Probabilità e Processi Stocastici (455AA)

Lezione 21

Dario Trevisan – *https://web.dm.unipi.it/trevisan*

3/12/2025

I teoremi limite

Struttura del capitolo

Concludiamo il corso discutendo alcuni dei principali teoremi limite in probabilità.

- definiamo le nozioni di convergenza per variabili aleatorie.

Struttura del capitolo

Concludiamo il corso discutendo alcuni dei principali teoremi limite in probabilità.

- definiamo le nozioni di convergenza per variabili aleatorie.
- presentiamo la **legge dei grandi numeri**, un risultato che collega il concetto di frequenza con la probabilità.

Struttura del capitolo

Concludiamo il corso discutendo alcuni dei principali teoremi limite in probabilità.

- definiamo le nozioni di convergenza per variabili aleatorie.
- presentiamo la **legge dei grandi numeri**, un risultato che collega il concetto di frequenza con la probabilità.
- vediamo come estendere la legge dei grandi numeri al caso di processi più generali, ottenendo teoremi detti *ergodici*.

Struttura del capitolo

Concludiamo il corso discutendo alcuni dei principali teoremi limite in probabilità.

- definiamo le nozioni di convergenza per variabili aleatorie.
- presentiamo la **legge dei grandi numeri**, un risultato che collega il concetto di frequenza con la probabilità.
- vediamo come estendere la legge dei grandi numeri al caso di processi più generali, ottenendo teoremi detti *ergodici*.
- il **teorema limite centrale** fornisce una precisazione della legge dei grandi numeri facendo intervenire variabili gaussiane

Struttura del capitolo

Concludiamo il corso discutendo alcuni dei principali teoremi limite in probabilità.

- definiamo le nozioni di convergenza per variabili aleatorie.
- presentiamo la **legge dei grandi numeri**, un risultato che collega il concetto di frequenza con la probabilità.
- vediamo come estendere la legge dei grandi numeri al caso di processi più generali, ottenendo teoremi detti *ergodici*.
- il **teorema limite centrale** fornisce una precisazione della legge dei grandi numeri facendo intervenire variabili gaussiane
- Le applicazioni dei teoremi limite comprendono i metodi Monte Carlo, cui accenniamo.

Struttura del capitolo

Concludiamo il corso discutendo alcuni dei principali teoremi limite in probabilità.

- definiamo le nozioni di convergenza per variabili aleatorie.
- presentiamo la **legge dei grandi numeri**, un risultato che collega il concetto di frequenza con la probabilità.
- vediamo come estendere la legge dei grandi numeri al caso di processi più generali, ottenendo teoremi detti *ergodici*.
- il **teorema limite centrale** fornisce una precisazione della legge dei grandi numeri facendo intervenire variabili gaussiane
- Le applicazioni dei teoremi limite comprendono i metodi Monte Carlo, cui accenniamo.
- Accenniamo infine ad altri teoremi limite, per gli eventi *estremi*.

Convergenza di variabili aleatorie

Dobbiamo precisare in che senso una successione di variabili aleatorie $(X_n)_{n=1}^{\infty}$ approssimi una variabile aleatoria limite X_∞, informalmente scriviamo

$$X_n \approx X_\infty \quad \text{o} \quad X_n \to X_\infty .$$

- Due approcci principali:

Convergenza di variabili aleatorie

Dobbiamo precisare in che senso una successione di variabili aleatorie $(X_n)_{n=1}^{\infty}$ approssimi una variabile aleatoria limite X_∞, informalmente scriviamo

$$X_n \approx X_\infty \quad \text{o} \quad X_n \to X_\infty.$$

- Due approcci principali:

1. la distanza $|X_n - X_\infty|$ diventa piccola, con grande probabilità, al crescere di n,

Convergenza di variabili aleatorie

Dobbiamo precisare in che senso una successione di variabili aleatorie $(X_n)_{n=1}^{\infty}$ approssimi una variabile aleatoria limite X_∞, informalmente scriviamo

$$X_n \approx X_\infty \quad \text{o} \quad X_n \to X_\infty.$$

- Due approcci principali:

1. la distanza $|X_n - X_\infty|$ diventa piccola, con grande probabilità, al crescere di n,
2. le leggi di X_n convergono verso la legge di X_∞, ad esempio confrontandone le densità, le CDF, le MGF, i momenti, ecc.

