n)+bx_2(n)] = aX_1(z)+bX_2(z)$

Traslazione  $\text{ZT}[x(n+k)] = X(z)z^k$

Convoluzione  $y(n) = \text{somma di conv tra } x(n) \text{ e } h(n)$   
 $Y(z) = X(z)H(z)$

La FT si ottiene valutando la ZT sul cerchio di raggio unitario  $z = e^{j\omega}$ :  $X(\omega) = X(z)|_{z=\exp(j\omega)}$

## SISTEMI

Un sistema è una trasformazione che in corrispondenza di un segnale di ingresso  $x(n)$ , produce un'uscita  $y(n)$ .

Nei sistemi LTI (Lineari Tempo Invarianti) la risposta del sistema  $h(n)$  al campione unitario  $\delta(n)$ , descrive completamente il sistema.  $y(n) = \sum_{k=-\infty}^{\infty} x(k) \cdot h(n-k) = \sum_{k=-\infty}^{\infty} x(n-k) \cdot h(k)$

I sistemi possono essere:

*FIR (Finite Impulse Response) se  $h(n)$  ha durata finita*

*IIR (Infinite Impulse Response) se  $h(n)$  è di durata non finita.*

Dato che  $h(n)$  riassume in sé tutte le caratteristiche del sistema LIT, è possibile testare su  $h(n)$  le proprietà di stabilità e causalità:

- Stabilità BIBO garantisce che la risposta del sistema non diverga de l'ingresso è limitato.  $\exists M : \sum_{n=-\infty}^{\infty} |h(n)| < M$   
Condizione necessaria e sufficiente a questo è:
- Causalità garantisce la possibilità di elaborare il segnale in linea. Condizione necessaria e sufficiente è  $h(n)=0$  per  $n < 0$ .

Esiste una sottoclasse di SLIT in cui il legame I/O è esprimibile come equazione alle

$$\sum_{k=0}^N a_k y(n-k) = \sum_{k=0}^M b_k x(n-k) \quad y(n) = -\sum_{k=1}^N a_k y(n-k) + \sum_{k=0}^M b_k x(n-k)$$

differenze: o in forma causale: Parte AR      Parte MA

**MATLAB**  
 $b=[b_0 \ b_1 \dots \ b_M]$   
 $a=[1 \ a_1 \dots \ a_N]$   
 $y=filter(b,a,x)$

I sistemi FIR, con  $h(n)$  di durata finita, possono essere descritti da equazioni non ricorsive ovvero con la sola parte MA.

I sistemi IIR devono essere descritti da equazioni alle differenze ricorsive, ovvero con parte AR+MA.

## FILTRI NUMERICI

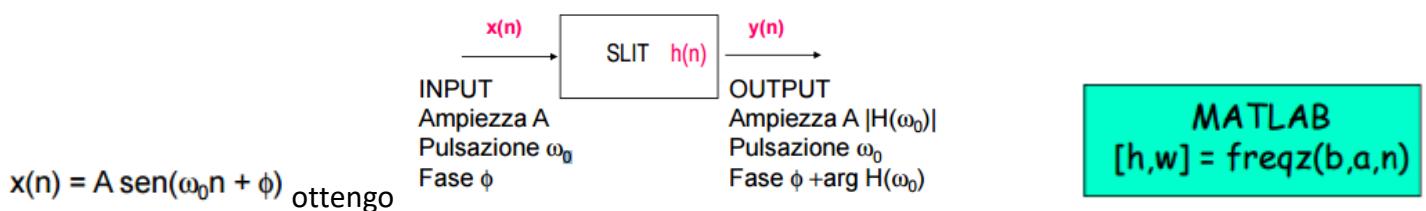
Un sistema a td lineare, invariante e stabile si dice filtro numerico. Viene applicato ad un segnale a tempo discreto per *modificare il suo contenuto in frequenza*, ad esempio per migliorare il SNR. Si lavora quindi nel

dominio delle frequenze.



$$Y(w) = H(w)X(w) \text{ con } H(W) = \text{FT}(h(n))$$

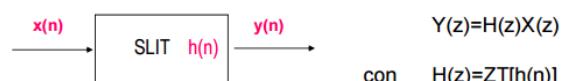
$H(w)$  è la risposta in frequenza del sistema. Si chiama così perché dato in ingresso un segnale sinusoidale



Noi vorremmo filtri che modificano solo il modulo, quindi con  $\arg H(w)=0$  per ogni  $w$ .

Questo si verifica se  $h(n)$  è pari ( $h(n)=h(-n)$ ) ovvero se il *sistema non è causale*.

Possiamo allora utilizzare una *fase lineare* che introduce nel segnale solo uno shift. Una fase lineare è del tipo  $\arg H(w)=kw$ . Con questa non si alterano le relazioni temporali tra le varie componenti del segnale. Questa condizione si verifica se  $h(n)$  è una sequenza simmetrica ( $h(N+n)=h(N-n)$ ).



Un sistema può essere rappresentato con la ZT:  $H(z)$  è la funzione di trasferimento del sistema

$$H(z) = \frac{\sum_{k=0}^M b_k z^{-k}}{\sum_{k=0}^N a_k z^{-k}}$$

I poli sono i valori di  $z$  per cui il denominatore di  $H(z)$  si annulla mentre gli zeri sono i valori per cui il numeratore si annulla. Poli e zeri possono essere reali o complessi coniugati.

MATLAB  
poli = roots(a)  
zeri = roots(b)

Anche con la ZT si possono valutare le proprietà del sistema:

- Stabilità → la ROC di  $H(z)$  contiene il cerchio di raggio unitario  $|z|=1$ :  $\exists M : \sum_{n=-\infty}^{\infty} |h(n)| < M$
- Causalità → la ROC di  $H(z)$  è la zona esterna ad un cerchio  $h(n)=0 \quad n < 0$

Per avere *stabilità e causalità* è necessario che tutti i poli siano all'interno del cerchio unitario.

I sistemi FIR hanno  $h(n)$  di durata finita e possono essere descritti da equazioni alle differenze non ricorsive (MA). Non hanno quindi parte regressiva e hanno solo poli nell'origine. Sono sempre stabili.

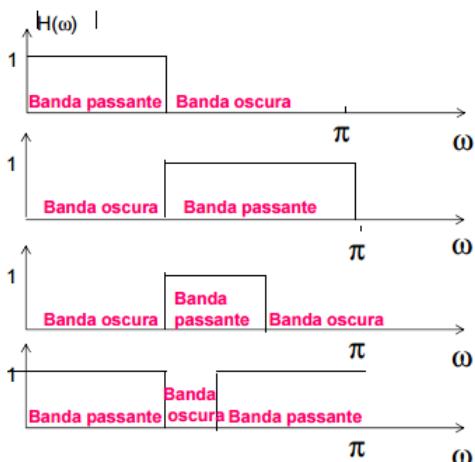
I sistemi IIR, hanno  $h(n)$  di durata non finita e devono essere descritti da equazioni alle differenze ricorsive (AR+MA). Hanno sicuramente almeno un polo fuori dall'origine per cui non è scontata la stabilità.

I filtri numerici sono sistemi a tempo discreto, lineari, invarianti alla traslazione e stabili. Vengono

rappresentati da equazione alle differenze  $b=[b_0 \ b_1 \dots \ b_M]$ ,  $a=[1 \ a_1 \dots \ a_N]$ ,  $y=filter(b,a,x)$ , risposta in frequenza  $[h,w] = freqz(b,a,n)$ , trasformata

MATLAB  
poli = roots(a)  
zeri = roots(b)

zeta e diagrammi zeri-poli nel piano complesso.



Passa basso

Passa alto

Passa banda

Elimina banda

FILTRI IDEALI → sono definiti in base al modulo della loro risposta in frequenza. Si evidenzia quali frequenze lasciano passare inalterate (banda passante) e quali si eliminano (banda oscura). Nella realtà si riescono ad ottenere solo approssimazioni del comportamento ideale, tanto più buone quanto è maggiore l'ordine del filtro.

Per ottenere filtri a fase sempre nulla (perdendo la causalità), si utilizza la tecnica Forward-Backward



$$Y(\omega) = H(\omega) H(-\omega) X(\omega) = |H(\omega)|^2 X(\omega)$$

<b>MATLAB</b> $b=[b_0 \ b_1 \dots \ b_M]$ $a=[1 \ a_1 \dots \ a_N]$ $y=filtfilt(b,a,x)$
--

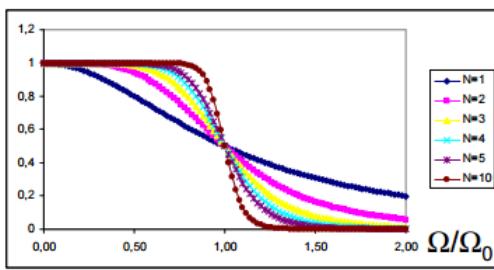
Il filtraggio F-B eleva al quadrato il modulo della risposta in frequenza del filtro H ed annulla la sua fase.

### PROGETTO DI FILTRI

Consiste nel trovare il filtro che approssima meglio le specifiche date.

Es. filtri di Butterworth

$$|H(\Omega)|^2 = 1/(1+(\Omega/\Omega_0)^{2N})$$



- **FILTRI IIR** si usano dei prototipi i cui parametri vengono adattati per soddisfare le specifiche. Sono ottenuti da filtri analogici (piano s) per trasformazione (dal piano s al piano z). Sono filtri passa basso.  $\Omega_0$  sceglie il punto di flessione o banda di taglio, N modula la banda di transizione.

- **FILTRI ELITTICI** consentono di realizzare un insieme di specifiche con il valore minimo dell'ordine. In matlab si realizzano con 2 comandi:

- $[N, W_n] = \text{Ellipord}(W_p, W_s, R_p, R_s)$  date le specifiche (es. per filtro passabasso banda passante, banda di transizione, ripple in banda passante, ripple in banda oscura) valuta l'ordine minimo (N) per le specifiche del filtro passate in input. Inoltre dà la frequenza naturale da passare al comando dopo (cut off frequency). Rp ed Rs sono i ripple e vanno passati in Db, Wp e Bs definiscono il tipo di filtro e vanno normalizzati a 1.  
 $[N, Wn] = \text{ELLIPORD}(Wp, Ws, Rp, Rs)$  returns the order N of the lowest order digital elliptic filter that loses no more than Rp dB in the passband and has at least Rs dB of attenuation in the stopband.  
Wp and Bs are the passband and stopband edge frequencies, normalized from 0 to 1 (where 1 corresponds to pi radians/sample). For example,  
Lowpass:  $Wp = .1, Ws = .2$   
Highpass:  $Wp = .2, Ws = .1$   
Bandpass:  $Wp = [.2 .7], Bs = [.1 .8]$   
Bandstop:  $Wp = [.1 .8], Bs = [.2 .7]$

- $[B, A] = \text{ELLIP}(N, Rp, Rs, Wn)$  calcola il filtro restituendo i coefficienti nei vettori di lunghezza N+1, B e A  
 $[B, A] = \text{ELLIP}(N, Rp, Rs, Wn)$  designs an Nth order lowpass digital elliptic filter with Rp decibels of ripple in the passband and a stopband Rs decibels down. ELLIP returns the filter coefficients in length N+1 vectors B (numerator) and A (denominator). The cutoff frequency Wn must be  $0.0 < Wn < 1.0$ , with 1.0 corresponding to half the sample rate. Use Rp = 0.5 and Rs = 20 as starting points, if you are unsure about choosing them.

If Wn is a two-element vector, Wn = [W1 W2], ELLIP returns an order 2N bandpass filter with passband  $W1 < W < W2$ .  
 $[B, A] = \text{ELLIP}(N, Rp, Rs, Wn, 'high')$  designs a highpass filter.  
 $[B, A] = \text{ELLIP}(N, Rp, Rs, Wn, 'stop')$  is a bandstop filter if Wn = [W1 W2].

- **FILTRI FIR** per questi filtri si specifica inizialmente l'andamento del modulo della risposta in frequenza per valori discreti di w:  $|H_d(\omega_i)| \quad i=1,2,\dots,Q$  Fissato l'ordine del filtro FIR, il suo modulo si può scrivere come  $|H(\omega)| = |b_0 + b_1 e^{j\omega} + b_2 e^{j2\omega} + \dots + b_Q e^{jQ\omega}|$ . Si cercano i valori dei coefficienti  $b_0, \dots, b_Q$  che minimizzano la distanza tra  $H_d(\omega_i)$  e  $H(\omega_i)$ , ad es. la somma dei quadrati delle distanze (si usano per questo tecniche di ottimizzazione). Si richiede fase lineare quindi  $b_k = b_Q \cdot k$ . In matlab si utilizza la funzione FIRPM.

`b=firpm(n,f,a)` restituisce un vettore `b` di  $N+1$  coefficienti secondo le caratteristiche di `f` e di `a`. Il filtro che si ottiene così ha fase lineare. `f` è il vettore di coppie di frequenze che specificano le bande. Ad esempio per un filtro passabasso  $f=[0, W_p, W_s, 1]$ . Le frequenze in `f` sono specificate nel range 0,1 dove 1 corrisponde alla frequenza di Nyquist. `a` specifica il tipo di filtro, contenendo le ampiezze nei punti specificati da `f`. Ad esempio per un filtro passa basso `a=[1,1,0,0]`. `A` e `f` devono essere della stessa dimensione.

`b=firpm(n,f,a,w)` utilizza i pesi nel vettore `w` per pesare l'errore. `W` ha un'entrata per banda (quindi la sua dimensione è metà di quella di `f` e `a`).

`b = firpm(n,f,a)` returns row vector `b` containing the  $n+1$  coefficients of the order  $n$  FIR filter whose frequency-amplitude characteristics match those given by vectors `f` and `a`.

The output filter coefficients in `b` obey a symmetry relation

Vectors `f` and `a` specify the frequency-magnitude characteristics of the filter:

- `f` is a vector of pairs of normalized frequency points, specified in the range between 0 and 1, where 1 corresponds to the Nyquist frequency. The frequencies must be in increasing order.
- `a` is a vector containing the desired amplitudes at the points specified in `f`.
- The desired amplitude at frequencies between pairs of points  $(f(k), f(k+1))$  for  $k$  odd is the line segment connecting the points  $(f(k), a(k))$  and  $(f(k+1), a(k+1))$ .
- The desired amplitude at frequencies between pairs of points  $(f(k), f(k+1))$  for  $k$  even is unspecified. The areas between such points are transition or "don't care" regions.
- `f` and `a` must be the same length. The length must be an even number.

`b=firpm(n,f,a,w)` uses the weights in `w` to weight the error. `W` has one entry per band (so it is half the length of `f` and `a`) which tells how much emphasis to put on minimizing the error in each band relative to the other bands.

`[b,err] = firpm(...)` returns the maximum ripple height in `err`

`[b,err,res] = firpm(...)` returns a structure `res` with the following fields.

### 3. VARIABILI AEATORIE

È il modello adatto a descrivere un esperimento "governato dal caso", quindi con qualche elemento di causalità, in cui quello che si osserva è un esito numerico, che non è quindi completamente prevedibile a priori. Ciò non vuol dire che l'esito sia completamente imprevedibile, ad es. nel lancio del dado so quali sono i possibili esiti, e so anche che, se il dado non è truccato, questi esiti sono equiprobabili.

Statistica descrittiva → si occupa dello studio di popolazioni, cioè di aggregati di individui, dove per individuo non si intende necessariamente un essere vivente o un individuo materiale, es. una popolazione può essere un insieme di misure. Gli elementi che ci interessano sono:

MEDIA  $x_m = \frac{1}{N} \sum_{i=1}^n x_i$  in MATLAB  $m=\text{mean}(x)$

MEDIANA valore dell'elemento in mezzo alla popolazione (media dei due valori in mezzo se N pari). La mediana è meno sensibile agli outliers in MATLAB con  $\text{hist}(x)$

ISTOGRAMMA per rappresentare la variabilità di una serie di campioni. In ascissa c'è il range di valori di campioni, diviso in intervalli. In ordinata c'è il numero di campioni in quel range. In MATLAB  $\text{hist}(x)$

VARIANZA  $s_x^2 = \frac{1}{N} \sum_{i=1}^n (x_i - x_m)^2$

DEVIAZIONE STANDARD Dev standard =  $\sqrt{\text{varianza}}$  in MATLAB  $x=\text{std}(y)$

Media e dev standard sintetizzano: il valore attorno al quale si dispongono i vari elementi del campione e quanto i vari elementi sono dispersi attorno al valore medio

*Una variabile aleatoria è una funzione che associa ad ogni campione di uno spazio di probabilità un valore numerico (discreto o continuo).*

Esistono va discrete che assumono un numero finito di valori, ciascuno associato ad una probabilità. In questo caso la probabilità è definita come il rapporto tra il numero di casi favorevoli e i casi possibili (con eventi equiprobabili).

Le va continue invece assumono valori appartenenti ad un insieme continuo. Per descrivere queste variabili conviene partire dalla funzione di distribuzione:  $F_A(a) = P[A \leq a]$ . La funzione distribuzione ha una importanza fondamentale per la descrizione statistica della v.a. A. Infatti dalla conoscenza della distribuzione si può ricavare la probabilità che A appartenga ad un generico insieme dell'asse reale  $P[a_1 \leq A \leq a_2] = P[A \leq a_2] - P[A \leq a_1] = F_A(a_2) - F_A(a_1)$

Si definisce densità di probabilità la derivata della funzione distribuzione:  $f_A(a) = dF_A(a)/da$ . Valgono le

$$\begin{aligned} f_A(a) &\geq 0 \\ \int_{-\infty}^{a_1} f_A(a) da &= F_A(a_1) \end{aligned}$$

seguenti proprietà:  $\int_{-\infty}^{+\infty} f_A(a) da = 1$ . Dalla conoscenza di  $f(a)$  si può calcolare la probabilità che A appartenga ad

$$\text{un certo insieme reale: } P[a_1 \leq A \leq a_2] = \int_{a_1}^{a_2} f_A(a) da$$

Mentre la  $P$  di una v.a. discreta in un certo valore  $a_1$  rappresenta la probabilità che essa assume il valore  $a_1$ , la ddp di una v.a. continua in un certo valore  $a_1$  NON rappresenta la probabilità che essa assume il valore  $a_1$ . Infatti la ddp di una v.a. continua NON è una probabilità ma una densità di probabilità, ovvero solo il suo integrale su un intervallo ha il significato di probabilità. Con la ddp si possono fissare degli intervalli di confidenza: fissato un livello di significatività  $\alpha$  possiamo cercare gli estremi  $x$  e  $y$  nei quali cade una percentuale  $\alpha$  di valori:

$$\begin{aligned} \int_x^y f_A(a) da &= \\ &= F_A(y) - F_A(x) = \alpha \end{aligned}$$

$$E[A] = \sum_{a_i \in A} a_i P[a_i] \quad \text{o} \quad E[A] = \int_{-\infty}^{+\infty} a \cdot f_A(a) da$$

Per una va definiamo la media o aspettazione come

Gode di linearità:  $E[b_0 + b_1 A_1 + \dots + b_n A_n] = b_0 + b_1 E[A_1] + \dots + b_n E[A_n]$

Il teorema dell'aspettazione è utile per conoscere la media di una funzione  $g(A)$  senza conoscerne la ddp ma

$$E[g(A)] = \sum g(a_i) P[a_i]$$

$$E[g(A)] = \int_R g(a) f_A(a) da$$

conoscendo solo la ddp di  $A$ :

Si definisce:

momento di ordine  $k$

$$m_A(k) = E[A^k]$$

varianza

$$\sigma_A^2 = E[(A - m_A)^2]$$

conoscendo solo media e SD ma non la ddp, si possono valutare solo dei bound per gli intervalli di confidenza con la diseguaglianza Chebychev:  $\text{Prob}[|a-m|>kSD] \leq 1/k^2$  ovvero  $\text{Prob}[|a-m|<kSD] \geq 1-1/k^2$

## MODELLI DI DISTRIBUZIONE DI VA

- Binomiale (Bernoulli)

$$\text{Prob}[X = k] = \binom{N}{k} p^k (1-p)^{N-k}$$

$$\text{media} = Np$$

$$\text{SD} = \sqrt{Np(1-p)}$$

$$A = [0 \quad 1] \quad P[0]=p \quad P[1]=1-p$$

- Gaussiana

$$\text{media} = m$$

$$\text{SD} = \sigma$$

$$f_A(a) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{a-m}{\sigma}\right)^2}$$

- Log-normale

$$f_A(a) = \frac{1}{a\sqrt{2\pi}s} e^{-\frac{1}{2}\left(\frac{\log(a)-m}{s}\right)^2}$$

$$\text{media} = e^{m+s^2/2}$$

$$\text{SD} = e^{2(m+s^2)} - e^{2m+s^2}$$

$$f_A(a) = 1/(a_{\max} - a_{\min}) \quad a_{\min} < a < a_{\max}$$

- Uniforme

$$= 0 \quad \text{altrove}$$

Il concetto di v.a. si può estendere al caso di due (o più) dimensioni, sia per v.a. discrete che continue, se ad ogni evento elementare associamo due o più funzioni.

