

# AUTOCORRELAZIONE ed ERRORI

nella SIMULAZIONE di processi di Markov

- In senso molto generale, ogni volta che facciamo un conto numerico, vorremmo essere in grado di riportare (un po' come si fa per un esperimento in laboratorio ...) due cose:

## RISULTATI ed ERRORI !

Spesso in laboratorio abbiamo dati indipendenti. Ci troviamo in un CASO SEMPLICE!  
Per PROCESSI di MARKOV le COSE saranno PIU' COMPLICATE!

Quando abbiamo un CAMPIONE di  $N$  DATI INDIPENDENTI  
(diciamo che questo campione sia contenuto in un VETTORE  $dd$  con  $\text{length}(dd) = N$ )

RISULTATO  $\longleftrightarrow \text{mean}(dd)$

ERRORE  $\longleftrightarrow \text{std}(dd) / \sqrt{\text{length}(dd)} = \frac{\text{std}(dd)}{\sqrt{N}}$

Questo è spesso il caso della raccolta di dati da misure in laboratorio: siccome i dati generati sono INDIPENDENTI (nell'ipotesi, ovviamente, che si tratti di misure dirette senza errori sistematici) posso applicare i risultati che discendono dalla LEGGE dei GRANDI NUMERI:

- Se osservo un fenomeno descritto da una certa distribuzione, la MEDIA ARITMETICA dei dati (notate che "parliamo Matlab": `df = mean`) è una STIMA CORRETTA e NON DISTORTA del VALORE MEDIO sulla distribuzione
- la VARIANZA della MEDIA è uguale alla VARIANZA del PROCESSO divisa per il numero di prove (cardinalità del mio campione)
- Poiché prendiamo come stima dell'ERRORE lo SCARTO (radice quadrata della Varianza), denominato  $\sigma$ , lo scarto della MEDIA è dunque  $\frac{\sigma}{\sqrt{N}}$ , dove  $\sigma$  è lo scarto del processo, come misurato dalla funzione "std" di Matlab.

N.B. L'unica sottigliezza in questo caso è nella determinazione NON DISTORTA della  $\sigma$ . (parliamo di unbiased estimators): poiché ci pensa Matlab a fare la cosa corretta, possiamo disinteressarcene...

→ COME DOBBIAMO CALCOLARE gli ERRORI nel caso di SIMULAZIONI di PROCESSI DI MARKOV?

La conclusione generale cui arriveremo è che la applicazione ingenua della semplice ricetta vista sopra (valida per un campione di dati indipendenti) ci conduce a SOTTOSTIMARE SISTEMATICAMENTE GLI ERRORI!

Un po' di notazioni:

- un processo è una successione di CONFIGURAZIONI del SISTEMA  $\{X_t | t=1 \dots T\}$ ;
- se misuro una funzione  $f(X)$  ottengo la successione  $\{f_t = f(X_t)\}$ ;
- CHIAMO  $\mu_f \equiv \langle f_t \rangle_\pi = \sum_x f(x) \pi_x$  il VALORE MEDIO di  $f_t$  sulla distribuzione asintotica  $\pi$

definisco FUNZIONE di AUTOCORRELAZIONE

$$\| C_{ff}(t) \equiv \langle f_s f_{s+t} \rangle_\pi - \langle f_s \rangle_\pi \langle f_t \rangle_\pi = \langle f_s f_{s+t} \rangle - \mu_f^2 \|$$

N.B. Correla due misure "istanti"  $t$  nel tempo della simulazione"  
(è funzione della distanza fra i due tempi! cfr la notazione:  $S$  non conta...)

N.B. 2 Mi aspetto che se  $C_{ff}(t^*) \simeq 0$  allora a distanza temporale  $t^*$  le mie misure siano decorrelate (al solito: valor medio del prodotto in queste condizioni  $\simeq$  prodotto dei valori medi...)

N.B. 3 Notare il pedice  $ff$ : la  $f_z$  di AUTOCORRELAZIONE è diversa per diverse funzioni  $f$ ...

Notare che  $C_{ff}(t) = \langle f_s f_{s+t} \rangle_\pi - \mu_f^2 = \sum_{xy} f(x) (W_{xy}^{tt} \pi_y - \pi_x \pi_y) f(y)$

e dunque  $C_{ff}(t) \xrightarrow[t \rightarrow \infty]{} 0$  purché  $W_{xy}^{tt} \xrightarrow[t \rightarrow \infty]{} \pi_x$  !

Quindi SU TEMPI LUNGI le MISURE RACCOLTE simulando il processo SI DECORRELANO: ma "QUANTO CI METTONO"?

Scrivo una FUNZ. di AUTOCORRELAZIONE NORMALIZZATA

$$\rho_{ff}(t) = \frac{C_{ff}(t)}{C_{ff}(0)} \quad (\text{ovvero } \rho_{ff}(0)=1)$$

TIPICAMENTE  $\rho_{ff}(t) \sim e^{-t/\tau}$

il che non ci stupisce: ricordate che perderemo memoria della distribuzione di probabilità iniziale  $P^{(0)}$  in  $P^{(N)} = W^N P^{(0)} \rightarrow \pi + O(|\hat{\lambda}|^N)$  ?  
 $N$  grande

$\hat{\lambda}$  è l'autovettore in modulo ( $\hat{\lambda} \neq 1$ ) più vicino a uno. Ora  $N$  è il nostro  $t$  e

$$|\hat{\lambda}|^N \rightarrow |\hat{\lambda}|^t = e^{\ln |\hat{\lambda}|^t} = e^{+t \ln |\hat{\lambda}|}$$

Poiché  $|\hat{\lambda}| < 1$ ,  $\ln |\hat{\lambda}| < 0$  e  $\ln |\hat{\lambda}| = -|\ln |\hat{\lambda}||$ , ovvero  $|\hat{\lambda}|^t = e^{-t |\ln |\hat{\lambda}||}$

che posso scrivere  $|\hat{\lambda}|^t = e^{-\frac{t}{\tau_{\text{exp}}}} \equiv e^{-t/\tau}$  ...

