

INSEGNAMENTO DI ANALISI DI DATI BIOLOGICI
LM in BIOINGEGNERIA – A.A.2023/2024

HOMEWORK 1

| | | | |
|---|-------------------------|------------------------------------|-------------------------|
| COGNOME: CASARIN | NOME: ALESSANDRO | MATRICOLA: 2086857 | DATA: 19/12/2023 |
| e-mail: alessandro.casarin.4@studenti.unipd.it | | allegati: casarin_h1_74.zip | |

ESERCIZIO H1_74

Nel file `smbg.mat` sono contenuti dati di Self Monitoring Blood Glucose (SMBG) di un soggetto diabetico. Le misurazioni SMBG sono state eseguite 3 volte al giorno (equispaziate, sempre alla stessa ora) per un periodo della durata di un mese (non sono presenti missing samples). Analizzando graficamente i dati, il diabetologo che ha in cura questo soggetto diabetico ritiene che possa esistere una ciclicità giornaliera nelle misurazioni, che verosimilmente ruota attorno ad un trend mensile. Possiamo pensare che la misurazione del valore glicemico possa essere strutturata così:

$$y(t) = m(t) + g(t) + v(t)$$

dove $m(t)$ rappresenta il trend mensile, $g(t)$ rappresenta la componente ciclica giornaliera e $v(t)$ è il rumore di misura. L'obiettivo di questo homework è quello di implementare uno script per separare le componenti $m(t)$ e $g(t)$ per cercare di confermare o meno se esistono una ciclicità giornaliera ed un trend mensile.

Estrarre il trend di lungo periodo usando una tecnica di approssimazione dati kernel smoother (kernel esponenziale).

Per ogni giornata ricostruire il relativo trend giornaliero (usare anche una quarta misura, l'ultima del giorno precedente o la prima del giorno successivo) e costruire quindi, dalle realizzazioni ottenute, le relative bande di variabilità.

SVOLGIMENTO

Introduzione e metodologia

Per la risoluzione del quesito, è necessario fare delle ipotesi nell'interpretazione dei dati: dal momento che i dati relativi alle misure di glucosio sono contenuti nel vettore `ys` e non è stato fornito un vettore relativo agli istanti di campionamento, ho ipotizzato che le misurazioni siano state fatte a partire dalle ore 00:00 del primo giorno, ogni 8 ore, per tutti e 30 i giorni del mese.

Il problema è stato affrontato in 3 passi:

- 1) Estrazione del trend mensile tramite approssimazione con kernel smoother di tipo esponenziale.
Non avendo indicazioni sulla funzione kernel da utilizzare, ho utilizzato una funzione composta da un esponenziale crescente per valori negativi dell'asse delle ascisse e un esponenziale decrescente per valori positivi dell'asse delle ascisse.
- 2) Estrazione del trend giornaliero tramite approssimazione parametrica con funzioni polinomiali di terzo grado applicata sulle tre misure della giornata e sulla prima misura del giorno successivo.

- 3) Costruzione delle bande di variabilità costituite dal 5° e 95° percentile delle realizzazioni dei trend giornalieri.

Presentazione del codice

Per realizzare quanto richiesto, ho realizzato due M-file, *h1_74.m* e *kernelExp.m*, disponibili nell'archivio *casarin_h1_74.zip*.

Lo script *h1_74.m* svolge sequenzialmente i punti descritti nell'introduzione, mostrando risultati differenti per valori diversi di alcuni parametri: λ per il kernel smoother, m (grado delle funzioni polinomiali) per l'approssimazione parametrica. I parametri più adatti sono stati poi scelti come soluzione dell'esercizio.

La function *kernelExp.m* assiste lo script principale restituendo i "pesi" da utilizzare per l'approssimazione con kernel esponenziale per ogni misura di glucosio. Per ottenere il trend mensile, per ogni misura viene richiamata questa function alla quale vengono passati come parametri il vettore dei tempi, l'istante di tempo relativo alla misura e il parametro λ che determina l' "apertura" della funzione kernel; la function restituisce il vettore dei "pesi" che verranno utilizzati per fare una media pesata e ottenere il valore desiderato.

La funzione implementata per il kernel esponenziale è riportata in Figura 1.

```
function weights = kernelExp(t, t_star, lambda)

t_star_idx = find(t==t_star);
t_sx = t(1:t_star_idx);
t_dx = t(t_star_idx:end);

weights = zeros(size(t));
weights(1:t_star_idx) = exp((t_sx - t_star) / lambda);
weights(t_star_idx:end) = exp(-(t_dx - t_star) / lambda);

end
```

Figura 1. Function *kernelExp*. Parametri: t è il vettore dei tempi, t_star è l'istante in cui centrare il kernel, $lambda$ è il parametro che determina l' "apertura" del kernel.

Il codice per l'estrazione del trend mensile è stato scritto facendo riferimento agli esercizi svolti nel laboratorio2. Dal momento che nella consegna non sono state fornite indicazioni sulla natura del rumore di misura, si è ipotizzato che quest'ultimo fosse a campioni scorrelati e varianza costante; la stima per l'approssimazione parametrica è stata fatta quindi tramite uno stimatore ai minimi quadrati non pesati. Le stime sono state calcolate per ogni giornata utilizzando una griglia virtuale composta da 24 ore, infine il trend è stato stimato facendo una media di tutte le realizzazioni delle varie giornate.