## STATISTICA INFERENZIALE

Parte dalla stima, ossia scoprire la distribuzione di una popolazione a partire da informazioni contenute in un campione casuale estratto da essa. Ho due alternative:

- È noto (o si ipotizza) il modello probabilistico, es. si tratta di una ddp gaussiana oppure uniformemente distribuita oppure, nel caso discreto, binomiale. Il problema di stima è allora riportato ad un problema di stima dei parametri del modello, cioè  $m$  e  $s$  nel caso gaussiano,  $a_{\min}$  e  $a_{\max}$  nel caso di v.a. uniformemente distribuita,  $p$  nel caso binomiale.
- Non si conosce il modello probabilistico, ovvero non si vogliono fare ipotesi su di esso.

Nel primo caso voglio stimare, ad esempio, la media e la varianza che caratterizzano completamente una va

gaussiana.

$$T = \frac{1}{N} \sum_{i=1}^N A_i$$

è uno stimatore non polarizzato della media;

$$T = \frac{1}{N} \sum_{i=1}^N (A_i - A_m)^2$$

è uno stimatore della varianza polarizzato e consistente

$$T = \frac{1}{N-1} \sum_{i=1}^N (A_i - A_m)^2$$

è uno stimatore della varianza non polarizzato e consistente.

Se la va non è gaussiana di volta in volta si devono definire gli stimatori.

Consideriamo ora il problema di stimare la ddp di una v.a. a partire da un insieme di suoi valori. Il problema è più complesso, perché si tratta di stimare non più un parametro (come media e SD) ma una funzione. Esistono vari approcci.

Metodo dell'istogramma → Si approssima la densità di probabilità da stimare con una funzione continua a tratti. Preso un generico intervallo, di estremi  $\alpha$  e  $\beta$ , indichiamo con  $p_i$  il valore dell'approssimazione

$$\int_{\alpha}^{\beta} f_A(a) da = \text{Prob}(a \in [\alpha, \beta]) = p_i * (\beta - \alpha) \quad p_i = \frac{\text{Prob}(a \in [\alpha, \beta])}{(\beta - \alpha)}$$

nell'intervallo  $[\alpha, \beta]$ .

Ora devo stimare  $\text{Prob}(a \in [\alpha, \beta])$  che chiamo  $P_{\alpha\beta}$ . Rappresenta la probabilità che un valore della variabile aleatoria preso a caso appartenga all'intervallo [alfa, beta]. Considero gli  $N$  dati a disposizione e chiamo  $X$  la va che conta il numero di volte che un elemento cade in [alfa, beta]. Se gli  $N$  dati sono indipendenti  $X$  ha

distribuzione binomiale:  $\text{Prob}[X = k] = \binom{N}{k} P_{\alpha\beta}^k (1 - P_{\alpha\beta})^{N-k}$ ,  $E[X] = NP_{\alpha\beta}$  ovvero  $P_{\alpha\beta} = E[X]/N$ . Quindi per misurare

$P_{\alpha\beta}$  dovremmo conoscere il valore medio della variabile  $X$  ma noi abbiamo a disposizione solo una

realizzazione. Questo ci fornirà una stima della probabilità cercata:  $\hat{P}_{\alpha\beta} = k/N$  e quindi  $\hat{p}_i = \frac{k/N}{(\beta - \alpha)}$ . Se confrontiamo con la definizione di istogramma data in precedenza, la ddp viene stimata dividendo l'istogramma per il numero totale di campioni e l'ampiezza dell'intervallo. La scelta degli intervalli è critica: per avere una buona approssimazione a gradino la differenza tra alfa e beta deve essere piccola ma per avere una buona stima degli  $E[X]$  questa deve essere grande → devo mediare nella scelta dell'intervallo tra queste due esigenze.

## TEST DI IPOTESI- T STUDENT

Test che permettono di prendere decisioni su una qualche caratteristica di una o più popolazioni partendo da campioni delle stesse. Ogni test:

1. Analisi dei dati disponibili → lavora su un certo tipo di dati (qualitativi o quantitativi, continui o discreti)
2. Verifica degli assunti di base → fa delle assunzioni e queste vanno verificate sui dati prima di applicare il test
3. Impostazione dell'ipotesi statistica → formula e impone l'ipotesi che vuole testare
4. Costruzione della statistica → per prendere decisioni riguardo all'ipotesi formulata
5. Determinazione della distribuzione della statistica → se l'ipotesi è corretta, la statistica ha una certa ddp
6. Definizione della regola di decisione → sono le condizioni per cui si accetta o si rifiuta l'ipotesi nulla, sulla base di una probabilità d'errore

Il test di Student serve per testare se due campioni provengono dalla stessa popolazione o da popolazioni diverse.

- 1) lavora su dati quantitativi misurabili su scala continua;
- 2) le assunzioni di base sono che i dati costituiscono due campioni indipendenti, di dimensioni N1 e N2, estratti da popolazioni gaussiane con medie ( $m_1$  e  $m_2$ ) e varianze note ( $s_1^2$  e  $s_2^2$ )
- 3) l'ipotesi statistica è:  $H_0: m_1 = m_2$  o in alternativa  $H_1: m_1 \neq m_2$

$$t = \frac{m_1 - m_2}{s_p \sqrt{\frac{1}{N_1} + \frac{1}{N_2}}}$$

- 4) si costruisce quindi la statistica:  $s_p^2 = \frac{(N_1-1)s_1^2 + (N_2-1)s_2^2}{N_1 + N_2 - 2}$  con  $s_p$  media ponderata tra  $s_1$  e  $s_2$ :

- 5) con ipotesi  $H_0$ : la v.a.  $t$  ha una distribuzione nota, detta distribuzione di Student, la cui forma dipende unicamente dai gradi di libertà  $df = N_1 + N_2 - 2$
- 6) per definire una regola di decisione si fissa un livello di significatività ad (es.  $\alpha = 5\%$ ) e si valuta l'intervallo di confidenza per la v.a.  $t \rightarrow$  Regola di decisione: Accetto l'ipotesi  $H_0$  se  $t$  cade all'interno dell'intervallo di confidenza Rifiuto l'ipotesi  $H_0$  se  $t$  cade al di fuori dell'intervallo di confidenza, perché sono valori poco probabili

con MATLAB  $p = tcdf(t, df)$  che calcola l'integrale da meno infinito a  $t$ . Il complemento a 1 di  $tcdf$  dà la

#### **DECISIONE**

$H_0$  assunta       $H_0$  rifiutata ?

probabilità che, sotto  $H_0$ , la variabile di Student sia maggiore del valore trovato  $t$ . Questa probabilità viene spesso chiamata *p-value*. Se assume valori bassi, l'ipotesi si rifiuta.

		$H_0$ vera	$H_0$ falsa ?
REALTA'	$H_0$ vera	VERI NEGATIVI	FALSI POSITIVI
	$H_0$ falsa ?	FALSI NEGATIVI	VERI POSITIVI
			Errore di tipo I
			Errore di tipo II

## 4. SEGNALI ALEATORI

I processi aleatori sono i modelli adatti a descrivere esperimenti casuali in cui quello che si osserva è una quantità che evolve nel tempo (segnaletico). È una funzione di due variabili,  $X(t;w)$ , il tempo ed un evento casuale.

Ad esempio, ogni volta che la misura di ECG viene ripetuta sullo stesso soggetto o su un diverso soggetto, alcuni fattori non controllabili e non noti fanno sì che tale misura abbia caratteristiche diverse (random).

Può essere allora utile rinunciare a descrivere l'esatta natura temporale del segnale, limitarsi ad una descrizione che evidenzia le proprietà consistenti, da un segnale all'altro.

Quando fissiamo il valore di  $w=w_1$  otteniamo un segnale deterministico detto segnale campione o realizzazione del processo. Se invece fissiamo  $t=t_1$ , quello che otteniamo è una v.a. Fissati due istanti distinti, si ottengono due distinte v.a., ovvero il vettore aleatorio  $X = [x(t_1, \omega) \ x(t_2, \omega)]$ .

Quindi, riassumendo,  $X(t, w)$  può rappresentare

- un insieme di funzioni delle variabili  $w$  e  $t$
- una funzione deterministica della variabile  $t$  detta funzione campione del processo ( $t$  variabile,  $w$  fissato)
- una variabile aleatoria ( $t$  fissato,  $w$  variabile)
- una n-upla di variabili aleatorie (fissati  $n$  valori di  $t$ ,  $w$  variabile)
- un numero reale ( $t, w$  fissati)

Se vogliamo utilizzare una descrizione parametrica di un processo lo specifichiamo mediante una funzione del tempo ed una v.a. e con la descrizione statistica della v.a.

Un P.A. viene spesso descritto in termini statistici, cioè si rinuncia ad una descrizione dettagliata di tutte le sue possibili realizzazioni e ci si accontenta di descrivere alcune caratteristiche da un punto di vista statistico. In genere si procede per gradi, assegnando prima le funzioni di distribuzione delle v.a. unidimensionali estratte in un certo istante di tempo (descrizione del primo ordine) poi le funzioni di distribuzione congiunte relative a tutte le coppie di v.a. unidimensionali (descrizione del secondo ordine) e così di seguito.

Descrizione statistica del PRIMO ORDINE → è relativa al processo in un generico istante  $t$  fissato. Sono date  $F_x(a; t) = P[X(t) \leq a]$

$f_x(a; t) = dF_x(a; t)/da$  che dipendono da  $a$  e da  $t$ , ora ci interessano **Media**  $m_x(t) = E[X(t)]$ , **Potenza statistica**  $P_x(t) = E[X^2(t)] - m_x(t)^2$ , **Varianza**  $\sigma_x^2(t) = E[(X(t) - m_x(t))^2] = P_x(t) - m_x(t)^2$ .

Tutti e tre sono segnali deterministici e se si dispone di una descrizione parametrica è facile ottenere media e varianza usando ad esempio il teorema dell'aspettazione. Se invece si dispone solo di un numero finito di  $N$  realizzazioni, si possono ottenere delle stime di queste funzioni, estendendo ad ogni tempo  $t$  quanto

visto per la stima di media e varianza di una v.a. a partire dalla misura di un duo campione:  $\hat{m}(t) = \frac{1}{N} \sum_{i=1}^N x_i(t)$ .

Descrizione statistica del SECONDO ORDINE → è relativa al processo in 2 generici istanti  $t_1$  e  $t_2$ , quindi è la descrizione statistica di 2 v.a.  $X(t_1, w)$  e  $X(t_2, w)$ . La funzione di distribuzione di probabilità e la ddp sono:

$$F_x(a_1, a_2; t_1, t_2) = \text{prob}[X(t_1) < a_1, X(t_2) < a_2]$$

$$f_x(a_1, a_2; t_1, t_2) = \frac{\partial^2 F_x(a_1, a_2; t_1, t_2)}{\partial a_1 \partial a_2}$$
 che dipendono da  $t_1$  e  $t_2$ .

Ci interessano:

**Autocorrelazione**  $R_x(t_1, t_2) = E[X(t_1)X(t_2)]$ , che è il valor medio del prodotto tra i valori assunti da ogni realizzazione ai tempi  $t_1$  e  $t_2$

$$\begin{aligned} \text{Autocovarianza } C_x(t_1, t_2) &= E[(X(t_1) - m_x(t_1))(X(t_2) - m_x(t_2))] \\ &= R_x(t_1, t_2) - m_x(t_1)m_x(t_2) \end{aligned}$$

Sono misure della "memoria" dei segnali, cioè di quanto il processo all'istante  $t_1$  è legato al valore che esso assume all'istante  $t_2$ . Se non c'è alcun legame tra i valori agli istanti  $t_1$  e  $t_2$  la autocovarianza vale 0.

-Descrizione statistica DI POTENZA → è una descrizione statistica parziale, basata sulle funzioni:

**Valor medio  $m_x(t) = E[X(t)]$** , **Autocorrelazione  $R_x(t_1, t_2) = E[X(t_1)X(t_2)]$** , o in alternativa autocovarianza  $C_x(t_1, t_2)$ .

Se disponiamo di un numero finito  $N$  di realizzazioni, si possono ottenere delle stime dell'autocorrelazione:

$$\hat{R}_x(t_1, t_2) = \frac{1}{N} \sum_{i=1}^N x_i(t_1)x_i(t_2)$$

In generale, la descrizione statistica di un p.a. dipende dal tempo, ma esistono in molti fenomeni aleatori alcuni parametri che risultano indipendenti dal tempo → stazionarietà

**DEF: Un processo aleatorio si dice stazionario in senso stretto se il suo comportamento statistico è invariante rispetto ad una traslazione dell'origine dei tempi.**

Si dice stazionario in senso lato o debolmente stazionario se l'invarianza rispetto ad una traslazione vale per la descrizione in potenza, pertanto: **Valor medio  $m_x(t) = E[X(t,A)] = m$**  → costante nel tempo

$$\begin{aligned} \text{Autocorrelazione } R_x(t_1, t_2) &= E[X(t_1) X(t_2)] \\ &= E[X(t_1) X(t_1 + \tau)] \\ &= R_x(\tau) \end{aligned} \quad \rightarrow \text{dipende da } \tau = t_2 - t_1$$

In un processo stazionario in senso lato, la funzione autocorrelazione  $R_x(\tau)$ , gode di importanti proprietà:

- 1) È una funzione reale pari  $R_x(\tau) = E[X(t) X(t+\tau)] = E[X(t-\tau)X(t)] = E[X(t)X(t-\tau)] = R_x(-\tau)$
- 2) Ha valore massimo nell'origine  $E[(X(t)+X(t+\tau))^2] = E[X(t)^2] + E[X(t+\tau)^2] + 2E[X(t)X(t-\tau)] = 2R_x(0) + 2R_x(\tau) \geq 0$   
da cui  $R_x(0) \geq |R_x(\tau)|$
- 3) Nell'origine coincide con la potenza statistica  $R_x(0) = E[X^2(t)] = P_x \geq 0$

**SIGNIFICATO DI AUTOCORRELAZIONE:** L'autocorrelazione è una misura di quanto in media sono collegati tra loro valori che distano di un tempo  $t$ . Più rapidamente va a zero la funzione di autocorrelazione, più blando è questo legame e quindi più "irregolari" sono le realizzazioni.

$$R_x(\tau) = C_x(\tau) = \sigma^2 \quad \begin{cases} \tau = 0 \\ \neq 0 \end{cases}$$

Se  $R_x(\tau) = \text{costante}$  abbiamo un processo stazionario (es  $X(t) = A$ ). Se invece  $R_x(\tau) \neq 0$  altrove non c'è alcun legame tra i valori in tempi diversi e si ha rumore bianco.

Se il processo è stazionario, il problema è più semplice:

- La media del processo è una costante
- La funzione di autocorrelazione dipende solo da  $\tau$

Però in teoria servono ancora tutte le possibili realizzazioni per ottenere media e autocorrelazione, o quanto meno un numero elevato di realizzazioni per ottenere una buona stima

**DEF: Un processo stazionario si dice ergodico se le sue proprietà statistiche possono essere ricavate da una sua qualsiasi realizzazione.**

Ogni realizzazione di un processo ergodico contiene in sé tutta l'informazione possibile sul processo in quanto una sorgente ergodica produce, nel corso di una realizzazione, tutte le situazioni ed i casi possibili per il processo con una frequenza pari alla probabilità di detti eventi. La media (temporale) di ogni

$$m_x = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} x_i(t) dt$$

$$R_x(\tau) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} x_i(t)x_i(t+\tau) dt$$

realizzazione coincide con il valor medio del processo. L'ergodicità ci permette di inferire sulle proprietà dell'intero processo, dall'unica realizzazione che spesso possiamo misurare.

## 5. ESTRAZIONE POTENZIALI EVOCATI

Il potenziale evocato (PE) non si può misurare separatamente dal segnale EEG di fondo. Se ipotizziamo un modello additivo, il segnale misurabile è espresso come somma di due componenti:  $y(t) = u(t) + v(t)$

$$y(t) = u(t) + v(t)$$

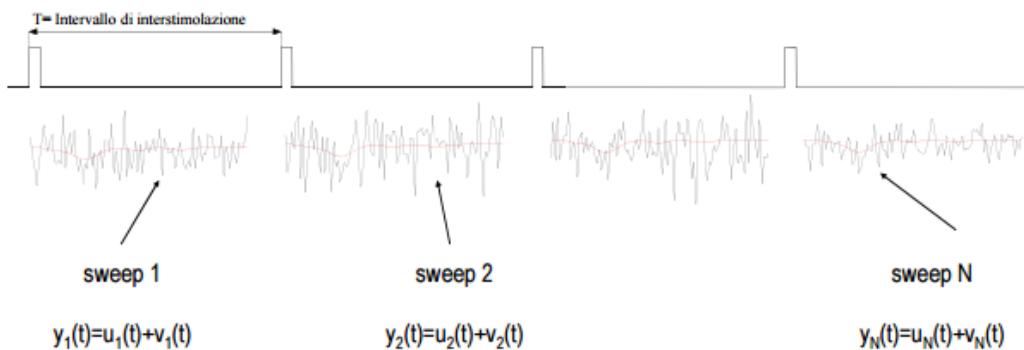
↑

segnale utile      Rumore EEG  
PE                  + errori di misura  
                        + artefatti

L'EEG di fondo ha ampiezze di qualche decina di mV.

Il PE, a seconda del tipo e della distanza del sito nervoso che lo genera dall'elettrodo di misura, ha ampiezze massime di ordine variabile tra le decine di  $\mu$ V e il mV.

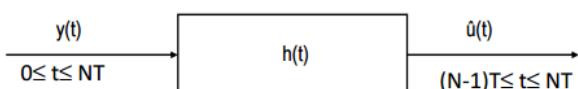
Quello che si fa è proporre al soggetto N stimoli identici e generalmente equispaziati nel tempo. Dopo ogni stimolo viene misurata una sweep (o epoca) di segnale, che contiene sia potenziale evocato che rumore EEG (attività spontanea).



$$\hat{u}(t) = \frac{1}{N} \sum_{i=1}^N y_i(t)$$

Una stima del potenziale evocato è data dalla media coerente dello sweep:  $\frac{1}{N} \sum_{i=1}^N v_i(t)$ . Aumentando N la stima migliora; questo perché le  $v_i(t)$  sono scorrelate dallo stimolo e la loro media è 0.

