### 1 Processi stocastici

Il termine "processo stocastico" può essere utilizzato a vari livelli di generalità.

Usualmente, si definisce processo stocastico a tempo discreto ogni successione  $X_1, X_2, \ldots, X_n, \ldots$  di variabili aleatorie. Scriveremo più sinteticamente  $(X_n)_{n\in\mathbb{N}}$ . Un esempio:  $X_n=$  temperatura massima a Pisa il giorno n-esimo del 2006. Un altro:  $X_n=$  valore di apertura del titolo AEM il giorno n-esimo del 2006. A volte si considerano anche processi a tempo discreto in cui il tempo varia sui numeri interi relativi:  $(X_n)_{n\in\mathbb{Z}}$ .

Si chiama processo stocastico a tempo continuo ogni famiglia  $(X_t)_{t\geq 0}$  di variabili aleatorie indicizzata dal tempo  $t\in [0,\infty)$ . Un esempio:  $X_t=$  valore del titolo AEM all'istante t. Con lo stesso nome si indicano anche i casi in cui il tempo varia su tutto l'asse reale:  $(X_t)_{t\in \mathbb{R}}$ ; oppure su un intervallo [0,T]:  $(X_t)_{t\in [0,T]}$  e così via per situazioni simili a queste.

E' però interessante in certe applicazioni considerare variabili aleatorie indicizzate da parametri più generali, ad esempio lo spazio, o lo spazio-tempo. Continueremo a chiamarli processi stocastici, oppure più specificamente  $campi \ aleatori$ . Ad esempio è un campo aleatorio una famiglia  $(U_{(t,x,y,z)})_{t,x,y,z\in\mathbb{R}}$  indicizzata da tempo e spazio. Un esempio concreto può essere la velocità dell'aria nel punto (x,y,z) dello spazio, all'istante t (nello studio delle predizioni atmosferiche si deve considerare questa grandezza aleatoria al variare di tempo e spazio).

Infine, per alcune applicazioni specifiche sono interessanti le famiglie di variabili aleatorie indicizzate da insiemi:  $(X_A)_{A\subset\mathbb{R}^d}$ . Ad esempio:  $X_A =$  quantità d'acqua piovana che cade nella regione A; oppure  $N_{[a,b]} =$  numero di chiamate telefoniche che arriva ad una centrale nel periodo di tempo [a, b].

In definitiva, volendo dare una definizione generale, un processo stocastico è una famiglia di variabili aleatorie indicizzata da un qualche insieme di parametri.

Fondamentale è il doppio punto di vista con cui si può studiare un processo stocastico. Per non fare discorsi inutilmente generali, riferiamoci al caso più semplice  $(X_n)_{n\in\mathbb{N}}$  e pensiamo specificamente al valore di apertura del titolo AEM il giorno n-esimo del 2006.

• Da un lato, possiamo fissare  $n = n_0$  (es. un giorno del futuro) e chiederci quale sia la distribuzione di probabilità di  $X_{n_0}$ , in particolare che valore ci aspettiamo mediamente e con quale deviazione standard.

• Dall'altro, possiamo osservare una realizzazione del processo, cioè una sequenza di valori specifici  $x_1, x_2, ...$  accaduti o che possono potenzialmente accadere.

La non ripetibilità di questo specifico esempio può risultare fuorviante. Pensiamo allora ad un esempio ripetibile. Indichiamo con  $N_t$  il numero di persone arrivate in un negozio nel tempo [0,t], dove con t=0 indichiamo l'istante di apertura, di un generico giorno di lavoro, e t varia tra t=0 e l'istante di chiusura t=T. Ecco i due diversi punti di vista:

- Da un lato possiamo chiederci quale sia in numero medio di persone arrivate dopo un'ora di lavoro; per calcolarlo servono infomazioni sulla distribuzione di probabilità di  $N_{\rm 1ora}$ .
- Dall'altro possiamo considerare ciò che è effettivamente successo in un ben preciso giorno: in quel giorno specifico avremo osservato una particolare funzione  $n(t), t \in [0,T]$ , il numero di persone arrivate quel giorno entro il tempo t; la funzione  $(n(t))_{t \in [0,T]}$  è una realizzazione di  $(N(t))_{t \in [0,T]}$ .

