AUTOCORRELAZIONE ed ERRORI

nella SIMULAZIONE di processi di Markov
(es: la nostra simulazione della CODA...)

Abbiamo più volte ripetuto che, al termine di un esperimento numerico (es: le nostre simulazioni su percolazione e code), vogliamo fornire

RISULTATI ed ERRORI!

★ Pensiamo alla simulazione della PERCOLAZIONE. 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 length(dd) = N)

RISULTATO ⟷ mean(dd)
ERRORE ⟷ std(dd)/sqrt(length(dd)) $= \dfrac{\text{std}(dd)}{\sqrt{N}}$

Questo è il caso dei nostri ESPERIMENTI sulla PERCOLAZIONE: siccome i dati generati sono INDIPENDENTI (nell'ipotesi, ovviamente, che il generatore di numeri casuali sia corretto), posso applicare i risultati discussi presentando la LEGGE dei GRANDI NUMERI:

- se osservo un fenomeno descritto da una certa distribuzione, la MEDIA ARITMETICA dei dati (notate che "parliamo Matlab": cfr fz mean) è una STIMA CORRETTA e NON DISTORTA del VALOR 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 σ, lo scarto della MEDIA è dunque $\frac{\sigma}{\sqrt{N}}$, dove σ 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 σ: (cfr note sulla Legge dei Grandi Numeri) poiché ci pensa Matlab a fare la cosa corretta, possiamo disinteressarcene...

$\longrightarrow$ COME DOBBIAMO CALCOLARE gli ERRORI nel caso di SIMULAZIONI di PROCESSI DI MARKOV?

- Abbiamo visto che nel caso della CODA la applicazione della ricetta poco sopra richiamata (std(dd)/sqrt(length(dd)) , valida per DATI INDIPENDENTI!) conduce a SOTTOSTIMARE gli ERRORI!

Un po' di notazioni:

- un processo è una successione di CONFIGURAZIONI del SISTEMA $\{X_t \mid t=1..T\}$;
- 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 VALOR MEDIO di f_t sulla distribuzione asintotica π

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. Correlo due misure "distanti 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 è $\simeq$ prodotto dei valori medi...)

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

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

e dunque $C_{ff}(t) \xrightarrow[t\to\infty]{} 0$ perché $W^{|t|}_{xy} \xrightarrow[t\to\infty]{} \pi_x$!

Quindi SU TEMPI LUNGHI 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 perdevamo memoria della distribuzione di probabilità iniziale $P^{(0)}$ in $P^{(N)} = W^N P^{(0)} \xrightarrow[N \text{ grande}]{} \pi + O(|\hat{\lambda}|^N)$?

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

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

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 riscrivere $|\hat{\lambda}|^t = e^{-\frac{t}{|\ln|\hat{\lambda}||^{-1}}} \equiv e^{-t/\tau}$ ...

DEFINISCO IL TEMPO di AUTORRELAZIONE ESPONENZIALE $\quad \tau_{exp,f} = \limsup_{t\to\infty} \frac{t}{-\log|\rho_{ff}(t)|}$

e $\tau_{exp} = \sup_f \tau_{exp,f}$ rappresenta dunque il TEMPO di RILASSAMENTO del MODO PIU' LENTO del SISTEMA.

Questo comporta che devo attendere un tempo ALMENO $t \sim \tau_{exp}$ 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 $\rho_{ff}(t) \sim e^{-t/\tau}$ ...)

QUANTO VALE LA VARIANZA ALL'EQUILIBRIO di

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

Questa è la quantità che misuro, perché so che $\overline{f} \xrightarrow[N\to\infty]{} \mu_f$. In sostanza, da $\overline{f}$ leggo il RISULTATO e la RADICE QUADRATA (scarto) della sua VARIANZA mi dà la stima dell'ERRORE, che è quello che vado cercando!

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

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 $N \gg \tau$

$$\mathrm{Var}_\pi(\bar{f}) \simeq \frac{1}{N}\left(2\,\tau_{int,f}\right) C_{ff}(0)$$

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

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

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

devo "spaziare le misure" per trovare quelle decorrelate!

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

è allora $\boxed{\mathrm{std}(dd) \,/\, \mathrm{sqrt}(N/(2\,\tau_{int,f}))}$

dove 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 $\tau_{int,f}$, il che chiede a sua volta di misurare la funzione di autocorrelazione di f.

IN SOSTANZA

- PRIMA di PRENDERE MEDIE devo FAR TERMALIZZARE il SISTEMA (per tempi di ordine τ_{exp}...)

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