Esiste se no un filtro COMB (a pettine) che lascia passare inalterate le componenti a frequenza  $K/T$  mentre le altre componenti vengono smorzate in proporzione a  $N$



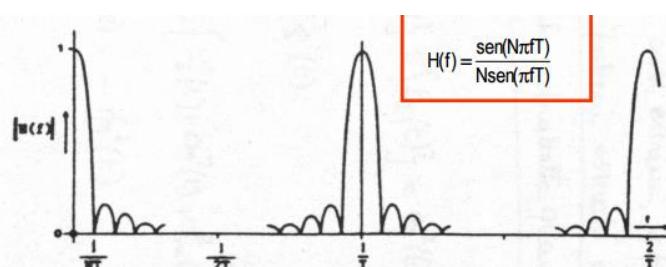
$$\hat{u}(t) = \frac{1}{N} \sum_{i=0}^{N-1} y(t - iT)$$

$$\hat{u}(t) = \frac{1}{N} \sum_{i=0}^{N-1} y(t) \otimes \delta(t - iT) = y(t) \otimes \frac{1}{N} \sum_{i=0}^{N-1} \delta(t - iT)$$

$$\hat{u}(t) = y(t) \otimes h(t)$$

$$h(t) = \frac{1}{N} \sum_{i=1}^{N-1} \delta(t - iT)$$

#### Filtro COMB (a pettine)



**Figure 4** The modulus of the transfer function of the equivalent filter

- il filtro ha N zeri tra 0 e  $1/T$
  - In  $y(t)$  la componente utile è periodica di periodo  $T$ , quindi rappresentabile nel dominio della freq con componenti a  $1/T$ =armonica fondamentale e ai suoi multipli  $k/T$ ,  $k=0,1,2,\dots$ .
  - Tali frequenze passano inalterate, in quanto  $H(k/T)=1$
  - Le frequenze diverse dai multipli della fondamentale vengono attenuate, tanto più quanto più è elevato N
  - Bisogna evitare che  $1/T$  e i suoi multipli siano vicini alla frequenza di alimentazione di 50Hz e ai suoi multipli

Questo stimatore necessita di alcune ipotesi:

la risposta evocata non varia :  $u_i(t) \equiv u(t) \forall i$

$$i) \hat{u}(t) = \frac{1}{N} \sum_{i=1}^N y_i(t) = \frac{1}{N} \sum_{i=1}^N u_i(t) + \frac{1}{N} \sum_{i=1}^N v_i(t) = u(t) + \frac{1}{N} \sum_{i=1}^N v_i(t)$$

$$ii) \text{ L'EEG di fondo è un processo a media nulla} \\ E[\hat{u}(t)] = u(t) + \frac{1}{N} \sum_{i=1}^N E[v_i(t)] = u(t)$$

iii)

EEG di fondo e' un processo *stazionario*, con varianza pari a  $\sigma_v^2$  indipendente da sweep a sweep ( $v_i$  indip. da  $v_j$ ), cioè  $v_1(t), v_2(t), \dots, v_N(t)$  sono N realizzazioni indipendenti di uno stesso PA

$$\text{var}[\hat{u}(t)] = E[(\hat{u}(t) - u(t))^2] = E\left[\left(\frac{1}{N} \sum_{i=1}^N v_i(t)\right)^2\right] = \\ \frac{1}{N^2} \sum_{i=1}^N E[v_i(t)^2] + \frac{1}{N^2} \sum_{i \neq j} \sum_{j=1}^N E[v_i(t)v_j(t)] = \frac{1}{N^2} \cdot N \cdot \sigma_v^2 = \frac{1}{N} \sigma_v^2$$

Valutare la varianza dello stimatore serve:

1) per dimensionare il numero N di repliche, necessarie per ridurre la varianza del rumore a livelli desiderati.

Se invece il potenziale evocato varia tra -10 e 10  $\mu\text{V}$

Il potenziale evocato varia tra -20 e 20  $\mu\text{V}$

almeno 89% dei valori dell'EEG sono compresi tra -60 e 60

La varianza dell'EEG è pari a 400  $\mu\text{V}^2$

$$\text{var}[\hat{u}(t)] = \frac{\sigma_v^2}{N} = \frac{400}{N} \approx 1 \rightarrow N=400$$

durata indicativa dell'esperimento: circa 7min

Dimensionare il numero N di repliche affinché all'uscita del sistema di averaging la SD del rumore sia ridotta a non più del 10% del valore massimo del segnale

Dovrà essere  $\text{var}[\hat{u}(t)] = \frac{\sigma_v^2}{N} = \frac{400}{N} \approx 4 \rightarrow N=100$   
durata indicativa dell'esperimento: 100 sec

se varia tra -1 e 1  $\mu\text{V}$

$$\text{var}[\hat{u}(t)] = \frac{\sigma_v^2}{N} = \frac{400}{N} \approx 0.01 \rightarrow N=40000$$

durata indicativa dell'esperimento: circa 11h

2) A calcolare gli intervalli di confidenza della stima, tramite la diseguaglianza di Chebychev,

$$k=2 \quad \text{Prob}\left[\hat{u}(t) - 2 \frac{\sigma_v}{\sqrt{N}} \leq u(t) \leq \hat{u}(t) + 2 \frac{\sigma_v}{\sqrt{N}}\right] \geq 75\%$$

$$k=3 \quad \text{Prob}\left[\hat{u}(t) - 3 \frac{\sigma_v}{\sqrt{N}} \leq u(t) \leq \hat{u}(t) + 3 \frac{\sigma_v}{\sqrt{N}}\right] \geq 89\%$$

Ad es. per k=3, si può dire con almeno 89% di probabilità che il potenziale evocato cade nella fascia compresa tra il valore stimato e +/-3 volte la deviazione standard della stima, che è pari alla deviazione standard del rumore EEG diviso per la radice del numero di repliche.

Le buone proprietà dello stimatore di averaging valgono se sono soddisfatte le ipotesi 1-2-3. L'ipotesi più critica è la prima, che è soddisfatta solo in prima approssimazione in quanto in realtà il potenziale evocato varia in relazione alle variazioni del sistema sensoriale in esame, per meccanismi di adattamento, abitudine, calo dell'attenzione, affaticamento. Questi problemi sono particolarmente evidenti in esperimenti con N elevato. D'altra parte N elevato garantisce migliori proprietà della stima.

Rimuoviamo l'ipotesi 1 e rivalutiamo l'operazione di averaging.

-Ipotesi 1bis: ogni risposta evocata  $u_i(t)$  è un P.A. (non stazionario!)  $u(t)$ , indipendente dal rumore.

$$\hat{u}(t) = \frac{1}{N} \sum_{i=1}^N y_i(t) = \frac{1}{N} \sum_{i=1}^N u_i(t) + \frac{1}{N} \sum_{i=1}^N v_i(t) = u(t) + \frac{1}{N} \sum_{i=1}^N v_i(t)$$

-Ipotesi 2bis: EEG è a media nulla,  $u(t)$  ha media pari a  $m(t)$

$$E[\hat{u}(t)] = E[u(t)] + \frac{1}{N} \sum_{i=1}^N E[v_i(t)] = m(t) \quad \text{Lo stimatore converge in media verso il valor medio del processo}$$

Stimiamo allora la media del processo. Consideriamo 2 casi limite:

- 1°caso limite: le risposte evocate hanno la stessa forma, ma variano di un fattore di scala  $u(t) = Au^\circ(t)$  con A v.a. e  $u^\circ(t)$  deterministico  $m(t) = E[u(t)] = E[A]u^\circ(t) \rightarrow$  La media del processo ha la stessa forma di tutte le realizzazioni, quindi ad es. i tempi di occorrenza dei picchi e la loro durata vengono correttamente stimati

- 2°caso limite: le risposte evocate hanno la stessa forma, ma differiscono nella latenza

Teorema dell'aspettazione

$u(t) = u^*(t - \tau)$  con  $\tau$  v.a. con ddp pari a  $f(\tau)$ .  
 $m(t) = E[u^*(t - \tau)] = \int_{-\infty}^{+\infty} f(\tau)u^*(t - \tau)d\tau$  → La media del processo ha una forma diversa dalla forma di tutti i potenziali evocati, perché è pari al prodotto di convoluzione tra la forma del singolo potenziale evocato e la ddp della latenza. Questo rende la situazione più critica

Se le risposte evocate sono tutte uguali, lo stimatore basato sulla media coerente tende in media verso questa comune risposta, con una varianza che scende a zero al crescere del numero di risposte. Questo vuol dire che aumentando il numero delle risposte è possibile stimare correttamente la risposta evocata. Se le risposte evocate sono diverse, lo stimatore basato sulla media coerente tende in media verso il valor medio delle risposte. Variazioni di scala tra le risposte sono meno critiche, in quanto la media recupera la forma comune a tutte le risposte. Variazioni di latenza sono più critiche, perché la media non coincide con la forma di nessuna risposta.

## 7. MODELLI ARMA PER EEG

I modelli arma vengono impiegati per l'analisi di segnali biologici a vari scopi. Consideriamo l'EEG come esempio di segnale e vediamo alcune possibili applicazioni:

- **SIMULAZIONE:** Applicando come ingresso ad un modello di ordine e con coefficienti adeguati una sequenza di rumore bianco si ottiene in uscita un segnale EEG indistinguibile dall'EEG originale. (Esaminando le coppie di segnali, non ci si deve aspettare una corrispondenza punto a punto, ma una somiglianza delle caratteristiche statistiche, in quanto i due segnali vanno considerati come realizzazioni diverse dello stesso processo stocastico. Si è però verificato che un neurologo formula la stessa diagnosi esaminando il segnale vero oppure quello simulato.)
- **COMPRESSEIONE:** è un'esigenza per registrazioni con lunghi periodi di osservazione. Dato che il modello AR permette di ricostruire un segnale simile all'originario alimentando il modello con un rumore bianco di potenza opportuna, si può pensare di memorizzare i parametri del modello AR che approssima un tratto di EEG, quindi usare il modello AR come compressore di informazioni. Si può procedere in due modi differenti:
  - Si memorizzano i parametri del modello (coefficienti ai e varianza del rumore di ingresso). Quando necessario, si ricostruisce un segnale simile a quello originario alimentando il modello con una sequenza di rumore bianco. Si realizzano buoni rapporti di compressione (40:1). È un metodo di compressione Lossy
  - Si memorizzano non solo i parametri del modello ma anche l'errore di predizione. Dato che la varianza di tale errore è comunque minore della varianza del segnale si ha una riduzione dell'ingombro di memoria. Il rapporto di compressione è di 2:1. Concettualmente questa tecnica decomponete i valori campionati in una parte correlata con i valori passati (combinazione lineare) e in una parte da essi scorrelata (errore di predizione). Poiché il segnale originario può essere ricostruito dalla sequenza degli errori di predizione, questa sequenza è una rappresentazione equivalente del segnale originario (compressione senza perdita di informazione) e sono sufficienti modelli di ordine basso (es.2) per avere buone prestazioni
- **SEGMENTAZIONE:** un segnale EEG è stazionario solo per intervalli di pochi secondi in quanto è influenzato dello stato metabolico, fisiologico e psicologico del soggetto. La segmentazione serve per dividere il segnale in segmenti stazionari. Un possibile approccio consiste nel dividere il segnale in segmenti sufficientemente brevi in modo tale da poter assumere che il segnale sia stazionario all'interno di ogni segmento. Ad esempio, nel caso di EEG, si segmenta il segnale in intervalli di 2 secondi, perché è noto che la stazionarietà è ragionevolmente soddisfatta in tale periodo.  
Un approccio diverso è invece quello della segmentazione adattativa, che consiste di suddividere il segnale in intervalli di durata variabile, all'interno dei quali valga l'ipotesi di stazionarietà. Si parla di "fixed window". E'allora necessario disporre di un criterio che consenta di capire quando, in un segmento di segnale di durata via via crescente, l'ipotesi non è più soddisfatta. I passi per ottenere una segmentazione adattiva sono:
  - a) a partire da un segmento di segnale di durata tale per cui la stazionarietà sia soddisfatta, (fixed window, es. t=0-2 sec) si mette a punto un modello ARMA.
  - b) Si fa passare una sequenza di dati successivi a tale intervallo (moving window, es. t=2-4sec) attraverso il filtro inverso. Si analizzano le caratteristiche della sequenza in uscita al filtro inverso. Se tale sequenza è ancora è una realizzazione di rumore bianco, a valor medio nullo e con una varianza simile a quella stimata, l'intervallo di stazionarietà può essere esteso. Si ripete l'analisi sull'intervallo successivo (es. t=3-4sec.)
  - c-d) Se la sequenza in uscita al filtro inverso non soddisfa più le ipotesi di partenza, allora l'intervallo di stazionarietà va ristretto all'intervallo precedente
  - e) Si ripete da a), con una nuova fixed window.

- RICONOSCIMENTO DI TRANSITORI: si utilizzano i modelli ARMA per riconoscere la presenza, all'interno del segnale EEG, di transitori (spikes), che sono indice di patologia.

Ipotesi:

$$x(k) = \sum_i s_i(k - k_i) + n(k)$$

EEG misurato

Spikes, ai tempi incogniti  $k_i$

EEG di fondo

Analisi:

si parte stimando da un segmento di dati i parametri del filtro AR (parametri del modello che descrive  $n$ ).

si fa poi passare il segnale da analizzare  $x(k)$

attraverso il filtro inverso, ottenendo in uscita

$$y(k) = x(k) + \sum_{i=1}^p a_i x(k-i)$$

Si considera poi un segmento del segnale  $y(k)$ , relativo ad un certo intervallo di tempo: se si tratta

di rumore bianco con caratteristiche simili a  $u(k)$ , allora in questo intervallo non c'è uno spike, in caso contrario viene riconosciuta la presenza di uno spike.

Per individuare con una certa precisione i tempi di occorrenza degli spikes, l'intervallo deve essere breve, quindi è necessario un test locale di verifica delle caratteristiche del rumore. Si usa spesso un test sul valore del residuo, o sul suo quadrato, ad es. si riconosce uno spike se  $|y_n| > k \sigma_u$ . Per poter applicare questo metodo è necessario che le forme d'onda da riconoscere siano poco frequenti, in modo che nella fase di stima parametrica il sistema stimi i parametri del rumore di fondo  $n(k)$  e non gli eventi da localizzare. Questa condizione è ben soddisfatta nel caso ad es. di ricerca di spikes epilettici, situazione in cui il metodo è stato applicato con successo.

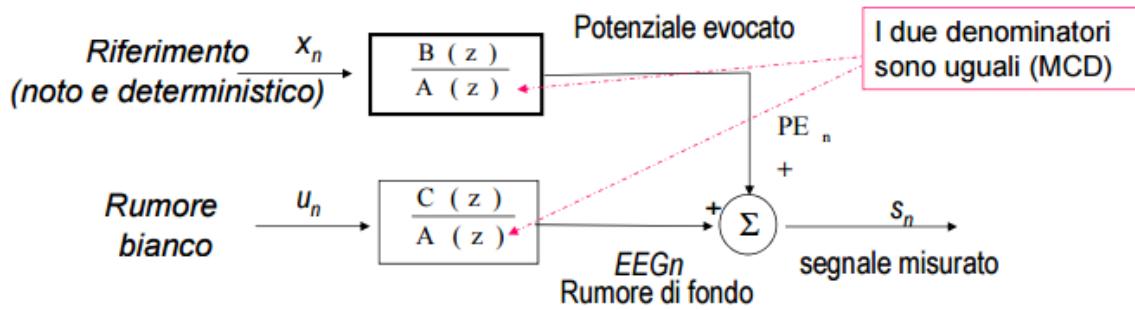
## IMPIEGO DI MODELLI ARMAX PER L'ESTRAZIONE DI POTENZIALI EVOCATI SINGLE SWEEP

Problemi:

A.) se il potenziale evocato varia in relazione alle variazioni del sistema sensoriale in esame (es. adattamento, abitudine, stato di attenzione, ...) si ottiene solo una risposta evocata media

B.) In teoria è bene mediare un grande numero di sweep, talvolta nella pratica ciò non è possibile (es. stimoli frequenti vs stimoli rari)

→ estrazione della single sweep con modello ARMAX



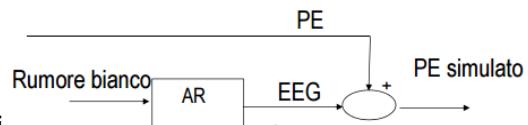
Per ogni sweep si formula un modello ARMAX

Per semplificare il procedimento di stima si suppone  $C(z)=1$

$$\begin{aligned} s_n &= -\sum_1^p a_i s_{n-i} + \sum_0^q b_i x_{n-i} + u_n \\ \hat{s}_n &= -\sum_1^p a_i s_{n-i} + \sum_0^q b_i x_{n-i} \\ J &= E[e_n^2] = [(s_n - \hat{s}_n)]^2 \\ a_i, b_i, \sigma_e^2 &: \min J \end{aligned}$$

Si ottiene un predittore lineare nei parametri  $a_i$  e  $b_i$ . Minimizzare la funzione costo porta a delle equazioni lineari, più complesse delle equazioni di Yule Walker ma comunque risolvibili analiticamente. Noti  $a_i$  e  $b_i$  si calcola PE single sweep come uscita del filtro  $B(z)/A(z)$  quando in ingresso c'è il riferimento  $x_n$ .

Per la scelta del riferimento si utilizzano riferimenti di popolazione o de media su tutte le sweep a disposizione (metodo off-line) o sulle sweep acquisite fino a quell'istante (metodo online). Per la scelta degli ordini del modello si utilizzano indici di parsimonia. La bianchezza dell'errore di predizione si valuta con il test di Anderson.



Questa tecnica si può validare applicandola a dati simulati

$$PE = \sum_{i=1}^4 A_i e^{-\frac{1}{2} \left( \frac{t-L_i}{\sigma_i} \right)^2}$$

PE è la combinazione lineare di 4 gaussiane

con ampiezza  $A_i$  e latenza  $L_i$  v.a. gaussiane

si ottengono risultati che validano la tecnica quando le sweep simulate e sstimate hanno media e SD simili.

Una possibile applicazione a dati reali è nello studio dell'attenzione (=focalizzare l'attività nervosa su uno stimolo) e della memoria prospettica (=ricordare azioni pianificate nel passato)

## 8. ANALISI SPETTRALE

La rappresentazione di segnali biologici nel dominio della frequenza è una valida alternativa alla rappresentazione usuale nel dominio del tempo. Un segnale potrebbe essere complesso da interpretare nel dominio del tempo mentre potrebbe avere uno spettro di più facile interpretazione anche perché lo spettro di un processo aleatorio è un segnale deterministico.

SEGNALI DETERMINISTICI:

$$E = \int_{-\infty}^{+\infty} x^2(t) dt$$

Dato un segnale deterministico continuo  $x(t)$  si definisce Energia del segnale:

$$E = \int_{-\infty}^{+\infty} x^2(t) dt = \frac{1}{2\pi} \int_{-\infty}^{+\infty} |X(\Omega)|^2 d\Omega$$

Dal teorema di Parseval l'energia si può anche definire a partire da  $X(\Omega) = FT[x(t)]$

$$S(\Omega) = \frac{1}{2\pi} |X(\Omega)|^2$$

Tale relazione porta definire la densità spettrale di energia

$S(\Omega)$  è una funzione che, integrata in una certo intervallo  $\Omega_1 - \Omega_2$  fornisce l'energia contenuta in  $x(t)$  e associata a tale banda.

Non tutti i segnali hanno energia finita. Per segnali ad energia infinita, conviene considerare la

$$P = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T x^2(t) dt \quad P(\Omega) = \frac{1}{2\pi} \lim_{T \rightarrow \infty} \frac{1}{2T} \left| \int_{-T}^{+T} x(t) e^{-j\Omega t} dt \right|^2$$

potenza e la densità spettrale di potenza

$P(\Omega)$  è una funzione che, integrata in un certo intervallo  $\Omega_1 - \Omega_2$  fornisce la potenza contenuta in  $x(t)$  e associata a tale banda.

Dalle definizioni, appare chiaro che se  $x(t)$  ha energia finita, la sua potenza è nulla. Viceversa, se la potenza è finita e diversa da zero, l'energia è infinita. Quindi in pratica o si è interessati all'energia, o si è interessati alla potenza, ma non ad energia e potenza contemporaneamente.

Consideriamo processi stazionari (è sufficiente la stazionarietà in senso lato). Le realizzazioni sono generalmente ad energia infinita, quindi si parla di spettro di potenza. Esiste una vasta letteratura, e si possono individuare due classi di metodi per stimare la  $P(\Omega)$ :

- Metodi non parametrici, o tradizionali, basati sulla FT
- Metodi parametrici, basati sull'impiego di modelli

Il Teorema di Wiener permette di definire  $P(\Omega)$  sulla base della  $f$  di autocorrelazione  $P(\Omega) = FT[R_x(\tau)]$ .

$$P(\Omega) = \lim_{T \rightarrow \infty} E \left[ \frac{1}{T} \left| \int_{-T/2}^{T/2} x(t) e^{-j\Omega t} dt \right|^2 \right]$$

NB ricompare l'aspettazione, perché le proprietà di ergodicità consentono di stimare l'autocorrelazione da una realizzazione, ma non lo spettro

*Densità spettrale di potenza  
di una singola realizzazione!*

METODO FT INDIRETTO (dati  $\rightarrow$  autocorrelazione  $\rightarrow$  spettro)

In pratica si hanno a disposizione  $N$  campioni  $x_0, \dots, x_{N-1}$  dai quali si può ottenere solo una stima di  $R_x(m)$  per alcuni valori di  $m$ :  $\hat{R}_x(m) \quad m=0, 1, \dots, M$ , si sfrutta la parità di  $R_x$  per estendere la stima ai valori negativi di  $m$  e si impone  $\hat{R}_x(m) = 0 \quad |m| > M$ .

$$\hat{R}_x(m) = \frac{1}{N-m} \sum_{n=0}^{N-1-m} x(n)x(n+m)$$

- 1° stimatore di  $R_x(m)$   $m > 0$ :

È non polarizzato, cioè  $E[\hat{R}_x(m)] = R_x(m)$  converge in media al valor vero

È consistente,  $\text{var}[\hat{R}_x(m)] \approx \frac{N}{(N-m)^2} [\dots]$ , la varianza tende a zero quando N tende a infinito. Ma, per N finito, la sua varianza cresce al crescere di m

- 2° stimatore di  $R_x(m)$   $m > 0$ : 
$$\hat{R}'_x(m) = \frac{1}{N} \sum_{n=0}^{N-1-m} x(n)x(n+m) = \frac{N-m}{N} \hat{R}_x(m)$$

È polarizzato  $E[\hat{R}'_x(m)] = \frac{N-m}{N} R_x(m)$ , ma asintoticamente non polarizzato

È consistente  $\text{var}[\hat{R}'_x(m)] \approx \frac{1}{N} [\dots]$ , la varianza tende a zero quando N tende ad infinito ma, per N finito, la varianza non cresce al crescere di m ma cresce l'errore di polarizzazione.