Si usa il termine realizzazione anche per storie non già accadute ma che potenzialmente possono accadere.

## 1.1 Esempi gaussiani

Una classe fondamentale è quella dei processi gaussiani. Un processo  $(X_n)_{n\in\mathbb{N}}$  si dice gaussiano se il vettore aleatorio

$$(X_{n_1}, X_{n_2}, ..., X_{n_k})$$

è gaussiano, per ogni  $n_k > ... > n_2 > n_1 \ge 0$ .

Un esempio banale ma utilissimo è il white noise gaussiano: una successione  $(X_n)_{n\in\mathbb{N}}$  di v.a. gaussiane indipendenti, di solito per convenzione a media nulla, con ugual varianza  $\sigma^2$ . Una sua realizzazione si ottiene col comando rnorm(1:n) in R. E' gaussiano anche il processo

$$X_n = a_n + \varepsilon_n$$

dove  $(a_n)$  è una successione di numeri reali e  $(\varepsilon_n)_{n\in\mathbb{N}}$  è un qhite noise gaussiano. Lo stesso vale per il processo  $(X_n)_{n\in\mathbb{N}}$  definito dall'equazione

$$X_n = aX_{n-1} + \varepsilon_n, \quad X_0 = b$$

con a, b numeri dati, e per molti altri esempi costruiti in modo *lineare* a partire da un white noise gaussiano. In un certo senso, quasi tutti i processi che abbiamo incontrato nelle nostre esemplificazioni con R erano gaussiani.

L'esempio più famoso della precedente categoria è la passeggiata aleatoria  $(random\ walk)$  definita da

$$X_n = X_{n-1} + \varepsilon_n, \quad X_0 = 0$$

o da  $X_n = \sum_{k=0}^n \varepsilon_k$ .

A tempo continuo, un processo  $(X_t)_{t\geq 0}$  si dice gaussiano se il vettore aleaotrio

$$(X_{t_0}, X_{t_1}, ..., X_{t_k})$$

è gaussiano, per ogni  $k \ge 1$  e  $0 \le t_0 < t_1 < ... < t_k$ .

Il white noise gaussiano a tempo continuo non è definibile in modo tradizionale; bisogna utilizzare la teoria delle distribuzioni (coinvolge ad es. la delta di Dirac). Si può ovviare usando lo schema matematico dei processi  $(X_{[a,b]})$  dipendenti da intervalli: si richiede che  $X_{[a,b]}$  sia una v.a. gaussiana di media zero e varianza b-a, e che  $X_{[a,b]}$  e  $X_{[c,d]}$  siano indipendenti se i due intervalli non si intersecano. Il significato intuitivo di  $X_{[a,b]}$  in questo caso è grosso modo la quantità di white noise nell'intervallo [a,b]. Invece il white noise ad un istante t non si può definire senza distribuzioni (intuitivamente sarebbe una sorta di delta di Dirac all'istante t).

Si può invece definire l'analogo della random walk, cioè il moto Browniano  $(B_t)_{t\geq 0}$ . Senza dare la definizione completa, limitiamoci ad elencare alcune sue proprietà fondamentali:

- è un processo gaussiano
- $B_0 = 0$
- $B_t B_s$  è una v.a.  $N(0, t s), t \ge s$
- gli incrementi  $B_{t_n} B_{t_{n-1}}, \ldots, B_{t_1} B_{t_0}$  sono indipendenti, per ogni n > 1 e  $0 < t_0 < t_1 < \ldots < t_n$ .

Purtroppo non è facile darne una definizione esplicita come nel caso discreto, in quanto non abbiamo un concetto facile di white noise. Però il moto Browniano si riesce a definire e costruire rigorosamente. A posteriori esso

è legato al processo  $X_{[a,b]}$  detto sopra (il cumulato del white noise) dalla relazione

$$X_{[s,t]} = B_t - B_s.$$

Volendo andare oltre, si può dire che il white noise è la derivata del moto Browniano:  $\frac{dB_t}{dt}$ . Però si dimostra che  $B_t$  non è derivabile nel senso usuale, ma solo nel senso delle distribuzioni.