Risultati e discussione

L'approssimazione tramite kernel smoother esponenziale è stata svolta con diversi valori del parametro λ ; dai risultati ottenuti è stato scelto il valore $\lambda=50$ perché, relativamente al tipo di kernel da me implementato, sembra essere il miglior compromesso tra approssimazione e integrità dei dati originali.

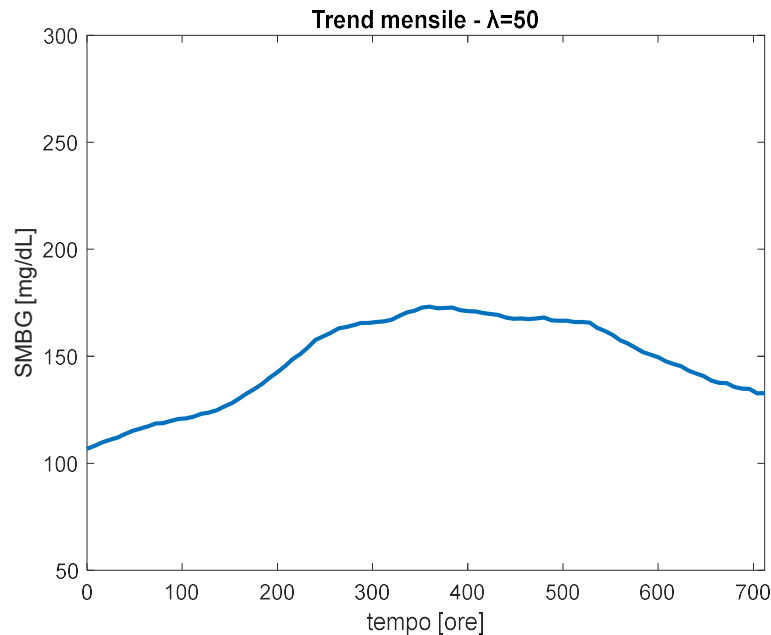


Figura 2. Risultato dell'approssimazione con kernel smoother di tipo esponenziale, con parametro $\lambda=50$. L'asse dei tempi è rappresentato in ore (a partire dalla prima misura effettuata all'ora 0).

Per quanto riguarda l'estrazione del trend giornaliero, le indicazioni sono di utilizzare le tre misure di glucosio giornaliere ed una quarta misura (ho scelto la prima del giorno successivo). Queste quattro misure sono state salvate in vettori e al fine di avere vettori organizzati allo stesso modo, il trend giornaliero è stato calcolato senza tenere conto delle misure relative all'ultimo giorno del mese, poiché non vi è a disposizione una misura successiva a tale giornata. Inoltre, al vettore di ogni giornata è stato sottratto il relativo vettore del trend mensile, in questo modo è possibile analizzare la ciclicità giornaliera senza il contributo di quella mensile.

L'approssimazione parametrica è stata svolta con funzioni polinomiali di ordine 2 e 3.

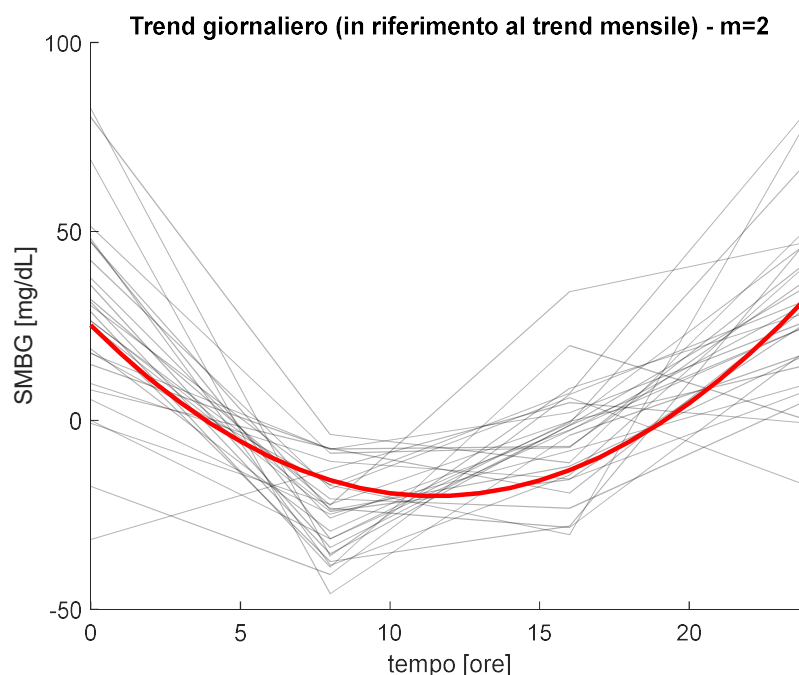


Figura 3. Risultato dell'approssimazione parametrica con funzioni polinomiali di ordine 2. Le linee nere sovrapposte rappresentano le misure originali dei vari giorni (in riferimento al trend mensile). La linea rossa rappresenta il trend giornaliero.