Abbiamo quindi 2 stimatori di  $P(w)$ :

$$\hat{P}_{BT}(\omega) = \text{FT}[\hat{R}_x(m)] = \sum_{m=-M}^M \hat{R}_x(m) e^{-j\omega m}$$

- 1° stimatore di  $P(w)$ :

È polarizzato  $E[\hat{P}_{BT}(\omega)] = \sum_{m=-M}^M E[\hat{R}_x(m)e^{-j\omega m}] = \sum_{m=-M}^M R_x(m)e^{-j\omega m} \neq P(\omega) = \sum_{m=-\infty}^{+\infty} R_x(m)e^{-j\omega m}$

Per quantificare l'errore di polarizzazione

$$E[\hat{P}_{BT}(\omega)] = \sum_{m=-\infty}^{+\infty} R_x(m)w(m)e^{-j\omega m} = \text{FT}[R_x(m)w(m)] = P(\omega) \otimes W(\omega)$$

$\rightarrow w(m)=1 \quad |m| \leq M; \quad 0 \quad |m| > M;$

Lo stimatore converge in media verso una versione distorta dello spettro.

$$\hat{P}'_{BT}(\omega) = \text{FT}[\hat{R}'_x(m)] = \sum_{m=-M}^M \hat{R}'_x(m) e^{-j\omega m}$$

- 2° stimatore di  $P(w)$ :

È polarizzato:  $E[\hat{P}'_{BT}(\omega)] = \sum_{m=-M}^M E[\hat{R}'_x(m)e^{-j\omega m}] = \sum_{m=-M}^M \frac{N-|m|}{N} R_x(m)e^{-j\omega m} \neq P(\omega) = \sum_{m=-\infty}^{+\infty} R_x(m)e^{-j\omega m}$

Per quantificare l'errore di polarizzazione:

$$E[\hat{P}'_{BT}(\omega)] = \sum_{m=-\infty}^{+\infty} R_x(m)w'(m)e^{-j\omega m} = \text{FT}[R_x(m)w'(m)] = P(\omega) \otimes W'(\omega)$$

$$\rightarrow w'(m) = (N-|m|)/N \quad |m| \leq M; \quad 0 \quad |m| > M$$

Lo stimatore converge in media verso una versione distorta dello spettro

## METODO FT DIRETTO (dati $\rightarrow$ spettro)

La formula corretta implica l'aspettazione e il passaggio al limite. Dato che si ha a disposizione un numero finito di campioni di una realizzazione, si omettono queste operazioni e si definisce lo stimatore (periodogramma):

$$\hat{P}_{PER}(\omega) = \frac{1}{N} \left| \sum_{n=0}^{N-1} x(n) e^{-j\omega n} \right|^2 = \frac{1}{N} |X(\omega)|^2$$

Si può dimostrare che  $\hat{P}_{PER}(\omega) = \hat{P}'_{BT}(\omega)$  con  $M=N-1$ .

Quindi ancora abbiamo una versione distorta dello spettro. Il periodogramma comporta gli stessi problemi di stima dei metodi precedenti.

## Aspetti implementativi :

### 1. Effetto della frequenza di campionamento

Supponiamo che il segnale discreto  $x(n)$  sia ottenuto per campionamento di un segnale continuo  $x(t)$  con passo  $T$  e frequenza di campionamento  $F_c=1/T$ . Quindi  $R_x(m)$  è la versione campionata di  $R_x(\tau)$ . Come gioca la frequenza di campionamento?

**Funzione di autocorrelazione continua**

$$R_x(\tau) \quad P(\Omega) = FT[R_x(\tau)] = \int_{-\tau_{\max}}^{\tau_{\max}} R_x(\tau) e^{-j\Omega\tau} d\tau$$

**Funzione di autocorrelazione discreta**

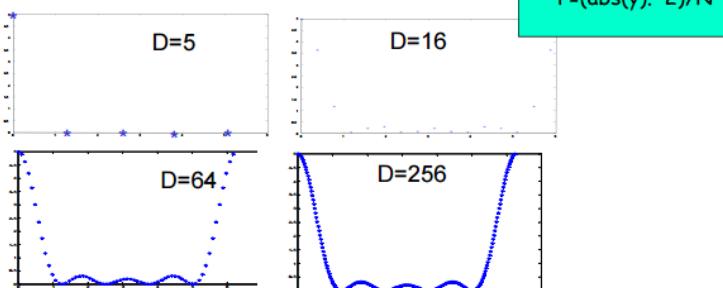
$$R_x(m) \quad P(\omega) = FT[R_x(m)] = \sum_{m=-M}^{M} R_x(m) e^{-j\omega m}$$

$P(\omega)$  è una funzione continua in  $\omega$  e periodica di periodo  $2\pi$ . Si può dimostrare (cfr le relazioni tra FT di un segnale continuo e FT del segnale discreto ottenuto per campionamento) che l'intervallo 0-2 $\pi$  di periodicità di  $P(\omega)$  corrisponde in frequenza ad un intervallo pari alla frequenza di campionamento!!!!

### Aspetti implementativi: 3. Zero padding

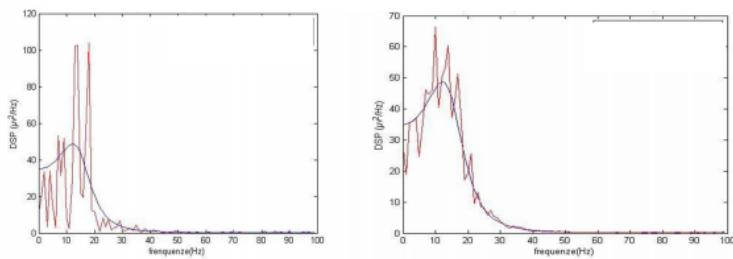
Consiste nell'aggiungere in coda alla sequenza un certo numero di zeri, per simulare un numero maggiore  $D$  (per motivi di efficienza dell'algoritmo è bene che sia una potenza di 2) di campioni. La DFT viene calcolata su  $D$  campioni, e quindi produce  $D$  valori, sempre equispaziati in 0-2 $\pi$ . Ciò non modifica la stima, ma fornisce un andamento più regolare dello spettro.

Esempio:  $x = \text{rect}(5)$



### Effetto dell'averaging

L'operazione di averaging regolarizza la stima, come evidenziata nei grafici sotto riportati, relativi ad uno studio di simulazione in cui in blu è rappresentato lo spettro «vero», in rosso la stima FT, a sinistra ottenuta da un unico segmento di 1 sec, a destra mediando gli spettri di 20 segmenti di 1 sec ciascuno.



Bisogna però trovare un compromesso tra tre esigenze contrastanti:

- mantenere  $L$  sufficientemente elevato per limitare i fenomeni di distorsione
- avere  $K$  sufficientemente elevato per ridurre la varianza dello stimatore
- tenere limitato l'intervallo totale  $KxL$  per soddisfare, almeno in prima approssimazione, l'ipotesi di stazionarietà

Il metodo di Welch offre una possibile soluzione basata sull'impiego di segmenti che si sovrappongono parzialmente, spesso del 50%.

### Aspetti implementativi: 2. DFT

La stima dello spettro viene calcolato per punti, via DFT, cioè:

$$\text{MATLAB} \\ P = \text{fft}(R, 2^*M+1)$$

**Metodo indiretto - 1° Stimatore**

$$\hat{P}_{BT}(k) = DFT[\hat{R}_x(m)] = \sum_{m=0}^{2M+1} \hat{R}_x(m) e^{-j\frac{2\pi}{2M+1}mk} = \hat{P}_{BT}(\omega) \Big|_{\omega=\frac{2\pi}{2M+1}k}$$

$K=0, 1, \dots, 2M+1$

**Metodo indiretto - 2° Stimatore**

$$\hat{P}'_{BT}(k) = DFT[\hat{R}'_x(m)] = \sum_{m=0}^{2M+1} \hat{R}'_x(m) e^{-j\frac{2\pi}{2M+1}mk} = \hat{P}'_{BT}(\omega) \Big|_{\omega=\frac{2\pi}{2M+1}k}$$

$$\text{MATLAB} \\ y = \text{fft}(x, N) \\ P = (\text{abs}(y) .^ 2) / N$$

**Periodogramma**

$$\hat{P}_{PER}(k) = \frac{1}{N} |DFT[x(n)]|^2 = \frac{1}{N} \left| \sum_{n=0}^{N-1} x(n) e^{-j\frac{2\pi}{N}nk} \right|^2 = \hat{P}_{PER}(\omega) \Big|_{\omega=\frac{2\pi}{N}k}$$

$$\text{MATLAB} \\ y = \text{fft}(x, N) \\ P = (\text{abs}(y) .^ 2) / N$$

### Aspetti implementativi: 4. Averaging

Consideriamo la stima con il metodo del periodogramma, ma considerazioni analoghe valgono anche per gli altri metodi. Si può dimostrare che se il segnale  $x(n)$  è un p.a. gaussiano, la varianza dello stimatore è data dalla seguente espressione

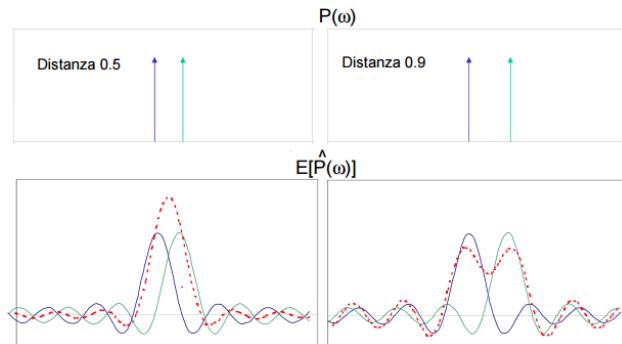
$$\text{var}(\hat{P}_{PER}(\omega)) \approx P_{PER}(\omega)^2 \left[ 1 + \left( \frac{\sin \omega N}{N \sin \omega} \right)^2 \right]$$

Quindi non tende a zero se il numero di campioni  $N$  tende ad infinito, ovvero il periodogramma non produce una stima consistente dello spettro. Per migliorare le proprietà della stima, si usa spesso suddividere i dati a disposizione in  $K$  intervalli ognuno di dimensione  $L$  ( $KL=N$ ) calcolare lo spettro su ciascun intervallo e quindi mediare. Se i vari segmenti sono tra di loro scorrelati, la media produce uno stimatore consistente dello spettro, con varianza che tende a zero quando  $K$  tende all'infinito:

$$\text{var}(\hat{P}_{ave}(\omega)) \approx \frac{1}{K} P_{PER}(\omega)^2 \left[ 1 + \left( \frac{\sin \omega L}{L \sin \omega} \right)^2 \right]$$

Il metodo FT ha però vari problemi:

1. Leakage: per effetto della convoluzione, c'è una perdita (leakage) di potenza verso le frequenze prossime.
2. Risoluzione limitata in frequenza: non si distinguono due righe spettrali che distano meno di metà dell'ampiezza del lobo centrale (es. M=5, 11 lobi di ampiezza 0.57) →
3. Mascheramento di componenti spettrali deboli: una componente spettrale debole può essere mascherata da una componente spettrale forte ad essa vicina
4. Alta sensibilità al rumore di misura



Il problema è che i metodi FT considerano sequenze finite (i dati per il periodogramma, la stima della funzione di autocorrelazione per i BT) e si assume che siano nulle al di fuori della finestra di stima.

### METODI PARAMETRICI

Si può pensare di fare delle ipotesi più ragionevoli, e approssimare la sequenza, fuori della finestra, con la sua estrapolazione fornita da un modello che sia una buona approssimazione del processo.

Procedura a 3 passi

- Selezione del modello
- Stima dei parametri del modello
- Calcolo dello spettro del modello



Nel caso di modelli ARMA lo spettro è facile da calcolare ed è una funzione continua di w:

$$\hat{P}_{\text{ARMA}}(\omega) = P_y(\omega) = |H(\omega)|^2 P_u(\omega)$$

$$H(z)|_{z=e^{j\omega}} = \frac{1 + \sum_{k=1}^M b_k e^{-j\omega k}}{1 + \sum_{k=1}^N a_k e^{-j\omega k}}$$

$$\sigma_u^2$$

Che modello scegliere?

Siccome i modelli AR sono facili da identificare (Y.W.) e, dal teorema di Wold, un modello ARMA (o MA) di ordine finito può essere sempre rappresentato con un modello AR in generale di ordine infinito, solitamente si scelgono modelli AR. MAI modelli MA, per i quali la f di autocorrelazione è comunque nulla al di fuori dell'intervallo da  $-q$  a  $q$ , dove  $q$  è l'ordine del modello.

Si può dim che, se i parametri del modello sono le soluzioni dell'eq di Yule Walker, vale la seguente relazione:

$$\hat{P}_{\text{AR}}(\omega) = \sum_{m=-\infty}^{+\infty} R_y(m) e^{-j\omega m} = \sum_{m=-\infty}^{-p-1} R_y(m) e^{-j\omega m} + \sum_{m=-p}^{+p} R_y(m) e^{-j\omega m} + \sum_{m=p+1}^{+\infty} R_y(m) e^{-j\omega m}$$

estrapolazione  
fornita dal  
modello

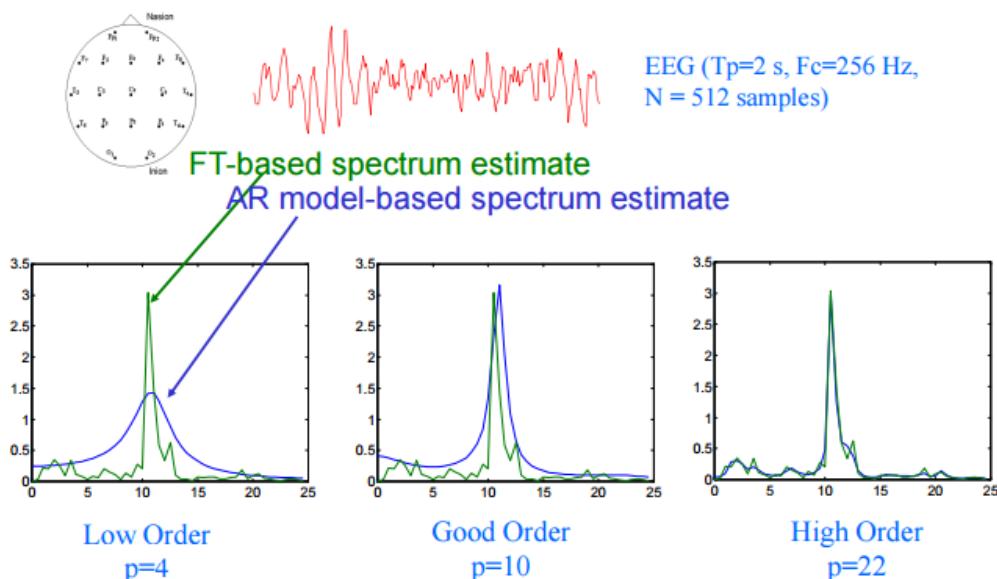
uguale alla f di  
autocorr  
stimata dai dati

estrapolazione  
fornita dal  
modello

Pertanto, il modello

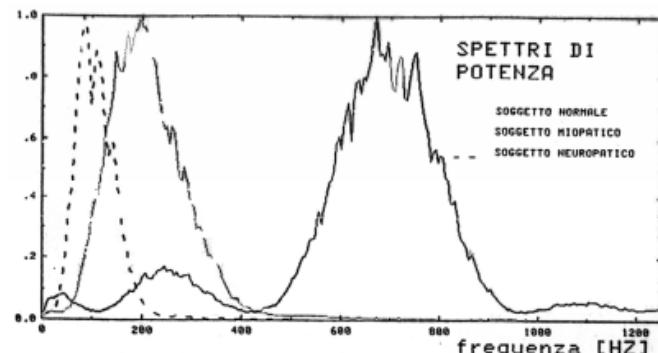
AR, identificato a partire dalle stime (dai dati) di  $R_s(m)$  per  $m$  compreso tra  $-p$  e  $p$ , consente di predire i suoi valori al di fuori di questo intervallo, e quindi evita di porli a zero, come invece avviene nei metodi FT indiretti. Questo anche spiega perché non si usano modelli MA.

Anche la scelta dell'ordine del modello è in questo caso cruciale e, oltre ai soliti metodi, si può utilizzare il confronto con i metodi FT



#### ESEMPI:

- 1) EMG: Lo spettro EMG del soggetto miopatico (linea scura) rivela frequenze più elevate di quello normale ( linea chiara) e neuropatico (linea tratteggiata). Quindi parametri estratti dal dominio della frequenza, es. frequenza di picco o frequenza media o frequenza mediana sono parametri di potenziale interesse clinico, perchè in grado di discriminare tra le varie classi di soggetti



- 2) PCG per la valutazione di protesi valvolare: Il fonocardiogramma (PCG) consente una valutazione funzionale di protesi valvolari cardiache da un punto di vista biomeccanico ed emodinamico, e quindi permette di individuare malfunzionamenti dovuti alla neoformazione di uno strato di tessuto fibroso o alla rottura di una parte della valvola impiantata. Si isola all'interno del segnale, il suono relativo alla valvola in esame utilizzando come riferimento segnali di ECG o CP (polso carotideo). Ad esempio, la calcificazione e la formazione di uno strato fibroso comportano un innalzamento della frequenza media (baricentro di frequenza). Pertanto si può pensare di usare tale parametro a fini diagnostici
- 3) Analisi spettrale del segnale di varibilità cardiaca: la durata del ciclo cardiaco viene misurata tra due picchi R e rappresentata nel tacogramma. Rappresentando le componenti in frequenza è possibile capire l'attività del sistema nervoso simpatico e parasimpatico:
  - 1) Banda VLF (Very Low Frequency) fra 0.01 e 0.04Hz • attività del Sistema Nervoso Simpatico • cambiamenti nella termoregolazione • preoccupazioni e pensieri ossessivi
  - 2) banda LF (Low Frequency) fra 0.04 e 0.15 Hz • attività del Sistema Nervoso Simpatico • attività di regolazione dei barocettori
  - 3) HF (High Frequency) fra 0.15 e 0.4 Hz. • attività del Sistema Nervoso Parasimpatico e del Vagale  
Il rapporto LF/HF è espressione del bilancio simpatico/vegetale.

- 4) Spettro dell'EEG: le componenti (alfa, beta, gamma theta, delta) sono definite proprio sulla base delle bande i frequenza. Possiamo quindi estrarre vari parametri (frequenza media o mediana. Potenza assoluta o relativa delle varie bande) usati per quantificare lo stato del soggetto, alterazioni indotte da patologie, effetto di farmaci....