Si chiamano equazioni differenziali stocastiche le equazioni formalmente scritte come

$$dX_t = b(X_t, t) dt + \sigma(X_t, t) dB_t.$$

Esse sono un analogo nel continuo di equazioni ricorsive nel discreto aventi il white noise discreto a secondo membro. Queste equazioni si possono studiare rigorosamente ed hanno il pregio, a livello computazionale, che la legge di  $X_t$  si può cercaqre di calcolare in linea di principio risolvendo una equazione alle derivate parziali di tipo parabolico, detta equazione di Fokker-Planck. Nel caso particolare in cui b è lineare in  $X_t$  e  $\sigma$  non dipende da  $X_t$ , la soluzione  $X_t$  di questa equazione (con dato iniziale ad es. deterministico) è un processo gaussiano.

A livello pratico, vale la pena di aggiungere che un grande pregio dei processi gaussiani è il fatto che sono individuati da pochi parametri. Quindi si prestano alla creazione di modelli abbastanza semplici e sono comodi dal punto di vista dell'identificazione rispetto a dati sperimentali. Per un processo gaussiano a tempo discreto basta conoscere  $E[X_n]$  e  $Cov(X_n, X_k)$  per ogni  $n \in k$ . Lo stesso vale a tempo continuo. La ragione di questa semplicità sta nell'espressione generale di un vettore gaussiano: esso è noto quando si conoscono la media delle sue componenti e la matrice di covarianza.

#### 1.2 Processo di Poisson

Un esempio non gaussiano di uso frequente è il processo di Poisson. Sulla falsariga delle proprietà del moto Browniano, senza entrare in tutti i dettagli, elenchiamo alcune proprietà fondamentali del processo di Poisson  $(N_t)_{t\geq 0}$ . di intensità  $\lambda > 0$ :

- $N_0 = 0$
- $N_t N_s$ , per  $t \geq s \geq 0$ , è una v.a. di Poisson di parametro  $\lambda \left( t s \right)$

• gli incrementi  $N_{t_n} - N_{t_{n-1}}, \ldots, N_{t_1} - N_{t_0}$  sono indipendenti, per ogni  $n \ge 1$  e  $0 \le t_0 < t_1 < \ldots < t_n$ .

Osserviamo in particolare, se si prende s = 0, dalle prime due proprietà discende che  $N_t$  è una v.a. di Poisson di parametro  $\lambda t$ .

Ricordiamo che una v.a. X si dice di Poisson di parametro  $\alpha > 0$  se può assumere solo i valori interi non negativi e vale  $P(X=k) = e^{-\alpha} \frac{\alpha^k}{k!}$ . Sottolineiamo (anche in contrasto col moto Browniano) il fatto che i valori del processo di Poisson  $(N_t)_{t\geq 0}$  sono i numeri interi non negativi: questo significa che se tracciamo il grafico di una generica realizzazione, tale grafico è una curva costante a tratti; ed inoltre è non decrescente in quanto gli incrementi sono anch'essi di Poisson; quindi il grafico è una sorta di scalinata ascendente, con salti posizionati irregolarmente. Si può dimostrare che gli intertempi tra un salto ed il successivo sono v.a. esponenziali di parametro  $\lambda$  indipendenti.

Viceversa, ricordiamo il teorema che afferma: data una successione  $(T_n)_{n\in\mathbb{N}}$  di v.a. esponenziali di parametro  $\lambda$  indipendenti, detto  $N_t$  il numero (aleatorio) di eventi di questa successione precedenti a t, cioè il più grande intero n tale che  $T_1+\ldots+T_n\leq t$ , allora  $N_t$  è una v.a. di Poisson di parametro  $\lambda t$ . Si può anche dimostrare che  $(N_t)_{t\geq 0}$  è un processo di Poisson di intensità  $\lambda$ . Pertanto il processo di Poisson serve ad esempio a descrivere il numero di arrivi ad un punto di servizio, se si può assumere che gli intertempi di arrivo siano esponenziali.

Ci sono varie generalizzazioni interessanti. Data una densità di probabilità f(x) (non è nemmeno necessario che sia normalizzata ad 1) si può usare questa al posto di  $\lambda$  nel seguente modo: nella seconda proprietà del processo di Poisson si richiede che

• 
$$N_t - N_s$$
, per  $t \ge s \ge 0$ , è una v.a. di Poisson di parametro  $\int_s^t f(x) dx$ .