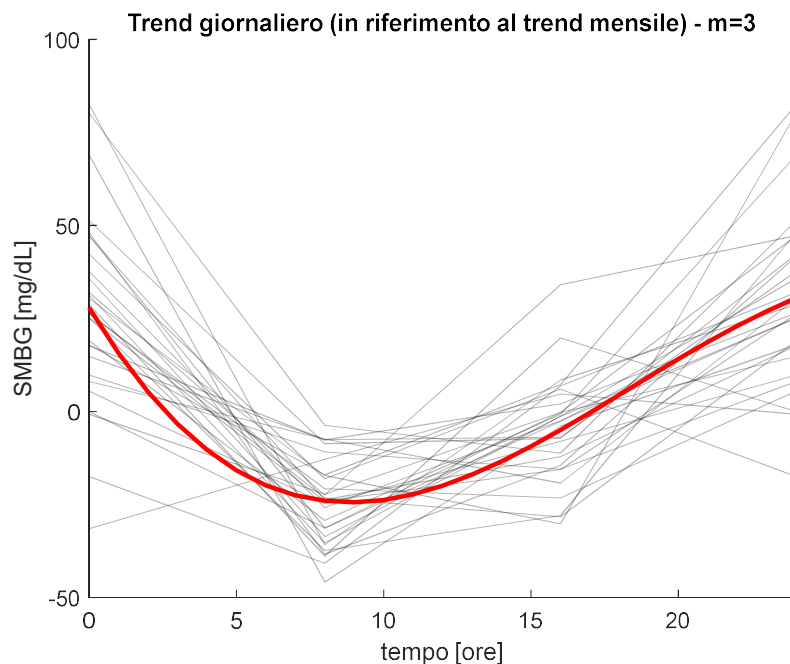


Figura 4. Risultato dell'approssimazione parametrica con funzioni polinomiali di ordine 3. Le linee nere sovrapposte rappresentano le misure originali dei vari giorni (in riferimento al trend mensile). La linea rossa rappresenta il trend giornaliero.

Come si nota dalle Figure 3 e 4, sembra esserci una ciclicità giornaliera e questa viene approssimata in modo migliore utilizzando polinomi di grado 3. Infatti, i residui ottenuti con grado 3 sono più piccoli rispetto a quelli ottenuti con grado 2.

A partire dalle realizzazioni ottenute per le diverse giornate, le bande di variabilità sono state costruite cercando il 5° e il 95° percentile della distribuzione dei valori per ogni ora relativa alla griglia virtuale (24 ore totali).

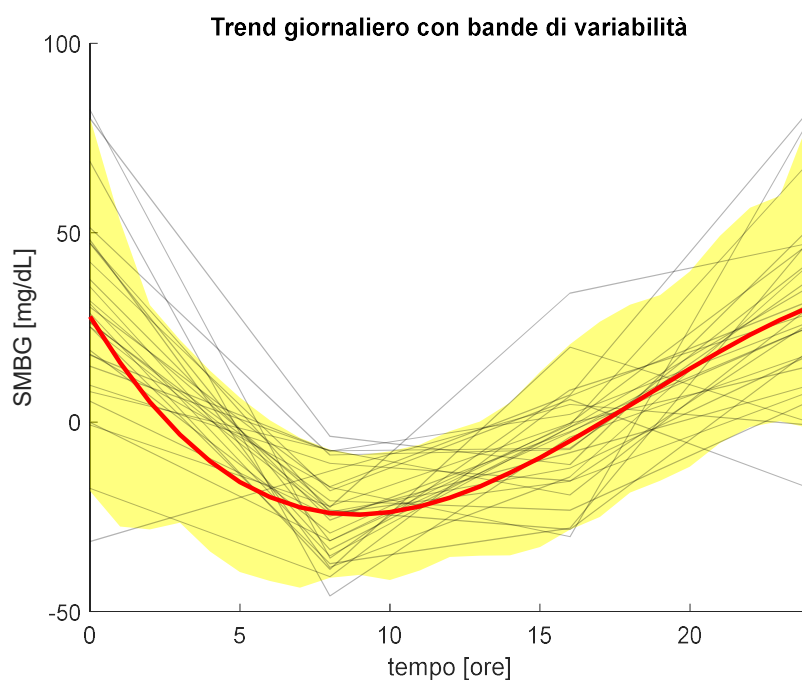


Figura 5. Risultato dell'approssimazione parametrica con funzioni polinomiali di ordine 3. Le bande gialle rappresentano le bande di variabilità.

I “take-home messages” da evidenziare sono i seguenti:

- È fondamentale fare delle ipotesi iniziali in caso di mancanza di informazioni specifiche nella consegna, in questo caso è stato necessario fare ipotesi sull’orario di acquisizione giornaliero delle misure e sulla natura dell’errore di misura.
- È consigliabile, inoltre, fare diversi esperimenti con metodi di approssimazione diversi e con una griglia di parametri variabile; in questo modo si avrà a disposizione più di una soluzione e si potrà scegliere quella più adeguata.