Inoltre le mappe EEG Permettono di rappresentare in modo immediato la distribuzione topografica nello scalpo delle potenze relative nelle varie bande. Vengono acquisiti diversi segnali EEG da una matrice di elettrodi e si calcolano le RP per ogni banda. Si utilizzano poi algoritmi di interpolazione per stimare il parametro su ogni pixel e si colorano in modo uniforme le regioni in cui il parametro assume valori simili. Un metodo utilizzato per l'interpolazione è il k-NN: di è la distanza tra  $(x_i, y_i)$  e  $(x_j, y_j)$  e si scelgono i K punti per i quali tali distanza è minima. K e m sono parametri da fissare

$$P(x, y) = \frac{\sum_{i=1}^K P(x_i, y_i) d_i^{-m}}{\sum_{i=1}^K d_i^{-m}}$$

In generale: Le mappe sono rappresentazioni convenzionali, molto efficaci perché di immediata comprensione, della distribuzione spaziale di un certo parametro Valori diversi del parametro vengono associati a colori diversi Serve la chiave di lettura: parametro e scala!! Per costruire la mappa si utilizza un algoritmo di interpolazione perché generalmente il parametro viene misurato/calcolato su un numero limitato di punti -- algoritmi diversi possono dare origine a mappe diverse Si possono confrontare mappe relative a soggetti normali e patologici, e rappresentare in una nuova mappa l'esito di elaborazioni o test statistici (es. il livello di significatività del test t su valori normali vs patologici per ogni pixel)

Potenza relativa nelle bande d/q/a/b	Frequenza media
$RP = 100 \frac{\int_{low}^{high} P(f) df}{\int_{0.5}^{25} P(f) df}$	$Fm = \frac{\int_{0.5}^{25} f \cdot P(f) df}{\int_{0.5}^{25} P(f) df}$

## 9. ANALISI SPETTRALE MULTIVARIATA

Spesso interessa quantificare la relazione mutua tra due segnali, che possono essere sia di natura diversa, es: variabilità del battito cardiaco e pressione sanguigna, che della stessa natura, es. EEG prelevato da diverse posizioni dello scalpo. Con riferimento al secondo esempio, l'analisi di coerenza permette di capire se due aree della corteccia cerebrale sono funzionalmente correlate, ad es. se hanno generatori sottocorticali comuni.

Se interpretiamo i due segnali come realizzazioni di due processi aleatori  $X$  e  $Y$  a tempo discreto, stazionari e a media nulla, la relazione mutua tra i due processi può essere quantificata sia nel dominio del tempo che della frequenza usando varie funzioni

- CROSS-CORRELATION  $R_{xy}(k) = E[x(n)y(n+k)]$

-In generale non è una funzione pari, ma dall'ipotesi di stazionarietà segue che:

$$R_{xy}(k) = E[x(n)y(n+k)] = E[x(n-k)y(n)] = E[y(n)x(n-k)] = R_{yx}(-k)$$

-se  $X$  e  $Y$  sono processi indipendenti e a media nulla:  $R_{xy}(k) = 0$

-se  $X=Y$  sono processi coincidenti:  $R_{xy}(k) = R_x(k) = R_y(k)$

-se i processi sono legati da una relazione lineare:  $R_{xy}(k) = h(k) \otimes R_x(k) \rightarrow$

$$R_{xy}(k) = E \left[ x(n) \sum_j h(j) x(n+k-j) \right] = \text{Dim:}$$

$$= \sum_j h(j) E[x(n)x(n+k-j)] = \sum_j h(j) R_x(k-j) = \\ = h(k) \otimes R_x(k)$$

-se  $Y$  è la somma di due componenti, la prima ( $Z$ ) legata da una relazione lineare a  $X$  e l'altra ( $E$ )

$$\text{indipendente da } X, \text{ allora } R_{xy}(k) = h(k) \otimes R_x(k) \rightarrow \text{Dim: } R_{xy}(k) = E[x(n)(z(n+k) + e(n+k))] = E[x(n)z(n+k)] + E[x(n)e(n+k)] = \\ = R_{xz}(k) = h(k) \otimes R_x(k)$$

- CROSS SPECTRUM  $P_{xy}(\omega) = \text{FT}[R_{xy}(k)]$

E' la FT della funzione di cross correlazione. Evidenzia le componenti spettrali comuni ai due processi.

-se  $X$  e  $Y$  sono indipendenti e a media nulla:  $P_{xy}(\omega) = 0$

-se  $X$  e  $Y$  sono processi coincidenti  $X=Y$   $P_{xy}(\omega) = P_x(\omega) = P_y(\omega)$

- se i processi sono legati da una relazione lineare:  $P_{xy}(\omega) = H(\omega)P_x(\omega)$

- se  $Y$  è la somma di due componenti, la prima ( $Z$ ) legata da una relazione lineare a  $X$  e l'altra ( $E$ ) indipendente da  $X$ , allora  $P_{xy}(\omega) = P_{xz}(\omega) = H(\omega)P_x(\omega)$

$$C_{xy}^2(\omega) = \frac{|P_{xy}(\omega)|^2}{P_x(\omega) \cdot P_y(\omega)}$$

- COERENZA

E' una funzione reale, che normalizza il modulo al quadrato dello spettro incrociato tra due processi rispetto agli spettri dei due processi (è la normalizzazione del cross spettro rispetto agli spettri dei due processi)

-se  $X$  e  $Y$  sono indipendenti e a media nulla:  $C_{xy}^2(\omega) = 0$

-se  $X$  e  $Y$  sono processi coincidenti  $X=Y$   $C_{xy}^2(\omega) = 1$

- se i processi sono legati da una relazione lineare:  $C_{xy}^2(\omega) = 1$

- se  $Y$  è la somma di due componenti, la prima ( $Z$ ) legata da una relazione lineare a  $X$  e l'altra ( $E$ )

$$C_{xy}^2(\omega) = \frac{|H(\omega)P_x(\omega)|^2}{P_x(\omega)P_y(\omega)} = \frac{P_z(\omega)}{P_y(\omega)}$$

indipendente da  $X$ , allora

Per stimare la coerenza si usano sia metodi basati sulla FT che metodi parametrici.

La stima della funzione coerenza con i metodi FT richiede alcune cautele. Infatti, se ad es. sotto ipotesi di ergodicità, si stimano gli spettri  $P_x$ ,  $P_y$ , con il metodo del periodogramma e si estendono principi analoghi (definizione di uno stimatore della cross correlazione, e poi sua FT) per il calcolo di  $P_{xy}$  si ottiene un risultato

$$\hat{R}_{xy}(k) = \frac{1}{N} \sum_{n=0}^{N-1-k} x(n)y(n+k) = \frac{1}{N} x(k) \otimes y(-k)$$

$$\hat{P}_{xy}(\omega) = \frac{1}{N} X(\omega)Y(-\omega)$$

$$\hat{P}_x(\omega) = \frac{1}{N} |X(\omega)|^2$$

$$\hat{P}_y(\omega) = \frac{1}{N} |Y(\omega)|^2$$

*Quindi per avere dei risultati informativi è indispensabile applicare ai metodi di stima degli spettri la tecnica di averaging per migliorare le proprietà degli stimatori!!!*



$$C_{xy}^2(\omega) = 1$$

non informativo

Con i metodi parametrici i due processi  $X$  e  $Y$  vengono descritti come le uscite di un modello AR multivariato, pilotato da due processi bianchi, a media nulla. Se ora indichiamo i due processi con  $y_1$  e  $y_2$ , il modello stabilisce un legame tra  $y_1$  all'istante  $n$  e non solo  $y_1$  in  $p$  istanti precedenti e il rumore  $u_1$  all'istante attuale,



ma anche  $y_2$  nei  $p$  istanti precedenti. Analogamente per  $y_2$ .

$$y_1(n) = -\sum_{k=1}^p a_k^{(1,1)} y_1(n-k) - \sum_{k=1}^p a_k^{(1,2)} y_2(n-k) + u_1(n)$$

$$y_2(n) = -\sum_{k=1}^p a_k^{(2,1)} y_1(n-k) - \sum_{k=1}^p a_k^{(2,2)} y_2(n-k) + u_2(n) \rightarrow \underline{y}(n) = \sum_{k=1}^p A_k \cdot \underline{y}(n-k) + \underline{u}(n)$$

$\underline{y}(n)$  è il vettore  $[2 \times 1]$  di processi aleatori

$A_k$  matrice  $[2 \times 2]$  dei coefficienti, es

Per identificare i parametri si possono scrivere equazioni analoghe a quelle di Yule Walker, in funzione di

$\underline{u}(n)$  è il vettore  $[2 \times 1]$  di rumore bianco a media nulla e matrice di varianza:  $\Sigma_u = \begin{vmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{vmatrix}$   $R(k) = E[\underline{y}(n)\underline{y}(n+k)]$  matrice di autocorrelazione  $[2 \times 2]$ , che ha sulla diagonale principale le correlazioni dei due

processi e fuori della diagonale le cross-correlazioni

Una volta identificato il modello, è possibile stimare la matrice  $[2 \times 2]$  delle densità spettrali

$$\underline{P}_y(\omega) = \begin{vmatrix} P_1(\omega) & P_{12}(\omega) \\ P_{21}(\omega) & P_2(\omega) \end{vmatrix}$$

usando la seguente relazione che lega la matrice della densità spettrale del vettore di processi in uscita  $P_y$  alla matrice di trasferimento  $H$  (valutata per  $z=e^{j\omega}$ ) e alla matrice della densità spettrale

$$P_y(\omega) = H(\omega) \cdot \Sigma_u \cdot H^T(\omega) \text{ con } H(\omega) = \left( I - \sum_{k=1}^p A_k Z^{-k} \right)^{-1}_{z=e^{j\omega}}$$

del vettore di rumore bianco in ingresso  $\Sigma_u \rightarrow$

matrice  $P_y$  è Hermitiana, gli elementi sulla diagonale sono reali e rappresentano gli auto spettri, mentre quelli fuori diagonale sono complessi e rappresentano i cross-spettri.

In modo del tutto analogo a quanto visto per lo spettro del segnale EEG, anche a partire dalla funzione coerenza si possono definire degli indici quantitativi, ad es

• Frequenza media (C\_MDF)

$$C_{MDF} = \frac{\int_{f_{max}}^{f_{max}} f \cdot C_{xy}^2(f) df}{\int_{0.5}^{f_{max}} C_{xy}^2(f) df}$$

Frequenza del massimo  $f_{max}$

Coerenza relativa nelle varie bande

$$RC = 100 \cdot \frac{\int_{f_{low}}^{f_{high}} C_{xy}^2(f) df}{\int_{0.5}^{f_{max}} C_{xy}^2(f) df}$$

La funzione coerenza, compresa tra zero e uno, misura il grado di legame (lineare) tra due segnali, in funzione della frequenza. E' una misura di sincronismo tra due segnali, cioè permette di individuare le frequenze in cui i due processi risultano avere un certo grado di associazione, ma non permette di introdurre alcuna misura di causalità, cioè di capire quale dei due segnali pilota l'altro. Questo rappresenta un limite della coerenza. Per ovviare a ciò, esistono:

- definizioni alternative della funzione coerenza, ad esempio la coerenza diretta (DTF):

$$DTF_{1 \rightarrow 2}(\omega) = \frac{\sigma_1 H_{21}(\omega)}{\sqrt{\sigma_2^2 |H_{22}(\omega)|^2 + \sigma_1^2 |H_{21}(\omega)|^2}}$$

A differenza dalla coerenza ordinaria, la DTF fornisce indicazioni sulla direzione dell'interazione tra due segnali

$$DTF_{2 \rightarrow 1}(\omega) = \frac{\sigma_2 H_{12}(\omega)}{\sqrt{\sigma_1^2 |H_{11}(\omega)|^2 + \sigma_2^2 |H_{12}(\omega)|^2}}$$

Oppure si utilizzano i principi di causalità di Granger. L'idea è che dati 2 segnali possiamo dire che x influenza y se:

- l'informazione di x è utile a predire y (1° criterio)
- includendo i valori passati di x nella stima regressiva di y, si migliora la qualità della stima, ovvero si riduce l'errore di predizione (2° criterio)

Il primo criterio si basa sul modello AR; si osservano gli elementi fuori dalla diagonale delle matrici Ak e per capire se, in presenza di una coerenza non trascurabile, è il primo segnale che pilota il secondo, oppure il

$$A_k = \begin{vmatrix} a_k^{(1,1)} & a_k^{(1,2)} \\ a_k^{(2,1)} & a_k^{(2,2)} \end{vmatrix}$$

secondo che pilota il primo, oppure ci sono controlli incrociati tra i due segnali.

Si possono avere 4 situazioni:

- y1 causa y2 ma non viceversa se almeno un elemento in posizione (2,1) è significativamente diverso da zero mentre tutti gli elementi (1,2) non lo sono
- y2 causa y1 ma non viceversa se almeno un elemento in posizione (1,2) è significativamente diverso da zero mentre tutti gli elementi (2,1) non lo sono
- y1 causa y2 e y2 causa y1 se almeno un elemento in posizione (1,2) e almeno un elemento in posizione (2,1) sono significativamente diversi da zero
- y1 e y2 sono scorrelati tra loro se tutti i coefficienti in posizione (1,2) e (2,1) non sono significativamente diversi da zero

$$GC_{x \rightarrow y} = \ln \frac{\text{var}(y)}{\text{var}(y|x)}$$

Il secondo criterio definisce l'indice di causalità:

Dove

- ln è il logaritmo naturale
- var(y) è la varianza del rumore di ingresso per il modello AR monovariato, che misura la possibilità di predire y sulla base della sua storia passata
- var(y|x) è la varianza del rumore di ingresso per il modello multivariato, che misura la possibilità di predire a partire dalla storia passata non solo di y ma anche di x.

- GC è sempre definito non negativo
- GC = 0 quando nessun legame causale esiste tra i segnali : var(y) = var(y|x)
- GC significativamente diverso da zero indica la presenza di una relazione di causalità da x a y
- Il valore di GC è una misura della forza del legame
- Una differenza significativa tra gli indici di causalità  $x \rightarrow y$  e  $y \rightarrow x$  indica il verso della relazione

per valutare la differente significatività si usa il test statistico F-test:

Ipotesi nulla  $H_0 : \text{var}(y) = \text{var}(y|x)$

### Statistica F

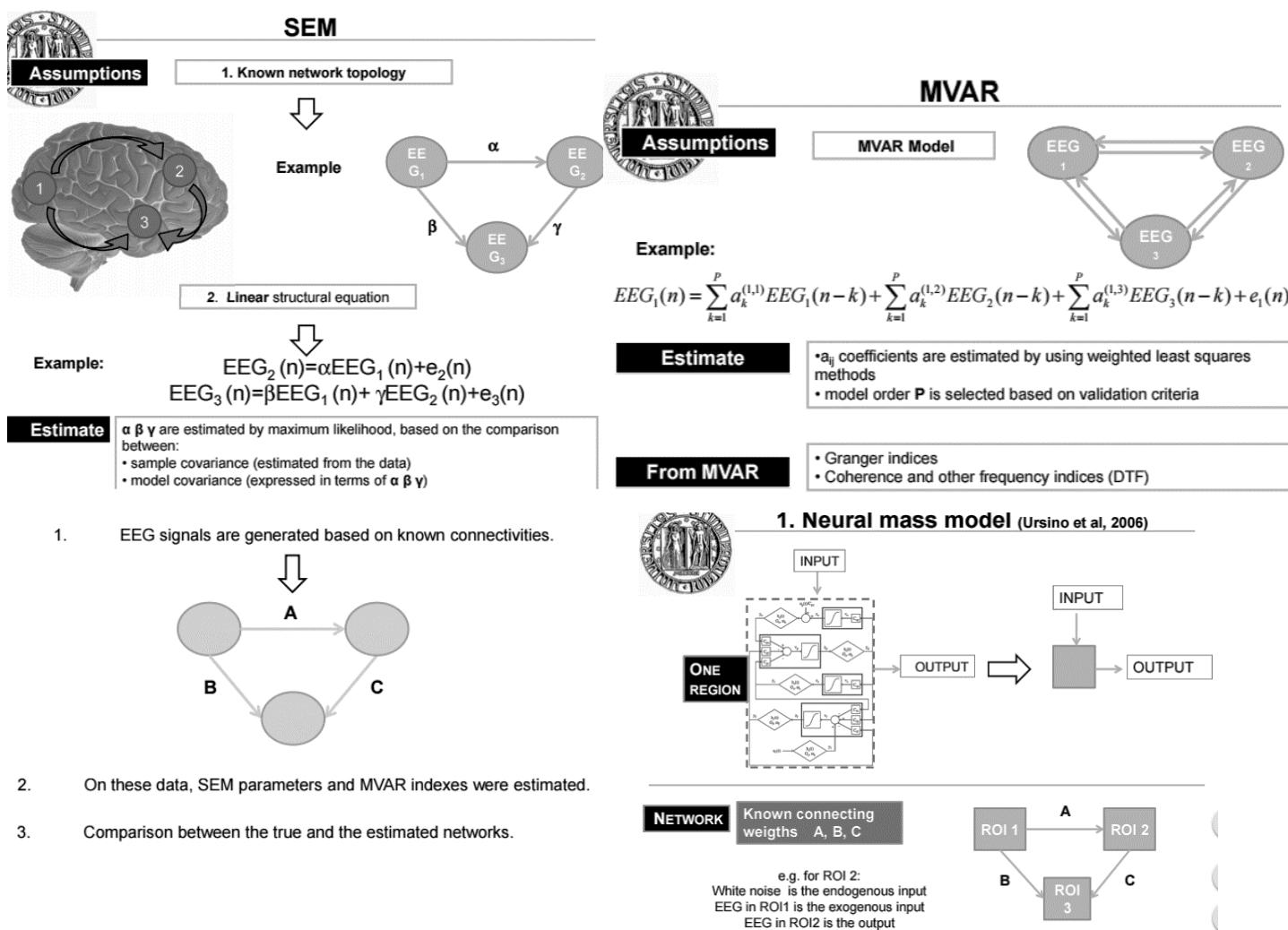
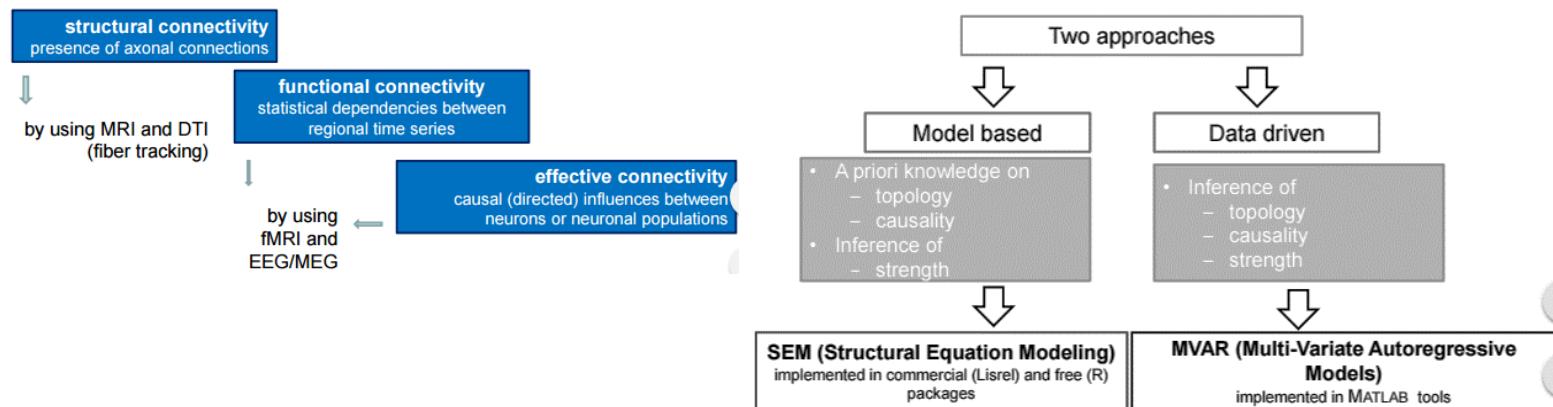
$$F = \frac{\frac{\text{var}(y) - \text{var}(y|x)}{p}}{\frac{\text{var}(y|x)}{(L-2p-1)}}$$

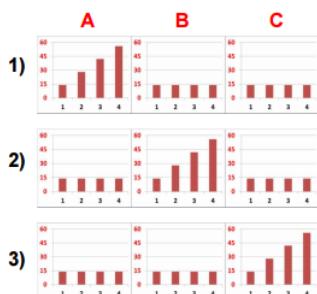
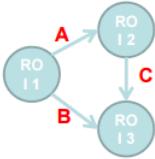
P = dimensione del modello MVAR  
L= numero di campioni

Sotto HO F è distribuita secondo la statistica [F di Fisher-Snedecor](#) con df pari a p e L-2p-1

Fissato un intervallo di significatività (es. Alfa=0,05) si valuta se accettare o rifiutare l'ipotesi nulla

esempio di applicazione: analisi di coerenza dell'EEG per lo studio della connettività cerebrale





Different networks are simulated, characterized by:  
different topologies  
different values of  
A, B, C

For each model:  
100 realizations are generated

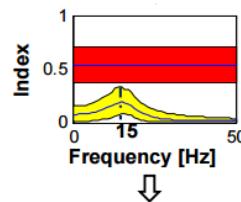
## 5. MVAR – GC and Frequency indexes

On each data set (triplet of EEGs from the three ROIs) a MVAR model is estimated

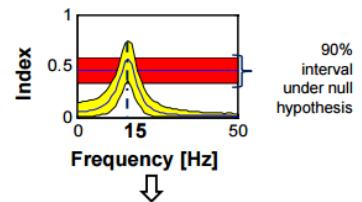
From MVAR:

GC is evaluated. It provides information on: 1. topology (significant GC means that a connection is present) 2. strength (CG value) 3. direction

Coherence is evaluated, and by elaborating on it, DTF and PDC are defined to measure the **LINEAR** coupling A → B at various frequencies.