Ne deriva il cosidetto processo di Poisson  $non\ omogeneo$ , avente funzione di intensità f. Serve ad esempio, per gli arrivi ad un servizio, a distinguere le ore di punta.

Un'altra generalizzazione è quella al caso bidimensionale (o multidimensionale). Però serve un cambio di punto di vista, un po' analogo a quello accennato sopra per il white noise a tempo continuo. Bisogna introdurre un processo  $(N_A)_{A\subset\mathbb{R}^2}$  indicizzato dagli insiemi A del piano e richiedere:

• 
$$N_A \sim \mathcal{P}(\lambda |A|)$$

•  $N_{A_1}, \ldots, N_{A_k}$  indipendenti se  $A_1, \ldots, A_k$  sono insiemi disgiunti.

Qui |A| indica l'area di A. Oppure si può generalizzare chiedendo

$$N_A \sim \mathcal{P}\left(\lambda \int_A f(x) dx\right).$$

Tali processi di Poisson, detti anche processi di punto (di tipo Poisson) nel piano, descrivono vari problemi. Possono indicare il numero di individui colpiti da una certa epidemia, al variare della zona in esame. Oppure, possono descrivere le posizioni aleatorie in cui avvengono determinati avvenimenti. Infatti, in un certo senso opportuno, è come se dietro un processo di Poisson del piano ci fosse una famiglia di punti aleatori  $(P_i)$ ,  $P_i \in \mathbb{R}^2$ , e la generica v.a.  $N_A$  fosse definita dal numero di punti che cadono in A (si pensi al caso unidimensionale e gli istanti legati ai tempi  $T_i$ ).

# 2 Grandezze associate ad un processo stocastico

Ad una singola v.a. X sono associate varie grandezze: media, varianza ed altri valori medi; varie probabilità, es.  $P(X \le x)$  al variare di x.

Considerando allora un processo stocastico  $(X_n)_{n\in\mathbb{N}}$  (lo stesso vale negli altri casi di processi), possiamo studiare le stesse cose al variare di n: ad esempio il valor medio

$$\mu_n = E\left[X_n\right]$$

e così via.

Fin qui nessuna novità. Le novità nascono quando si introducono grandezze che descrivono il processo preso simultanemente per più valori dei suoi parametri. Una grandezza di questo tipo è la covarianza tra gli istanti n e k:

$$Cov(X_n, X_k) = E[(X_n - \mu_n)(X_k - \mu_k)].$$

Ricordiamo che, a parte il problema dell'unità di misura, la covarianza è un indicatore del legame tra due variabili aleatorie. Quindi  $Cov(X_n, X_k)$  misura il legame tra i tempi n e k. Ad esempio, se  $(X_n)_{n\in\mathbb{N}}$  è un white noise (successione di v.a. indipendenti), allora  $Cov(X_n, X_k) = 0$  per  $n \neq k$ . Invece, se  $(X_n)_{n\in\mathbb{N}}$  è il valore di apertura di un titolo, è plausibile che  $Cov(X_n, X_{n+1})$  sia diverso da zero.

Altre grandezze che dipendono da vari istanti del processo simultaneamente sono le probabilità condizionali:

$$P\left(X_{n+1} = y | X_n = x\right);$$

le probabilità di storie:

$$P(X_{n+1} = x_{n+1}, X_n = x_n, ..., X_0 = x_0)$$

ed altre simili. Tutte le grandezze che dipendono da più tempi colgono in una misura o un'altra qualcosa della dinamica del processo.

### 2.1 Stima empirica delle grandezze

Interessante è il problema della stima sperimentale delle grandezze associate ad un processo. Si deve osservare che già il semplice valor medio  $\mu_n = E[X_n]$  è una successione di numeri, non un singolo numero, per cui il problema della sua stima può apparire molto arduo.