DEFINISCO IL **TEMPO di AUTOCORRELAZIONE ESPONENZIALE**

$$\tau_{\text{exp}, f} = \limsup_{t \rightarrow \infty} \frac{t}{-\log |\rho_{ff}(t)|}$$

e  $\tau_{\text{exp}} = \sup_f \tau_{\text{exp}, f}$  rappresenta dunque il TEMPO di RILASSAMENTO

del MODO PIU' LENTO del SISTEMA.

Questo comporta che devo attendere un tempo ALMENO  $\pm \tau_{\text{exp}}$  per considerare il SISTEMA TERMALIZZATO...

... ma questo non ci dice ancora nulla sugli errori...

Definisco ora il **TEMPO di AUTOCORRELAZIONE INTEGRATO**

$$\tau_{int,f} = \frac{1}{2} \sum_{t=-\infty}^{\infty} \rho_{ff}(t) = \frac{1}{2} + \sum_{t=1}^{\infty} \rho_{ff}(t)$$

(con questa normalizzazione - cfr il fattore  $\frac{1}{2}$  -  $\tau_{int,f} \sim \tau_{exp,f}$  se  $\rho_{ff}(t) \sim e^{-t/2} \dots$ )

QUANTO VALE LA VARIANZA ALL'EQUILIBRIO di

$$\bar{f} \equiv \frac{1}{N} \sum_{t=1}^N f_t \quad ?$$

Questa è la quantità che misuro, purché so che  $\bar{f} \xrightarrow{N \rightarrow \infty} \mu_f$ . In sostanza,

da  $\bar{f}$  leggo il RISULTATO e la RADICE QUADRATA (scarto) della sua VARIANZA mi dà la stima dell'ERRORE, che è quello che vado cercando!

$$\begin{aligned} \text{Var}_{\pi}(\bar{f}) &= \langle \bar{f}^2 \rangle - \mu_f^2 = \frac{1}{N^2} \langle \sum_{t,s=1}^N f_t f_s \rangle - \mu_f^2 \\ &= \frac{1}{N^2} \sum_{t,s=1}^N \left[ \langle f_t f_s \rangle - \mu_f^2 \right] \\ &= \frac{1}{N^2} \sum_{r,s=1}^N \langle f_f(r-s) \rangle = \frac{1}{N^2} \sum_{t=-N+1}^{N-1} (N-|t|) C_{ff}(t) \\ &= \frac{1}{N} \sum_{t=-N+1}^{N-1} \left( 1 - \frac{|t|}{N} \right) C_{ff}(t) \\ &= \frac{1}{N} 2 \langle f_f(0) \rangle \frac{1}{2} \sum_{t=-N+1}^{N-1} \left( 1 - \frac{|t|}{N} \right) \rho_{ff}(t) \end{aligned}$$

Ora, quando  $N \gg t, \tau$  (ovviamente è quello che voglio! molte misure!!)

$$\frac{1}{2} \sum_{t=-N+1}^{N-1} \left( 1 - \frac{|t|}{N} \right) \rho_{ff}(t) \sim \frac{1}{2} \sum_{t=-\infty}^{\infty} \rho_{ff}(t)$$

Il che vuol dire che per un numero di misure  $N \gg \tau$

$$\text{Var}_{\pi}(\bar{f}) \approx \frac{1}{N} (2 \tau_{\text{int},f}) C_{ff}(0)$$

dove  $C_{ff}(0)$  è la varianza di  $f$  all'equilibrio...

E' come dire che  $\text{Var}_{\pi}(\bar{f}) = \frac{1}{N_{\text{eff}}} \text{Var}_{\pi}(f)$  dove  $N_{\text{eff}} \equiv \frac{N}{2 \tau_{\text{int},f}}$

ovvero la varianza della media è quella della quantità che voglio misurare abbattuta NON di  $\frac{1}{N}$  ma di  $\frac{1}{N_{\text{eff}}}$ , dove  $N_{\text{eff}} < N$  per effetto di  $\tau_{\text{int},f}$ :

devo "spaziare le misure" per trovare quelle decorelate!

L'ERRORE che faremo (ad esempio per la lunghezza della coda all'equilibrio)

è allora

$$\frac{\text{std}(\text{dd})}{\sqrt{N/(2 \tau_{\text{int},f})}}$$

dove  $\text{dd}$  è un campione di  $N$  misure di  $f$  sul processo.

Sul campione stesso (in generale sul campione più grande possibile) avrò per prima cosa determinato  $\tau_{\text{int},f}$ , il che chiede a suo volta di misurare la funzione di autocorrelazione di  $f$ .

## IN SOSTANZA

- PRIMA di PRENDERE MEDIE devo FAR TERMALIZZARE il SISTEMA (per tempi di ordine  $\tau_{\text{exp}}...$ )

- UNA VOLTA ALL'EQUILIBRIO, devo prima determinare la FUNZIONE di AUTOCORRELAZIONE; poi da questa devo calcolare  $\tau_{\text{int},f}$ ; INFINE posso DETERMINARE l'ERRORE.