No significative coupling



Significative coupling at ≈15 Hz



## Conclusion

TOPOLOGY

STRENGTH

DIRECTED/INDIRECTED CAUSALITY

SEM	known	✓	known
-----	-------	---	-------

Granger Causality	✓	✓	✓
-------------------	---	---	---

Frequency indices	✓	✓	✗
-------------------	---	---	---

**Granger causality and frequency indexes are not competitive but complementary techniques**

→ Use together!

## 10. RAPPRESENTAZIONI TEMPO-FREQUENZA

L'analisi di Fourier, che gioca un ruolo fondamentale per l'analisi spettrale segnali biologici, decompone il segnale nelle sue componenti periodiche, seno e coseno.

$$x(t) = \text{FT}^{-1}[X(f)] = \frac{1}{2\pi} \int_{-\infty}^{+\infty} X(f) e^{j2\pi f t} df$$

Per un certo valore  $f_1$  di frequenza,  $X(f_1)$  rappresenta il peso di seno e coseno a frequenza  $f_1$  nel segnale  $x(t)$ . Tale peso viene calcolata attraverso la trasformata inversa calcolata in corrispondenza della frequenza  $f_1$ :

$$X(f_1) = \int_{-\infty}^{+\infty} x(t) e^{-j2\pi f_1 t} dt$$

quindi moltiplicando il segnale per l'esponenziale complesso (seno e coseno) alla frequenza  $f_1$  e poi integrando su tutto il dominio temporale.

Non si fornisce quindi nessuna informazione sull'eventuale localizzazione nel tempo delle componenti periodiche o sulla variazione nel tempo dell'ampiezza delle componenti spettrali.

Una prima soluzione è rappresentata dalla Short Time Fourier Transform (STFT) che si basa sulla segmentazione del segnale attraverso una finestra temporale e sull'analisi in frequenza di ogni segmento, supposto stazionario.

STFT trasforma un segnale monodimensionale che varia nel tempo in un segnale bidimensionale del tempo e della frequenza che analizza le variazioni nel tempo del contenuto spettrale del segnale. Pertanto: fissata una certa finestra rettangolare centrata al tempo  $t$ ,  $X(t,f)$  evidenzia le componenti spettrali del segmento di segnale contenuto nella finestra. Come vedremo più avanti, vale anche il duale, cioè fissata un intervallo di frequenza centrale in  $f$ ,  $X(t,f)$  evidenzia come variano nel tempo le componenti dello spettro relative alle frequenze in questo intervallo.

Il modulo al quadrato della STFT fornisce una stima dello spettro tempo-frequenza.

La dimensione (e la forma) della finestra utilizzata influenza lo spettrogramma. Finestre rettangolari, molto semplici da implementare, comportano una stima spettrale che converge in media verso il prodotto di convoluzione dello spettro vero per una funzione sinc

Si considera una finestra  $w(\tau)$  reale, es. finestra rettangolare, finestra gaussiana ecc.

Fissato un istante  $t$ :

- si costruisce la finestra traslata in  $t$ :  $w(\tau - t)$
- si moltiplica il segnale  $x(\tau)$  per la finestra  $w(\tau - t)$

$$X(t,f) = \text{STFT}[x(t)] = \text{FT}[x(\tau)w(\tau - t)] = \int_{-\infty}^{+\infty} x(\tau)w(\tau - t)e^{-j2\pi f \tau} d\tau$$

$X(t,f)$  risulta formata dalle componenti dello spettro relative al segnale in un intorno del tempo  $t$

PROPRIETA' DELLA STFT:

- Traslazione nel tempo corrisponde ad uno shift della fase  $x(t-t_0) \xrightarrow{\text{STFT}} X(t-t_0)$
- Dualmente  $x(t) \exp(j2\pi f_0 t) \xrightarrow{\text{STFT}} X(t, f-f_0)$

$$X(t,f) = e^{-j2\pi f t} \int_{-\infty}^{+\infty} x(\tau)w(\tau-t)e^{-j2\pi f(\tau-t)} d\tau$$

Dalla definizione di STFT si può scrivere

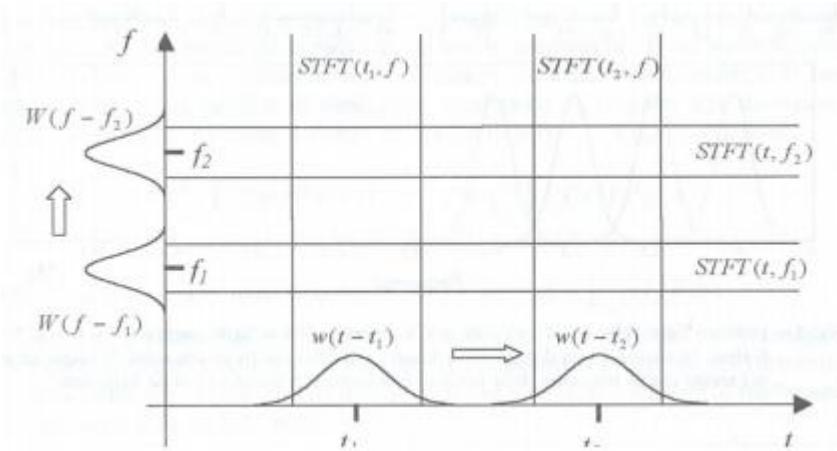
Si consideri una frequenza  $f_1$ . L'integrale è la convoluzione tra  $x(t)$  e  $w(-t)e^{j2\pi f_1 t}$ , ovvero:

$$X(t, f_1) = e^{-j2\pi f_1 t} x(t) \otimes [w(-t)e^{j2\pi f_1 t}]$$

Fattore di demodulazione

uscita di un filtro passa-banda con risposta impulsiva pari alla portante a frequenza  $f_1$  modulata dalla finestra  $w$ , quindi con risposta in frequenza pari a  $W(f-f_1)$

Al variare di  $f_1$ , NON varia la forma del filtro ma varia la frequenza centrale: ciò consente di interpretare la STFT come banco di filtri



$$X(t, f) = \int_{-\infty}^{+\infty} x(\tau) e^{-j2\pi f\tau} \cdot w(\tau - t) d\tau = \int_{-\infty}^{+\infty} x(\tau) \cdot h_{t,f}(\tau) d\tau$$

con  $h_{t,f}(\tau) = e^{-j2\pi f\tau} w(\tau - t)$

ad ogni istante  $t_1$  la STFT è la FT del segnale selezionato dalla finestra  $w(t-t_1)$   
Ad ogni frequenza  $f_1$  la STFT isola le componenti del segnale viste attraverso il filtro  $W(f-f_1)$ .

Pertanto, STFT può essere vista come la proiezione del segnale  $x(t)$  su una famiglia di funzioni  $h$ , derivate dalla  $w$  per mezzo di:

- traslazione nel tempo  $w(t-\tau)$
- traslazione in frequenza  $e^{-j2\pi f\tau}$

$$\int_{-\infty}^{+\infty} w(\tau) d\tau = 1$$

ANTITRASFORMATA STFT: Se la funzione  $w$  ha area unitaria: il segnale  $x(t)$  può essere ricostruito:

$$x(t) = \int_{-\infty}^{+\infty} \left[ \int_{-\infty}^{+\infty} X(\tau, f) e^{j2\pi f\tau} df \right] d\tau$$

antitrasformata della STFT vista come funzione della sola frequenza

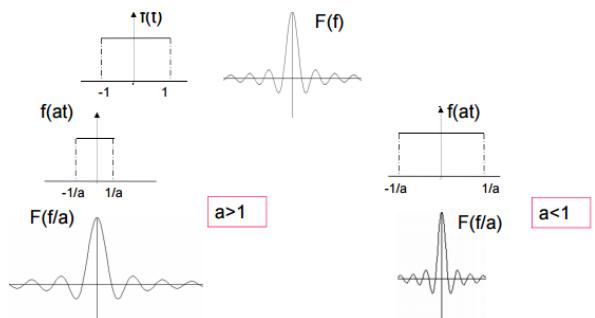
La STFT è lo spettro locale in un intorno del tempo  $t$ , ottenuto selezionando una parte di segnale con la finestra di analisi  $w(t)$ . Per avere una buona risoluzione nel tempo, serve una finestra "stretta". D'altra parte l'interpretazione della STFT come banco di filtri mostra che per avere una buona risoluzione in frequenza è necessario che  $W(f)$  abbia banda stretta. Queste due esigenze sono contrastanti, in quanto ad una variazione (compressione o espansione) della durata [ $w(at)$ ] corrisponde una variazione opposta (espansione o compressione) della banda [ $W(f/a)$ ]:

$$\text{FT}[w(at)] = \frac{1}{|a|} W\left(\frac{f}{a}\right)$$

$a$  è un qualsiasi numero reale positivo che rappresenta il fattore di scala.

Quando si cambia la scala ad una funzione, cioè si passa da  $f(t)$  a  $f(at)$ , questa

- se  $a < 1$  si contrae nel tempo e si espande in frequenza
- se  $a > 1$  si contrae nel tempo e si espande in frequenza



Analogamente, Quando si passa da  $f(t)$  a  $f(t/a)$ :

- se  $a < 1$  si espande nel tempo e si contrae in frequenza
- se  $a > 1$  si contrae nel tempo e si espande in frequenza

Le relazioni viste suggeriscono che durata e banda di un segnale sono inversamente proporzionali. Dato che la durata di  $w$  e/o la banda di  $W$  potrebbero essere infinite, conviene definire una misura di concentrazione (o dispersione), legata alla media quadratica

$$\Delta t^2 = \frac{1}{E_w} \int_{-\infty}^{+\infty} t^2 |w(t)|^2 dt$$

$$\Delta f^2 = \frac{1}{E_w} \int_{-\infty}^{+\infty} f^2 |W(f)|^2 df$$

con

$$E_w = \int_{-\infty}^{+\infty} |w(t)|^2 dt = \int_{-\infty}^{+\infty} |W(f)|^2 df$$

Si può dim:

$$\Delta t \Delta f \geq \frac{1}{4\pi}$$

**Principio di indeterminazione di Heisenberg**

Nella STFT si usa sempre la stessa finestra, sia al variare di  $t$  che al variare di  $f$ , quindi i valori di risoluzione  $\Delta t$  e  $\Delta f$  sono costanti nell'intero piano tempo-frequenza. Ciò è abbastanza innaturale: il concetto di frequenza è associato a quello di oscillazione e per poter affermare che è presente un'oscillazione ad una certa frequenza con ragionevole precisione è necessario osservare un fenomeno per un intervallo che includa uno o più periodi. Se ad es. siamo interessati a rilevare un ritmo circadiano, ( $f=1.16 \text{ Hz}$ ) dobbiamo osservare per uno o più giorni, mentre se siamo interessati a ritmi veloci quali l'interferenza a 50Hz è sufficiente una finestra della durata di decimi di secondo.

Per ovviare a questi problemi, J Morlet introdusse l'idea base della trasformata wavelet, che consente di decomporre il segnale in funzioni base: compresse nel tempo alle alte frequenze, perché sono sufficienti finestre di durata limitata ( $\rightarrow$  non serve alta risoluzione temporale) per seguire variazioni rapide del segnale dilatate nel tempo alle basse frequenze, quando sono necessarie finestre ampie per riconoscere variazioni lente del segnale  $P$ .

$$X(t,f) = \int_{-\infty}^{+\infty} x(\tau) \cdot h_{t,f}(\tau) d\tau$$

Riprendiamo l'espressione della STFT riscritta come  $X(t,f) = \int_{-\infty}^{+\infty} x(\tau) \cdot h_{t,f}^*(\tau) d\tau$  con  $h$  derivata dalla finestra  $w$  per mezzo di: traslazione nel tempo e traslazione in frequenza  $\rightarrow h_{t,f}(\tau) = w(\tau - t) e^{-j2\pi f\tau}$

$$X(t,f) = \int_{-\infty}^{+\infty} x(\tau) \cdot h_{t,f}^*(\tau) d\tau$$

Nella trasformata Wavelet l'espressione è la stessa ma  $h$  è derivata da una finestra di riferimento  $h_0$  (mother wavelet)  $h_0(\tau) = w(\tau) e^{-j2\pi f_0 \tau}$ .

Tale mother wavelet viene modificata per mezzo di: • traslazione nel tempo • cambiamento di scala  $h_{t,a}(\tau) = \frac{1}{\sqrt{a}} h_0\left(\frac{\tau-t}{a}\right) = \frac{1}{\sqrt{a}} w\left(\frac{\tau-t}{a}\right) e^{-j2\pi f_0\left(\frac{\tau-t}{a}\right)}$   $h$  è quindi una funzione del tempo e del fattore di scala  $a$ , con  $a = f/f_0$ . Il termine  $1/\sqrt{a}$  è introdotto per normalizzare l'energia della wavelet.

Nell'analisi dei segnali in cui il concetto di frequenza è di primaria importanza, come nel caso del segnale EEG, è opportuno far ricadere la scelta della forma della wavelet su una famiglia di funzioni ben caratterizzate

in frequenza, come sono le wavelet con inviluppo gaussiano: Per tali wavelet  $h_0(t) = e^{-\frac{1}{2} \left(\frac{t}{\sigma_t}\right)^2} e^{j2\pi f_0 t}$  a cui corrisponde  $H_0(f) = \sqrt{2\pi\sigma_t^2} e^{-2\pi^2\sigma_t^2(f-f_0)^2}$ . I parametri da settare sono  $\sigma_t$  (legato alla durata della finestra) e  $f_0$  (legato al numero di oscillazioni)

Abbiamo già visto che quando si cambia la scala ad una funzione, cioè si passa da  $f(t)$  a  $f(t/a)$ , questa • si

espande se  $a > 1$  • si contrae se  $a < 1$ . Allora nella WT, qui riscritta come  $WT(t,a) = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} x(\tau) \cdot h^*\left(\frac{\tau-t}{a}\right) d\tau$  all'aumentare di  $a$  la funzione  $h$  si espande nel tempo (peggiora la risoluzione temporale) e la WT può tener conto di variazioni lente del segnale (migliora la risoluzione in frequenza). Si ha quindi filtraggio con un banco

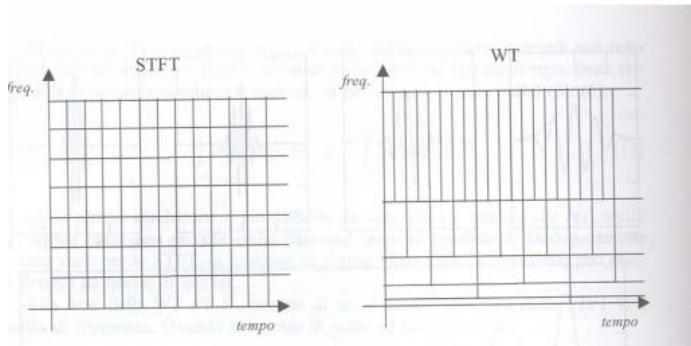
di filtri con risposta all'impulso  $h$  la cui estensione aumenta con  $a$ . Si può ritornare alla frequenza attraverso la relazione  $a=f^*/f$  dove  $f^*$  è la freq centrale della FT della wavelet di base (mother wavelet).

La risoluzione dipende dalla funzione wavelet scelta e può essere valutata ricorrendo alle misure  $\Delta t$  e  $\Delta f$  viste in precedenza. Se indichiamo con  $\Delta t_m$  e  $\Delta f_m$  i valori della wavelet madre ( $a=1$ ) per la generica funzione si

$$\Delta t = a \cdot \Delta t_m$$

$\Delta f = \frac{1}{a} \cdot \Delta f_m$   
ha Pertanto il prodotto rimane costante, e al solito inferiormente limitato da  $4\pi$ . Le wavelet con inviluppo gaussiano (wavelet di Morlet) raggiungono il limite inferiore.

### Suddivisione nel piano tempo-frequenza



Sotto quali condizioni una funzione può essere una mother wavelet?

- Energia finita
  - Media nulla
  - La trasformata wavelet di un segnale 1d è una rappresentazione 2d del segnale: è necessario che non si perda informazione nella trasformata. A tale scopo è necessario che sia verificata la seguente espressione:  

$$c_g = \int \frac{|H(f)|^2}{|f|} df < +\infty$$
- Con  $c_g$ : costante di ammissibilità, è legato alla wavelet scelta. Ciò implica che la trasformata di Fourier:  $H(f)$  deve avere un valore 0 alla frequenza  $f=0$ .

# 11. METODI DI CLASSIFICAZIONE AUTOMATICA IN MEDICINA

Problema: dati un evento x  
Un numero M di classi w<sub>1</sub>, w<sub>2</sub>, w<sub>3</sub>...w<sub>M</sub>

Assegnare l'evento ad una di queste classi

Esempi: Dal segnale ECG misurato in un soggetto, definire se il soggetto è normale, oppure ha ipertrofia ventricolare oppure è in corso un infarto

Da misure attraverso spettrometria di massa del proteoma di un soggetto, definire se il soggetto è normale o ha una certa forma di tumore

Da una immagine al microscopio riconoscere e contare cellule con caratteristiche diverse per quantificare emocromo e formula leucocitaria

La definizione delle classi è uno degli aspetti principali di questo problema. In alcuni (rari) casi le classi sono definite in modo preciso e la classificazione non è ambigua. Più in generale non esiste una definizione "esatta" di classi (vd esempi precedenti). Considereremo nel seguito la situazione in cui le classi sono caratterizzate attraverso un processo di apprendimento, basato sull'osservazione di un numero di esempi di cui è nota con precisione l'appartenenza ad una specifica classe. Tale insieme di esempi viene detto training set e con riferimento a queste situazioni si parla di classificazione supervised. Ci sono poi delle situazioni in cui si dispone di un numero di esempi di cui però non è nota la classe di appartenenza. Un primo obiettivo (cluster analysis) è allora di riconoscere la presenza in questo insieme di gruppi di dati con caratteristiche simili, ognuno dei quali viene ricondotto ad una classe. Si parla allora di classificazione unsupervised.