Vengono in aiuto alcuni teoremi validi per particolari processi, detti stazionari ed ergodici. Rimandiamo la loro trattazione alle dispense sulla legge dei grandi numeri e teorema ergodico (richiameremo qualcosa nel prossimo paragrafo). Qui ricordiamo solo l'idea ed alcuni esempi. L'idea è che da una singola realizzazione, quindi da una registrazione sperimentale, si possano stimare varie grandezze, a patto che il processo sia stazionario (in realtà ergodico). La stazionarietà può essere accettata ad occhio vedendo alcune realizzazioni; l'ergodicità purtroppo va ipotizzata alla cieca.

Ricordiamo che per un processo stazionario  $(X_n)_{n\in\mathbb{N}}$  la media  $\mu$  non dipende da n. Se abbiamo una realizzazione sperimentale  $x_1,...,x_n$  ed il processo è ergodico, allora la media aritmerica

$$\frac{x_1 + \dots + x_n}{n}$$

tende a  $\mu$  (nei sensi opportuni stabiliti dai teoremi rigorosi). Si noti che qui  $x_1, ..., x_n$  non è un campione ottenuto con n esperimenti indipendenti, ma ottenuto da un singolo esperimento di durata temporale n. Quindi i valori  $x_1, ..., x_n$  possono provenire da v.a. correlate. Nonostante questo vale la legge dei grandi numeri sopra ricordata.

In ipotesi di stazionarietà, il valor medio  $E[(X_n - \mu)(X_{n+k} - \mu)]$  non dipende da n. Lo chiamiamo funzione di autocovarianza ed indichiamo con

$$R(k)$$
:

$$R(k) = E[(X_0 - \mu)(X_k - \mu)].$$

Per essa a volte si omette la centralizzazione con  $\mu$ , cioè si considera la funzione  $E[X_0X_k]$ , essendo poi  $R(k) = E[X_0X_k] - \mu^2$ .

La funzione di autocorrelazione (libera da problemi di unità di misura) è invece

$$\frac{E\left[\left(X_{0}-\mu\right)\left(X_{k}-\mu\right)\right]}{\sigma^{2}}$$

dove  $\sigma^2 = Var[X_n]$ .

Fatta questa premessa, se abbiamo una realizzazione sperimentale  $x_1, ..., x_n$  ed il processo è ergodico, allora

$$\frac{1}{n-k} \sum_{i=1}^{n-k} x_i x_{i+k} \to E[X_0 X_k].$$

La precisione della stima peggiora al crescere di k, per cui non si deve pretendere di calcolare R(k) per troppi valori di k, se la stringa sperimentale non è sufficientemente lunga.

Con R abiamo calcolato la funzione di autocorrelazione col comando acf(x). Essa ha fornito (si ricordi l'esempio della previsione di un titolo con una certa ciclicità) un'indicazione del periodo delle oscillazioni visibili ad occhio, pur deturpate da rumore: se il periodo è P, la acf calcolata per k=P deve essere marcatamente positiva. Infatti  $X_P$  e  $X_0$  saranno positivamente correlate, quindi  $\frac{E[(X_0-\mu)(X_P-\mu)]}{\sigma^2}$  sarà vicino ad 1. Analogamente, si può argomentare che la serie sperimentale di partenza  $x_1, ..., x_{n_0}$  e la sua traslata di ordine P,  $x_{P+1}, ..., x_{P+n_0}$  (bisogna limitarsi ad un  $n_0$  tale che  $P+n_0 \leq n$ ) devono somigliare, quindi  $\sum_{i=1}^{n-k} x_i x_{i+P} \sim \sum_{i=1}^{n-k} x_i^2$  che conduce ad un valore marcatamente positivo della acf.

Per essere precisi, ciò che fa R quando calcola acf di ordine k non è semplicemente  $\frac{1}{n-k}\sum_{i=1}^{n-k}x_ix_{i+k}$  ma è la correlazione empirica tra le stringhe  $x_1,...,x_{n-k}$  e  $x_{1+k},...,x_n$ , ovvero

$$\frac{\frac{1}{n-k}\sum_{i=1}^{n-k}x_ix_{i+k} - \widehat{\mu}_{n-k}\widehat{\mu}_{n-k}^{(+)}}{\widehat{\sigma}_{n-k}\widehat{\sigma}_{n-k}^{(+)}}$$

dove

$$\widehat{\mu}_{n-k} = \frac{1}{n-k} \sum_{i=1}^{n-k} x_i, \quad \widehat{\mu}_{n-k}^{(+)} = \frac{1}{n-k} \sum_{i=1}^{n-k} x_{i+k}$$

$$\widehat{\sigma}_{n-k} = \sqrt{\frac{1}{n-k} \sum_{i=1}^{n-k} x_i^2}, \quad \widehat{\sigma}_{n-k}^{(+)} = \sqrt{\frac{1}{n-k} \sum_{i=1}^{n-k} x_{i+k}^2}.$$

Naturalmente se il processo stocastico di partenza non è stazionario (ed ergodico), il software calcola ugualmente una acf ma ogni sua interpretazione è fasulla.

### 2.2 Processi stazionari

Un processo a tempo discreto  $(X_n)_{n\in\mathbb{N}}$  si dice *stazionario in senso lato* se i suoi momenti di ordine due (esistono finiti e) non variano nel tempo. Tra i momenti di ordine due vanno però annoverate anche le covarianze. Quindi, per essere precisi, si richiede che:

- $E[X_n]$  non dipende da n
- $\bullet \ Cov\left(X_{n},X_{n+k}\right)$ non dipende dan,per ogni $k=0,1,\ldots$

In particolare, per k=0, si trova che  $Var[X_n]$  non dipende da n, e neppure a posteriori  $E[X_n^2]$ .

A tempo continuo,  $(X_t)_{t\geq 0}$ , si chiede che  $E[X_t]$  e  $Cov(X_t, X_{t+s})$  non dipendano da t, per ogni  $s\geq 0$ .

Un processo a tempo discreto  $(X_n)_{n\in\mathbb{N}}$  si dice stazionario in senso stretto se la legge del vettore aleatorio

$$(X_n, X_{n+1}, ..., X_{n+k})$$

non dipende da n, ma solo da k=0,1,... Ne discende che è stazionario in senso lato. Il viceversa non è vero in generale. Però è vero per processi gaussiani: un processo gaussiano stazionario in senso lato è anche stazionario in senso stretto. La ragione è semplice: media e covarianza determinano tutto del processo.

A tempo continuo,  $(X_t)_{t\geq 0}$ , si chiede che la legge del vettore aleatorio

$$(X_{t+s_0}, X_{t+s_1}, ..., X_{t+s_k})$$

non dipenda da t, ma solo da  $0 \le s_0 < s_1 < ... < s_k$ . I legami col concetto si stazionarietà in senso lato sono identici al caso discreto.

E' facile estendere queste definizioni al caso di processi dipendenti da parametri più generali.

Sotta la sola ipotesi di stazionarietà in senso lato ed un'opportuna ipotesi di decadimento della covarianza per distanze grandi, vale un teorema ergodico, per il cui enunciato rimadiamo alle dispense sulla legge dei grandi numeri.

Invece, sotto l'ipotesi di stazionarietà in senso stretto, più un'ipotesi di ergodicità un po' complessa da formulare (ed estremamente difficile da verificare rigorosamente in esempi concreti) vale un teorema ergodico molto forte, che grossolanamente dice che con probabilità uno vale la seguente affermazione:

$$\lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} \varphi(X_k) = E[\varphi(X)]$$

per ogni funzione  $\varphi$  con un minimo di regolarità  $(E[|\varphi(X)|] < \infty)$ . Qui X è una generica v.a. con la stessa legge delle  $X_k$  (ad es.  $X_1$ ). Il teorema è molto più forte di quello delle dispense sulla LGN in quanto lì si afferma solo che  $\frac{1}{n}\sum_{k=1}^{n}X_k$  converge a E[X], mentre qui convergono infinite altre espressioni interessanti. Ad esempio  $\frac{1}{n}\sum_{k=1}^{n}(X_k-\mu)^2$  converge alla varianza di X (basta prendere  $\varphi(x)=(x-\mu)^2$ ). Oppure, detto  $N_A$  il numero delle  $X_k$ , k=1,...,n, che cadono in un insieme A, vale che  $\frac{N_A}{n}$  tende a  $P(X \in A)$  (basta prendere  $\varphi(x)=1$  se  $x \in A$ , zero altrimenti). Questi teoremi sono alla base della possibilità di stimare grandezze del processo tramite una sola traiettoria.