## CLASSIFICAZIONE SUPERVISED

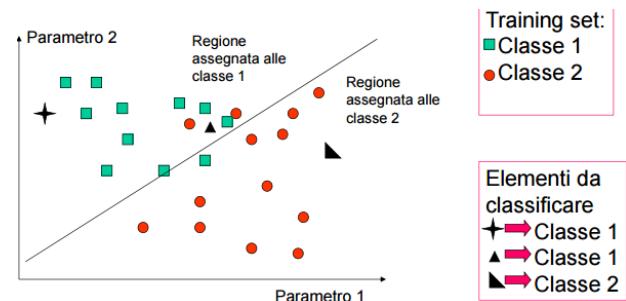
Due approcci: geometrico o matematico → esistono vari metodi: statistici, reti neurali, alberi decisionali, support vector machines ecc

Sintattico o linguistico → Si basa sull'assunzione che eventi complessi possono essere descritti come composizione ricorsiva di eventi più semplici, così come in un linguaggio un periodo è composto da più frasi, a loro volta composte da più parole. Le regole di composizione (grammatica), auspicabilmente diverse in ogni classe, vengono apprese attraverso l'osservazione del training set. L'elemento da classificare viene scomposto e la sua grammatica esplorata. L'elemento viene assegnato alla classe che ha una grammatica più simile E' un approccio attraente ma difficile da tradurre in pratica, infatti non sono molte le applicazioni in cui l'approccio è stato utilizzato con successo

Approccio geometrico: Ogni elemento da classificare è rappresentato con un vettore di parametri (features) Questi valori vengono definiti come un punto nello spazio dei parametri. Se il vettore dei parametri è scelto in modo adeguato, gli elementi del training set permettono di ripartire lo spazio dei parametri, attraverso opportuni criteri (vedremo tecniche bayesiane) in regioni assegnati alle varie classi. Spesso

le classi non sono separate in modo netto. Quindi non si può escludere che il classificatore commetta degli errori

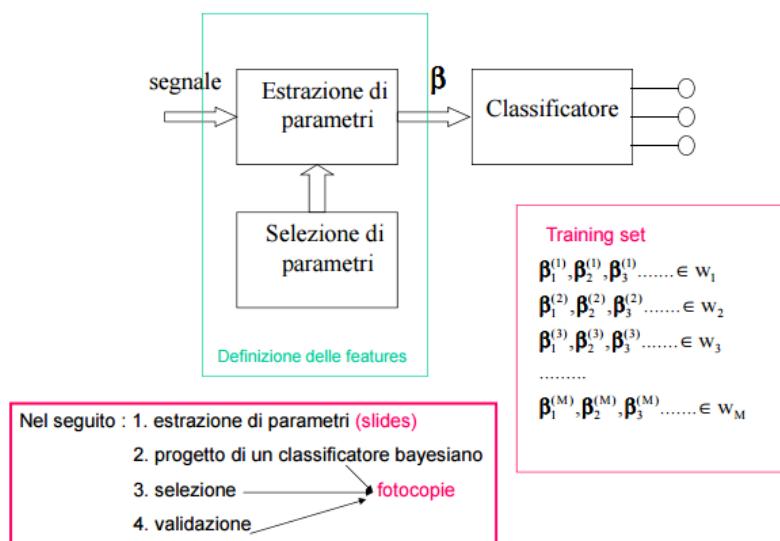
-definizione dei parametri: scelti in modo da ottimizzare la separazione tra classi. Si parte da un numero elevato per poi ridurli attraverso selezioni



- progetto del classificatore: Dal training set, viene individuata la ripartizione dello spazio di parametri (ad es. nella figura precedente si è usato un classificatore lineare) sulla base di opportuni criteri (es. di tipo statistico, o basati su misure di distanza).
- validazione del classificatore: Si testa la capacità del classificatore di dare una risposta corretta, quantificando le sue prestazioni su un insieme di dati di cui è nota la classe di appartenenza (testing set)
- uso del classificatore: si applica il classificatore all'evento x da classificare

Il vettore dei parametri deve essere scelto in modo da ottimizzare la separazione tra le classi. Quindi la scelta dei parametri è critica, e non ci sono indicazioni generali su come scegliere i parametri. Può essere allora vantaggioso usare un numero elevato di features. D'altra parte, più sono numerose le features, più complessa è la fase di progetto del classificatore e più numeroso dovrebbe essere il training set. Inoltre la prestazione del classificatore si degrada se le features sono ridondanti Allora si parte definendo un gran numero di features, e poi si lascia ad una fase successiva detta di selezione (esistono varie tecniche allo scopo) il compito di isolare le poche features, o combinazioni di features, in grado di discriminare tra le varie classi

### Schema del progetto di un classificatore



#### ESTRAZIONE DI PARAMETRI:

Supponiamo che l'evento da classificare sia un segnale  $x(n)$ ,  $n=0,1,\dots,N_{tot}-1$  i cui campioni sono contenuti nel vettore  $y$  di dimensione  $N_{tot}$

L'estrazione del vettore  $\beta$  che contiene le  $N$  features può essere visto come la funzione  $\beta=F(y)$  con  $F$  funzione in genere non lineare.

Si distinguono due tipi di parametri:

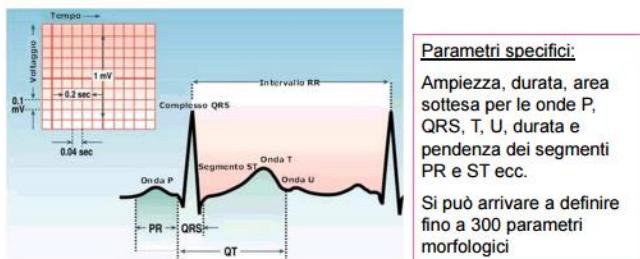
- Specifici : ampiezza, durata, pendenza.. vantaggi: conservano il significato fisiopatologico svantaggi: • Relazione non lineare con il segnale • Difficile misurarli con uguale precisione • Tra loro dipendenti (ridondanti) • Difficile quantificare la loro informazione utile Quindi arbitraria la scelta del tipo e del numero

-Non Specifici: da trasformazione lineari dei dati (es. PCA, Fourier,...)

vantaggi & svantaggi : complementari ai precedenti

Spesso ai parametri estratti da un segnale vengono aggiunti parametri relativi al soggetto, ad es, età, fattori di rischio, esiti di altri esami, trattamenti ricevuti. Alcune di queste variabili sono continue (es, età) altre sono discrete (il soggetto ha o non ha ricevuto un trattamento). Inoltre le varie features possono assumere valori diversi in termini di ordine di grandezza, anche in relazione alle unità di misura adottate. Soluzione → Normalizzazione dei parametri con funzioni come divisione per il massimo valore assunto nel training set o divisione per il valore medio assunto nel training set

## Esempio1: ECG



Alcuni problemi: che periodo scegliere?

mediare alcuni periodi e poi estrarre i parametri?

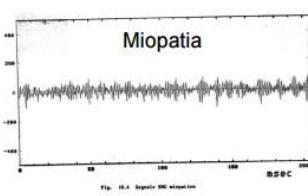
estrarre i parametri da più complessi e poi mediare?

Esistono dei margini di soggettività nel modo in cui i parametri sono definiti  
Importanza di standardizzare il modo in cui i parametri vengono estratti!!!

## EMG: parametri (specifici) nel dominio del tempo

Aampiezza media

MIOPATIA ↓ NEUROPATIA ↑

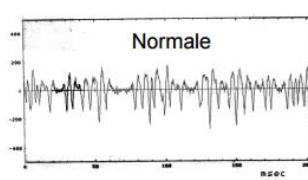


Numero di massimi positivi sopra una certa soglia

MIOPATIA ↑ NEUROPATIA ↓

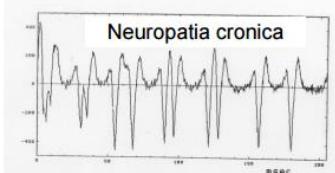
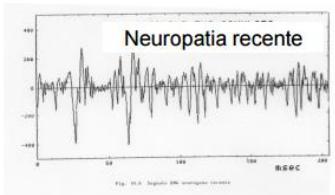
Numero di attraversamenti della linea di base

MIOPATIA ↑ NEUROPATIA ↓



Rapporto tra numero di max positivi e ampiezza media

MIOPATIA ↑ NEUROPATIA ↓



## 12. CLASSIFICATORE DI BAYES

Si basa su una descrizione probabilistica delle classi e dei valori dei parametri delle varie classi.

Prendiamo l'esempio più semplice: 2 classi ( $w_1$ = pazienti sani  $w_2$ = pazienti malati)

1 parametro ( $\beta$ =temperatura corporea)

Indichiamo con  $P(w_1)$  la probabilità a priori che un paziente sia sano e con  $P(w_2)$  la probabilità che sia malato (che appartenga alla classe  $w_2$ ). Vale  $P(w_1) + P(w_2) = 1$ , le due classi sono mutuamente esclusive.

Consideriamo  $\beta$  come una v.a. continua e quindi possiamo definire  $P(\beta|w_1)$  e  $P(\beta|w_2)$ .

Dopo aver misurato la temperatura di un paziente da classificare e aver trovato che questo valore è pari ad un certo  $\beta$ , come questa misura influenza la probabilità che il soggetto sia sano o malato? Dalla regola di

$$P(\omega_i|\beta) = \frac{P(\beta|\omega_i)P(\omega_i)}{P(\beta)} \text{ con } i=1,2$$

bayes: dove  $p(\beta) = \sum_{i=1}^2 p(\beta|\omega_i)P(\omega_i)$

e la regola di scelta è:

$$\beta \rightarrow \omega_1 \quad \text{se} \quad P(\omega_1|\beta) > P(\omega_2|\beta)$$

$$\beta \rightarrow \omega_2 \quad \text{se viceversa} \quad P(\omega_1|\beta) < P(\omega_2|\beta)$$

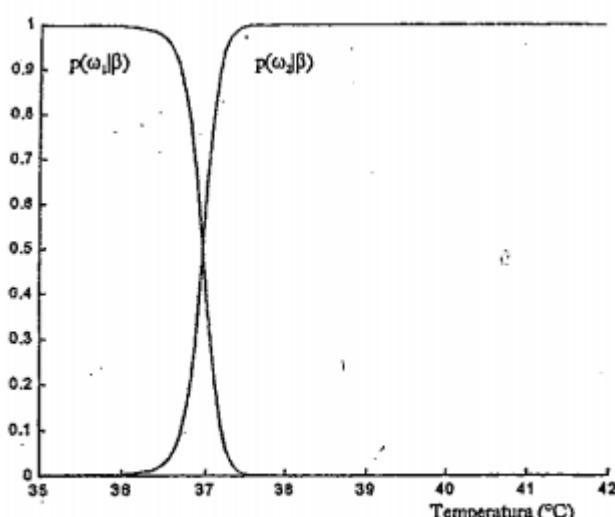


Fig. 1.2. Densità di probabilità a posteriori.

Indichiamo con  $H$  l'ascissa dell'intersezione tra le due

$$\beta \rightarrow \omega_1 \quad \text{se} \quad \beta < H$$

$$\beta \rightarrow \omega_2 \quad \text{se} \quad \beta > H$$

La scelta di  $H$  come soglia è quella che minimizza l'errore; infatti data una generica soglia  $TH$ , abbiamo due possibilità d'errore:

$$\text{Caso 1: } \beta \in \omega_1 \quad \text{ma} \quad \beta > TH, \text{ quindi } \beta \rightarrow \omega_2$$

$$\text{Caso 2: } \beta \in \omega_2 \quad \text{ma} \quad \beta < TH, \text{ quindi } \beta \rightarrow \omega_1 \rightarrow$$

$$P(\text{errore}) = \int_{-\infty}^{+\infty} P(\text{errore}|\beta)p(\beta)d\beta =$$

$$= \int_{-\infty}^{TH} P(\omega_2|\beta)p(\beta)d\beta + \int_{TH}^{+\infty} P(\omega_1|\beta)p(\beta)d\beta$$

è la f da

$$\min P(\text{errore}) = \int_{-\infty}^{+\infty} \min[P(\omega_1|\beta), P(\omega_2|\beta)]p(\beta)d\beta$$

minimizzare →

Il classificatore basato su questa regola si chiama classificatore a errore minimo.

Generalizziamo ora al caso di  $M$  classi  $w_1, w_2, \dots, w_M$  e un evento da classificare rappresentato da un vettore di

dimensione  $N$ :  $\beta = [\beta_1, \beta_2, \dots, \beta_N]$ . Supponiamo note le probabilità a priori  $P(\omega_i) \quad i=1, \dots, M$  con  $\sum_{i=1}^M P(\omega_i) = 1$  e le ddp  $P(\beta|\omega_i) \quad \text{con } i=1, \dots, M$ .

Assumiamo che ad ogni decisione  $\beta \rightarrow w_i$  seguì un'azione  $a_i$  e si definisce come  $\lambda(a_i|w_j)$  il costo dell'azione  $a_i$  quando la classe vera di appartenenza è  $w_j$ . Essendo  $P(w_j|\beta)$  la probabilità che la classe a cui appartiene  $\beta$  sia  $w_j$ , il costo complessivo associato alla decisione  $a_i$  è rappresentato da

che viene chiamato rischio condizionato. Il rischio totale legato all'intero procedimento di decisione è dato da

$$R(a_i|\beta) = \sum_{j=1}^M \lambda(a_i|\omega_j)P(\omega_j|\beta)$$

La regola di decisione allora assegna l'elemento  $\beta$  alla classe che comporta rischio minore

$$\beta \rightarrow \omega_i \quad \text{se} \quad R(a_i|\beta) < R(a_j|\beta) \quad \text{con } j=1, \dots, M, \quad j \neq i$$

Ad esempio:

$\alpha_1$  = terapia/astensione dal lavoro       $\lambda_{12}$  = costo della terapia/astensione dal lavoro per un soggetto sano

$\alpha_2$  = nessuna cura       $\rightarrow \lambda_{21}$  = costo di escludere dalla terapia un soggetto affetto da influenza

$$R(\alpha_1 | \beta) = \lambda_{11} P(\omega_1 | \beta) + \lambda_{12} P(\omega_2 | \beta)$$

A cui seguono i costi condizionati:  $R(\alpha_2 | \beta) = \lambda_{21} P(\omega_1 | \beta) + \lambda_{22} P(\omega_2 | \beta)$ . La regola di decisione è quindi:

se  $R(\alpha_1 | \beta) < R(\alpha_2 | \beta)$  allora  $\beta \rightarrow \omega_1$

altrimenti  $\beta \rightarrow \omega_2$  esplicitando la funzione rischio, scegliere la classe  $w_1$  significa che la seguente disequazione è soddisfatta:

$$\frac{P(\omega_1 | \beta)}{P(\omega_2 | \beta)} > \frac{\lambda_{12} - \lambda_{22}}{\lambda_{21} - \lambda_{11}}$$

$R(\alpha_1 | \beta) = (\lambda_{11} P(\omega_1 | \beta) + \lambda_{12} P(\omega_2 | \beta)) < R(\alpha_2 | \beta) = \lambda_{21} P(\omega_1 | \beta) + \lambda_{22} P(\omega_2 | \beta)$  che equivale a

Il secondo membro dipende solo dai costi ed è detto rapporto di verosimiglianza.

$$\lambda(\alpha_i | \omega_i) = \lambda_i = \begin{cases} 0 & \text{se } i = j \\ 1 & \text{se } i \neq j \end{cases}$$

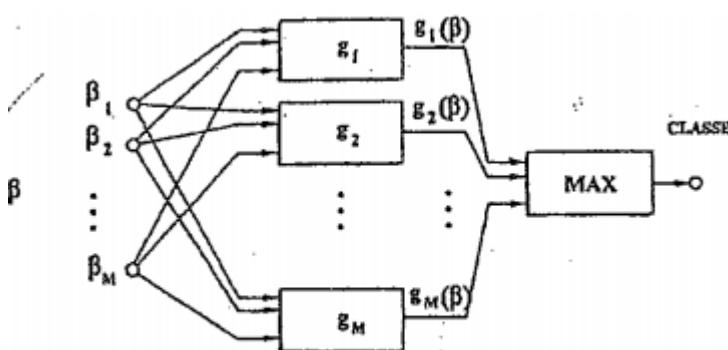
Il caso particolare in cui si scelgono i costi come si traduce nell'espressione del

$$R(\alpha_i | \beta) = \sum_{\substack{j=1 \\ j \neq i}}^M P(\omega_j | \beta)$$

rischio condizionato relativo all'azione  $\alpha_i$  dato da , ed essendo  $\sum_{j=1}^M P(\omega_j | \beta) = 1$ , il rischio diventa  $R(\alpha_i | \beta) = 1 - P(\omega_i | \beta)$ . Quindi scegliere la classe a minimo rischio equivale a scegliere la classe con la massima probabilità a posteriori. Quindi il classificatore a errore minimo può essere visto come il caso particolare del classificatore a rischio minimo per la scelta particolare dei pesi.

Possiamo pensare ai classificatori Bayesiani come ad un insieme di  $M$  funzioni in  $\beta$

$g_i(\beta)$  con  $i = 1, \dots, M$  ciascuna associata alle varie classi.



Si possono identificare varie funzioni discriminanti. Ad esempio per il classificatore di Bayes a errore minimo si identifica la funzione discriminante con la probabilità a posteriori, ponendo:

$$g_i(\beta) = P(\omega_i | \beta) \quad i = 1, \dots, M$$

Si possono anche considerare trasformazioni monotone per ottenere altre funzioni discriminanti, ad

esempio utilizzando la funzione logaritmo si ottiene  $g_i(\beta) = \ln [p(\beta | \omega_i)] + \ln [P(\omega_i)]$

Un classificatore, in generale, ha come effetto quello di suddividere lo spazio dei parametri  $R^N$  in regioni contigue, separate da iperpiani di dimensione  $(N-1)$  che sono il luogo dei punti per i quali vale

$$g_i(\beta) = g_k(\beta) \quad \text{per qualche } i \neq k$$

CLASSIFICATORE AD ERRORE MINIMO CON DDP GAUSSIANA

$$p(\beta | \omega_i) = \frac{1}{(2\pi)^{\frac{N}{2}} |\Sigma_i|^{\frac{1}{2}}} e^{-\frac{1}{2}(\beta - \mu_i)^T \Sigma_i^{-1} (\beta - \mu_i)}$$

Supponiamo  $p(\beta | \omega_i) \in N(\mu_i, \Sigma_i)$  →

$$g_i(\beta) = \ln [p(\beta | \omega_i)] + \ln [P(\omega_i)] \rightarrow$$

$$g_i(\beta) = -\frac{N}{2} \ln(2\pi) - \frac{1}{2} \ln |\Sigma_i| - \frac{1}{2} [(\beta - \mu_i)^T \Sigma_i^{-1} (\beta - \mu_i)] + \ln [P(\omega_i)]$$

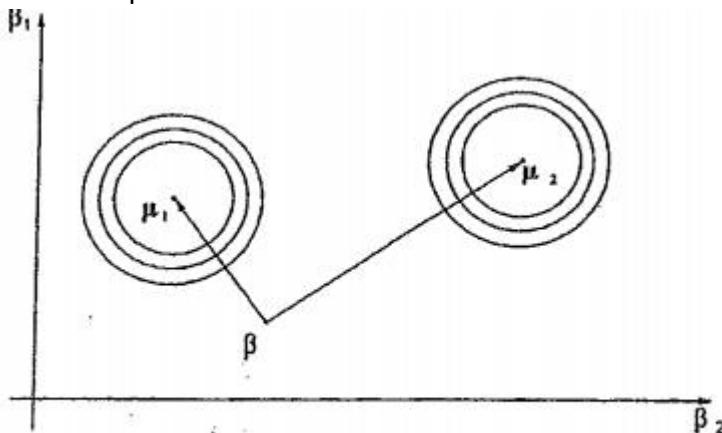
che sono funzioni quadratiche del vettore  $\beta$ .

Consideriamo il caso particolare in cui  $\Sigma_i = \sigma^2 I$  per  $\forall i$ , tutti i parametri hanno la stessa varianza e sono statisticamente indipendenti. Geometricamente questo significa che i valori delle varie classi si distribuiscono all'interno di ipersfere uguali per tutte le classi  $i$  i cui centri sono i punti di  $R^N$  individuati dai vettori delle medie  $\mu$ . Nella espressione della funzione discriminante i termini che non dipendono da una particolare classe non hanno un peso nel discriminare e quindi possono essere trascurati:

$$g_i(\beta) = -\frac{1}{2\sigma^2} [(\beta - \mu_i)^T (\beta - \mu_i)] + \ln [P(\omega_i)] = -\frac{1}{2\sigma^2} \|\beta - \mu_i\|^2 + \ln [P(\omega_i)]$$

$$g_i(\beta) = -\frac{1}{2\sigma^2} \|\beta - \mu_i\|^2$$

sotto l'ipotesi di classi equiprobabili diventa → classificatore di Bayes a minima distanza euclidea: associa il punto nello spazio dei parametri  $\beta$  alla classe la cui media si trova alla minima distanza da  $\beta$ .



#### Classificatore a minima distanza euclidea nel caso di due classi.

Questo classificatore viene detto classificatore di Bayes lineare in quanto le funzioni discriminanti possono

$$g_i(\beta) = -\frac{1}{2\sigma^2} \|\beta - \mu_i\|^2 \Rightarrow$$

essere espresse come funzioni lineari del vettore dei parametri  $\beta$ :

Trascurando il termine  $\beta^T \beta$  che essendo indipendente dalla classe non discrimina, si ottiene:

$$g_i(\beta) = -\frac{1}{2\sigma^2} (\beta^T \beta - 2\mu_i^T \beta + \mu_i^T \mu_i)$$

$$g_i(\beta) = a_i^T \beta + b_i$$

con:  $a_i = \frac{1}{\sigma^2} \mu_i$ ,  $b_i = -\frac{1}{2\sigma^2} \mu_i^T \mu_i$

Un altro caso particolare si ottiene se  $\Sigma_i = \Sigma \quad \forall i$ . In questo caso le classi hanno la stessa matrice di covarianza. Trascurando i termini che non dipendono dalla classe, tutti uguali tra loro, le funzioni

$$g_i(\beta) = -\frac{1}{2} [(\beta - \mu_i)^T \Sigma^{-1} (\beta - \mu_i)] + \ln [P(\omega_i)]$$

discriminanti hanno la forma: . Il termine  $[(\beta - \mu_i)^T \Sigma^{-1} (\beta - \mu_i)]$  è

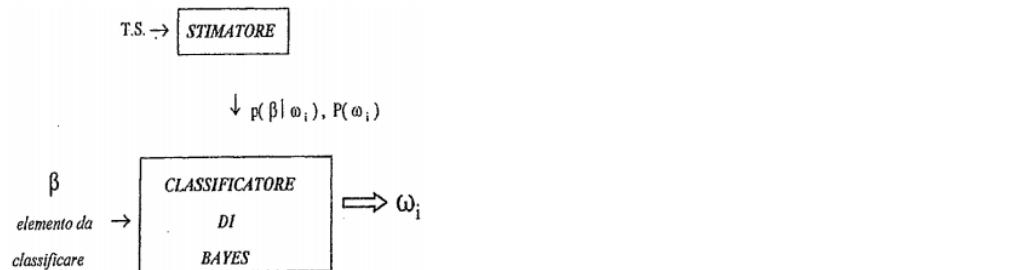
la distanza di Mahalanobis tra il vettore  $\beta$  da classificare ed i vettori delle medie  $\mu$ . Se le classi sono quiprobabili si ottiene il classificatore di Bayes a minima distanza di Mahalanobis.

### STIMA DELLA PROBABILITÀ A PRIORI

Il metodo di Bayes impone la conoscenza di  $P(\beta | \omega_i)$  e di  $P(\omega_i)$  con  $i=1,2,\dots,M$ . nella pratica raramente ci si trova in queste condizioni e si devono mettere a punto procedimenti di stima di tali grandezze da utilizzare in una fase precedente a quella del classificatore vero e proprio. Un possibile approccio è quello di utilizzare un insieme di campioni che costituiscono il training set: un insieme di elementi simili a quelli da studiare ma di cui si conoscono le classi → stima con supervisore

Quindi: a partire dal training set si stimano  $P(\beta | \omega_i)$  e  $P(\omega_i)$

Si assumono queste stime come valori veri di  $P(\beta | \omega_i)$  e  $P(\omega_i)$  e si mette a punto il classificatore



Spesso non è necessario stimare  $P(\omega_i)$  perché è già nota (ad es. l'incidenza di una patologia  $\omega_i$  può essere nota da dati a priori sulla popolazione). In alternativa, utilizzando il training set, si può stimare la probabilità delle singole classi dalle frequenze campionarie, cioè se  $Q$  è il numero di campioni del training

set e di questi  $Q_1 \in \omega_1, Q_2 \in \omega_2$  ecc le probabilità sono  $\hat{P}(\omega_1) = \frac{Q_1}{Q}, \hat{P}(\omega_2) = \frac{Q_2}{Q}, \dots, \hat{P}(\omega_M) = \frac{Q_M}{Q}$ .

La stima di  $P(\beta | \omega_i)$  è più complessa e varia a seconda dei casi.

### -STRUTTURA PROBILISTICA NOTA A PRIORI

Si tratta in questo caso di una stima parametrica. Se ad esempio  $(p(\beta | \omega_i) \in N(\mu_i, \Sigma_i))$ , i parametri da stimare sono il vettore media  $\mu$  e la matrice di covarianza  $\Sigma$  →  $\varphi = [\mu_1, \mu_2, \dots, \mu_N, \Sigma_{11}, \Sigma_{21}, \Sigma_{22}, \dots, \Sigma_{NN}]$  di dimensione  $\frac{N(N-1)}{2} + 2N$ .

Il problema si può semplificare assumendo che i campioni del training set siano vettori aleatori indipendenti. Si prende in considerazione una classe alla volta, dividendo il problema in  $M$  sottoproblemi. Consideriamo come esempio la stima di  $p(\omega_1 | \beta)$ : dati  $Q_1$  vettori indipendenti  $\beta_1, \beta_2, \dots, \beta_{Q_1}$  appartenenti alla classe  $\omega_1$ , nota la struttura della funzione densità di probabilità  $p(\beta | \omega_1)$ , si può scrivere che

$$P(\beta_1, \beta_2, \dots, \beta_{Q_1} | \omega_1, \varphi) = \prod_{j=1}^{Q_1} p(\beta_j | \omega_1, \varphi)$$

ora la stima parametrica di

massima verosimiglianza consiste nello scegliere i vari parametri in modo da massimizzare tale funzione:

$$\hat{\varphi} = \arg \left[ \max_{\varphi} P(\beta_1, \beta_2, \dots, \beta_{Q_1} | \omega_1, \varphi) \right].$$

Nel caso della distribuzione gaussiana si ottiene

$$\hat{\mu}_i = \frac{1}{d} \sum_{j=1}^d \beta_j;$$

$$\hat{\Sigma}_i = \frac{1}{d} \sum_{j=1}^d [(\beta_j - \hat{\mu}_i)(\beta_j - \hat{\mu}_i)^T]$$

ma non sempre esiste una soluzione analitica.

## -STRUTTURA PROBABILISTICA IGNOTA

Consideriamo una classe singola ( $w_1$ ): il vettore dei parametri  $\beta$  è un vettore aleatorio di dimensione  $N$  la cui descrizione statistica  $p(\beta)$  rappresenta l'oggetto di stima ( $\rightarrow$ sottointeso che cerco  $P(\beta|w_1)$ )

Consideriamo una regione  $R^N$  dello spazio; si può calcolare la probabilità  $Pr$  che  $\beta$  appartenga a questa

$$P_R = P[\beta \in R] = \int_R p(\beta) d\beta$$

regione: . Se  $p(\beta)$  è continua,  $Pr$  può essere intesa come il valore medio di  $p(\beta)$

$$P_R = \int_R p(\beta) d\beta \approx \bar{p}_R, \quad V \Rightarrow \bar{p}_R \approx \frac{P_R}{V}$$

nell'intervallo  $R$  moltiplicato per il volume della regione e quindi:

Devo stimare  $Pr$  e per questo utilizzo il T.S. la probabilità che  $k$  elementi del T.S. appartengano alla regione

$R$  ha distribuzione binomiale ed è:  $Prob\left[\begin{array}{l} k \text{ elementi} \in R \\ (Q_1 - k) \text{ elementi} \notin R \end{array}\right] = \binom{Q_1}{k} (P_R)^k (1 - P_R)^{Q_1 - k}$ . Si ha  $E[k] = Q_1 P_R$  e quindi

$P_R = \frac{E[k]}{Q_1}$ . Contando quanti elementi del T.S. appartengono a  $w_1$  e cadono nella regione  $R$ , indicando con  $k$

tale numero si può ottenere  $\hat{P}_R = \frac{k}{Q_1} \rightarrow \hat{P}_R(\beta) = \frac{k}{Q_1 V}$ . In questo modo si converge in probabilità verso il valore medio della regione  $p(\beta)$ . Per ottenere una stima puntuale di  $p(\beta)$  si deve scegliere  $V \rightarrow 0$  ma ciò comporta  $k=0$ . C'è dunque un errore di stima dovuto al fatto che il numero di campioni è limitato e un errore di approssimazione dovuto al fatto che le regioni hanno volumi finiti.

Tre condizioni sono necessarie perché  $\hat{P}_R(\beta)$  converga a  $p(\beta)$ : 1.  $\lim_{Q_1 \rightarrow \infty} V_Q = 0$ , 2.  $\lim_{Q_1 \rightarrow \infty} k_Q = \infty$ , 3.  $\lim_{Q_1 \rightarrow \infty} \frac{k_Q}{Q_1} = 0$ .

Ci sono due modi per ottenere sequenze di regioni che soddisfano queste condizioni:

- 1) METODO DI PARZEN data l'osservazione  $\beta \in w_1$  è possibile dire che  $p(\beta|w_1)$  è non nulla e, essendo continua, non dovrebbe essere nulla neanche nelle vicinanze. Si esprime questa informazione con una funzione detta kernel che assume il valore massimo in corrispondenza dell'osservazione  $\beta_i$  e via

via scema. In generale il kernel è esprimibile con  $\varphi(\beta, \beta_i^{(k)})$  che soddisfa  $\int_{R^N} \varphi(\beta, \beta_i^{(k)}) d\beta = 1$

$$\varphi(\beta, \beta_i^{(k)}) = \begin{cases} \frac{1}{h} & |\beta - \beta_i^{(k)}| \leq \frac{h}{2} \\ 0 & \text{altrove} \end{cases}$$

ad esempio nel caso  $\dim \beta = 1$ , il kernel rettangolare ha l'espressione dove  $h$  è un parametro da scegliere (ogni kernel ha infatti un grado di libertà)

$$\hat{p}(\beta|w_1) = \frac{1}{Q_1} \sum_{i=1}^{Q_1} \varphi([\beta, \beta_i^{(1)}])$$

la stima della ddp è allora che dipende dalla scelta del tipo di kernel e da  $h$ . con valori elevati di  $h$  si ha bassa risoluzione, con valori bassi si ha variabilità statistica.

- 2) METODO K-NN con il metodo di parzen la scelta del volume delle celle è un punto delicato che incide sull'andamento della stima di  $p(\beta)$ . Con questo metodo si scegli invece il volume della cella in modo che al suo interno cadano in ogni caso un certo numero  $k$  di campioni del TS, ad esempio  $k = \sqrt{Q}$ .

allora la stima di  $p(\beta)$  è ancora data da  $\hat{p}(\beta) = \frac{k/Q}{V}$  dove  $k$  e  $Q$  sono costanti mentre  $V$  dipende da  $\beta$ . Quindi, dato un  $\beta$  da classificare, consideriamo una regione di volume  $V$  composta da  $K$  elementi del TS. Indichiamo con

$k_1$  il numero di elementi che appartengono a  $w_1$

....

$k_m$  numero di elementi di  $w_m$

$Q_1$  elementi di  $w_1$

Vale ovviamente  $\sum_{i=1}^M k_i = k$

...

Qm elementi di  $w_m$

$$\hat{p}(\beta | \omega_i) = \frac{k_i}{Q_i V} \quad \text{con } i = 1, \dots, M$$

Si ottiene allora . Applicando il classificatore ad errore minimo,

$$g_i(\beta) = p(\beta | \omega_i) P(\omega_i)$$

Sostituendo si ottiene  $g_i(\beta) = \frac{k_i}{Q_i V} \cdot \frac{Q_i}{Q_{tot}} = \frac{k_i}{V Q_{tot}}$  ed eliminando i termini che non discriminano la funzione:  $g_i(\beta) = k_i$  che da la regola di decisione:

$$\beta \rightarrow \omega_i \quad \text{se} \quad k_i > k_j \quad \text{per} \quad \forall j = 1, \dots, M \quad \text{con} \quad j \neq i.$$

Cioè si assegna  $\beta$  alla classe che presenta più campioni nella regione in cui si trova  $\beta$  stesso. Questo metodo usa in continuazione il TS (a differenza di altri metodi che lo usano solo in fase di progettazione) e quindi questo costringe a mantenere il TS in memoria. Questo metodo è quindi vantaggioso se si deve aggiornare il TS perché non è necessario modificare l'algoritmo di decisione.

## SELEZIONE DEI PARAMETRI

All'aumentare dei parametri aumenta la complessità dello stimatore. Bisogna quindi, nella fase di selezione, partire dai molti parametri estratti e ridurli per stabilire quali siano discriminanti.

### Selezione:

dal vettore di parametri  $\beta = [\beta_1, \dots, \beta_N]^T$

ad un nuovo vettore di parametri  $y = [y_1, \dots, y_D]^T$

con  $N = \dim(\beta) < D = \dim(y)$

Selezione diretta :  $y_i = \beta_i$

Trasformazioni :  $y_i = F(\beta_1, \dots, \beta_N)$

-TRASFORMAZIONE LINEARE DI FISHER è una trasformazione lineare dal vettore  $\beta = [\beta_1, \dots, \beta_N]^T$  a

$y = [y_1, \dots, y_D]^T$  con  $D = M - 1$ . ( $M = \#$  classi). La trasformazione è scelta in modo che  $y$  sia la scelta ottima rispetto ad un criterio tra le classi.

### Caso 1 $M=2$

$$y = \rho_1 \beta_1 + \rho_2 \beta_2 + \dots + \rho_N \beta_N = \rho^T \beta$$

$y$  è uno scalare che rappresenta la proiezione di  $\beta$  su una retta che ha  $[\rho_1 \rho_2 \dots \rho_N]^T$  come versore.

Definiamo:

Classe n° elementi v.medio dispersione intra-

$$\begin{array}{cccc} \omega_1 & Q_1 & \mu_1 & W_1 \\ \omega_2 & Q_2 & \mu_2 & W_2 \end{array}$$

$$\text{dispersione inter} = B = (\mu_1 - \mu_2)(\mu_1 - \mu_2)^T$$

Per le proiezioni

Classe n° elementi v.medio dispersione intra-

$$\begin{array}{cccc} \omega_1 & Q_1 & \rho^T \mu_1 & \rho^T W_1 \rho \\ \omega_2 & Q_2 & \rho^T \mu_2 & \rho^T W_2 \rho \end{array}$$

$$\text{dispersione inter} = b = \rho^T B \rho$$

L'indice di separabilità è  $J = \rho^T B \rho / \rho^T (W_1 + W_2) \rho$  che assume valore elevato se le medie delle due classi sono lontane tra loro e i valori nelle due classi sono concentrati attorno ai valori medi.

**Problema:** trovare  $\rho$  che rende massimo  $J$

**Soluzione:**  $\rho$  è autovettore della matrice  $W^{-1}B$

quindi

$$W^{-1}B \rho = \lambda \rho$$

ovvero  $W^{-1}(\mu_1 - \mu_2)(\mu_1 - \mu_2)^T \rho = \lambda \rho$

quindi  $\rho$  nella direzione di  $W^{-1}(\mu_1 - \mu_2)$

In quanto  $(\mu_1 - \mu_2)^T \rho$  è uno scalare

di  $\rho$  interessa la direzione e non il modulo!

Si può generalizzare al caso di M classi

$$\mathbf{y} = T^T \beta$$

$$J = \det(T^T B T) / \det(T^T (W_1 + \dots + W_M) T)$$

$$\text{oppure } J = \text{tr}(T^T B T) / \text{tr}(T^T (W_1 + \dots + W_M) T)$$

**Problema:** trovare la matrice  $T$  che rende massimo  $J$

**Soluzione:** nel primo caso,  $T$  ha come colonne gli autovettori della matrice  $W^{-1}B$

-TRASFORMAZIONE K-L è una trasformazione lineare dal vettore  $\beta = [\beta_1, \dots, \beta_N]^T$  a  $\mathbf{y} = [y_1, \dots, y_N]^T$  con  $\mathbf{y} = T \beta$ .

e  $T$  matrice che ha come righe vettori ortonormali  $\Phi$ , per cui  $T^T = T^{-1}$  ovvero

$$\beta = T^{-1} \mathbf{y} = T^T \mathbf{y} = \Phi_1 y_1 + \Phi_2 y_2 + \dots + \Phi_N y_N$$

$$y_1 = \Phi_1^T \beta$$

.....

$$y_N = \Phi_N^T \beta$$

il problema è trovare la trasformazione  $T$  in modo che,

definito con  $\beta_d$  l'approssimazione ottenuta con i primi  $D$  termini,  $\beta_d = \Phi_1 y_1 + \Phi_2 y_2 + \dots + \Phi_D y_D + \Phi_{D+1} b_{D+1} + \dots + \Phi_N b_N$ ,

sia minimo l'errore quadratico medio di approssimazione:  $\text{Errore} = E[(\beta - \beta_d)^2]$

**Soluzione:**

Indicata con  $\Sigma_\beta$  la matrice di varianza-covarianza di  $\beta$  e con  $\lambda_1, \lambda_2, \dots, \lambda_N$  i suoi autovalori ordinati da quello di valore più elevato via via a quelli di valore più basso  $\lambda_1 > \lambda_2 > \dots > \lambda_N$ .

$\Phi_1, \Phi_2, \dots, \Phi_N$  sono gli autovettori corrispondenti

Si può dimostrare che:

$$\text{Errore} = \lambda_{D+1} + \dots + \lambda_N$$

$$\Sigma_y = T \Sigma_\beta T^{-1} = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_N)$$

Le componenti principali sono tra

$$\text{Tr}(\Sigma_y) = \text{tr}(\Sigma_\beta)$$

loro indipendenti!!

Pertanto:

1° componente principale:

$y_1 = \Phi_1^T \beta$  è lo scalare che approssima al meglio  $\beta$   
ha varianza pari a  $\lambda_1$   
spiega  $\lambda_1 / (\lambda_1 + \lambda_2 + \dots + \lambda_N)$  di varianza totale

1° e 2° componente principale:

$y_1 = \Phi_1^T \beta$   
 $y_2 = \Phi_2^T \beta$  è il vettore di dim 2 che approssima il meglio  $\beta$   
ha varianza pari a  $\lambda_1 + \lambda_2$   
spiega  $\lambda_1 + \lambda_2 / (\lambda_1 + \lambda_2 + \dots + \lambda_N)$  di varianza totale

Ecc..

## -SELEZIONE DIRETTA

- RICERCA ESAUSTIVA:
- 1) si calcolano tutte le possibili combinazioni di D parametri presi da  $\beta$
  - 2) per ognuno si calcola l'accuratezza (1-prob. Errore) del classificatore
  - 3) si sceglie la combinazione che massimizza l'accuratezza

Si hanno a disposizione molti parametri e quindi il numero di combinazioni possibili può risultare elevatissimo e la messa a punto dei classificatori può essere onerosa da un punto di vista computazionale. Si può allora semplificare il punto 2 con una ricerca sequenziale in avanti:

- 1) si sceglie  $D=1$  e allora i parametri sono  $\beta_1, \beta_2, \dots, \beta_N$  e si calcolano gli indici  $J_1, J_2, \dots, J_N$
- 2) si scegli la combinazione che discrimina meglio (es.  $\beta_1$ )
- 3) incremento  $D$  ma scelgo solo le combinazioni che hanno come primo elemento quello scelto al punto 2
- 4) mi fermo quando trovo un indice soddisfacente

## VALIDAZIONE DEL CLASSIFICATORE

In sede di progetto, prima di passare al suo utilizzo, è molto importante la validazione del classificatore, ovvero avere una misura delle sue prestazioni attraverso una stima dell'errore di classificazione e poter comparare diversi classificatori.

Si può calcolare l'errore in modo analitico, come fatto per il classificatore ad errore minimo, ma questo può risultare difficoltoso.

La situazione ottimale è poter disporre di due TS: il primo si utilizza per mettere a punto il classificatore mentre il secondo per stimare l'errore di classificazione. È possibile successivamente calcolare la percentuale di classificazioni corrette e costruire delle tabelle da utilizzare per stimare la bontà del classificatore. Per valutare la bontà di tale stima si esegue il test con  $n$  elementi e se come risultato si

ottiene un numero di classificazioni errate pari a  $K$ , la stima della probabilità d'errore è data da  $\hat{P}(\text{errore}) = \frac{k}{n}$ . Se chiamiamo  $p$  la vera probabilità d'errore, la probabilità che su  $Q$  elementi,  $k$  siano classificati in modo

$$\text{errato è } P\left[\begin{array}{c} k \text{ errati} \\ (Q - k) \text{ corretti} \end{array}\right] = \binom{Q}{k} p^k (1-p)^{n-k} \rightarrow \frac{E[k]}{Q} = p \rightarrow \hat{p} = \frac{k}{Q} \quad (\text{la stima migliora al crescere di } Q)$$

A partire da questa descrizione statistica si può costruire un monogramma e ottenere gli intervalli di confidenza delle stime.

La situazione idela è quindi quella di avere molti dati da dividere in 2 TS. Se però ne ho pochi, devo scegliere tra l'avere una buona stima dell'errore e un classificatore poco attendibile o viceversa.

Una tecnica spesso usata nel caso di TS non particolarmente numeroso è basata sul metodo leave one out che dato un S composto da  $\beta_1, \beta_2, \dots, \beta_Q$ , considera un primo TS con  $\beta_2, \dots, \beta_Q$  e usa  $\beta_1$  per testarlo. Fatto questo si rispetta l'operazione con  $\beta_1, \beta_3, \dots, \beta_Q$  e si usa  $\beta_2$  per il test. Si procede così per  $Q$  volte e si

contano quanti  $k$  elementi sono stati classificati male e si ottiene  $\hat{p} = \frac{k}{Q}$ . Lo svantaggio è che si stimano  $Q$  classificatori molto simili ma non uguali.