Convergenza in probabilità e in media quadratica

Siano $(X_n)_{n=1}^{\infty}$ e X_∞ variabili aleatorie a valori in $\mathbb{R}^d$. Diciamo che X_n converge verso X_∞

- *in probabilità* se per ogni $\varepsilon > 0$, si ha

$$\lim_{n\to\infty} P(|X_n - X_\infty| \leq \varepsilon) = 1,$$

oppure, equivalentemente,

$$\lim_{n\to\infty} P(|X_n - X_\infty| > \varepsilon) = 0;$$

Convergenza in probabilità e in media quadratica

Siano $(X_n)_{n=1}^{\infty}$ e X_∞ variabili aleatorie a valori in $\mathbb{R}^d$. Diciamo che X_n converge verso X_∞

- *in probabilità* se per ogni $\varepsilon > 0$, si ha

$$\lim_{n\to\infty} P(|X_n - X_\infty| \leq \varepsilon) = 1,$$

  oppure, equivalentemente,

$$\lim_{n\to\infty} P(|X_n - X_\infty| > \varepsilon) = 0;$$

- *in media quadratica* se vale

$$\lim_{n\to\infty} \mathbb{E}\left[|X_n - X_\infty|^2\right] = 0.$$

$\varepsilon > 0$ fissato

La convergenza in media quadratica implica quella in probabilità:

$$P(|X_n - X_\infty| > \varepsilon) = P(|X_n - X_\infty|^2 > \varepsilon^2) \underset{\text{Markov}}{\leq} \frac{\mathbb{E}\left[|X_n - X_\infty|^2\right] \to 0}{\varepsilon^2},$$

- se il membro a destra è infinitesimo per $n \to \infty$ anche quello a sinistra lo è.

Un criterio per la convergenza ad una costante

in media quadratica

Siano $(X_n)_{n=1}^{\infty}$ variabili aleatorie a valori in $\mathbb{R}^d$. Allora X_n converge verso una costante $c \in \mathbb{R}^d$ se e solo se

$$\mathbb{E}[X_n] \to c \quad \text{e} \quad \Sigma_{X_n} \to 0.$$

$$X_n \approx c_n \pm \sigma_n$$

Dimostrazione

Caso scalare $X_n \in \mathbb{R}$ $X_\infty = c$

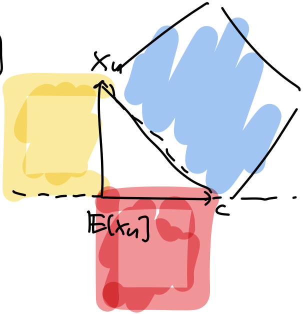
$$\boxed{\mathbb{E}[|X_n - c|^2]} = \mathbb{E}\Big[\big| \underbrace{(X_n - \mathbb{E}[X_n])}_{\text{centrata}} + \underbrace{(\mathbb{E}[X_n]-c)}_{\text{costante}} \big|^2 \Big]$$

$$= \mathbb{E}[(X_n-\mathbb{E}[X_n])^2] + \mathbb{E}[(\mathbb{E}[X_n]-c)^2] +$$

$$+ 2\, \mathbb{E}[(X_n - \mathbb{E}[X_n])(\mathbb{E}[X_n]-c)]$$

$$\mathbb{E}[X_n-\mathbb{E}[X_n]] = 0$$

$$= \boxed{\underbrace{Var(X_n)}_{\text{Varianza}}} + \boxed{\underbrace{(\mathbb{E}[X_n]-c)^2}_{\text{bias}^2}}$$



Convergenza in legge

L'idea sarebbe di confrontare le densità delle X_n con la densità del limite X_∞. Ma come fare se le X_n sono discrete e X_∞ è continua?

Siano $(X_n)_{n=1}^{\infty}$ e X_∞ variabili aleatorie a valori in $\mathbb{R}^d$. Diciamo che X_n converge verso X_∞ *in legge se*

- nel caso $d = 1$, si ha

$$\lim_{n\to\infty} \mathrm{CDF}_{X_n}(t) = \mathrm{CDF}_{X_\infty}(t)$$

per ogni $t \in \mathbb{R}$ eccetto al più i punti t in cui $\mathrm{CDF}_{X_\infty}(t)$ ha una discontinuità di tipo salto (ossia $P(X = t) > 0$)

Convergenza in legge

L'idea sarebbe di confrontare le densità delle X_n con la densità del limite X_∞. Ma come fare se le X_n sono discrete e X_∞ è continua?

Siano $(X_n)_{n=1}^\infty$ e X_∞ variabili aleatorie a valori in $\mathbb{R}^d$. Diciamo che X_n converge verso X_∞ *in legge* se

- nel caso $d = 1$, si ha

$$\lim_{n\to\infty} \mathrm{CDF}_{X_n}(t) = \mathrm{CDF}_X(t)$$

  per ogni $t \in \mathbb{R}$ eccetto al più i punti t in cui $\mathrm{CDF}_X(t)$ ha una discontinuità di tipo salto (ossia $P(X = t) > 0$)

- nel caso generale $d \geq 1$, si ha

$$\lim_{n\to\infty} \mathrm{MGF}_{X_n}(t) = \mathrm{MGF}_{X_\infty}(t)$$

  per ogni t in cui $\mathrm{MGF}_{X_\infty}(t)$ sia finita, supponendo che $\mathrm{MGF}_X(t)$ sia finita per t sufficientemente piccolo.

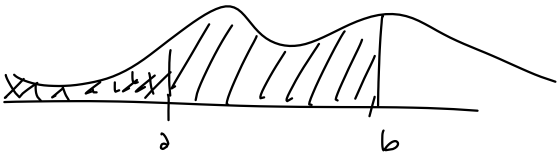
In alternativa, si può richiedere la convergenza delle funzioni caratteristiche per ogni $\omega \in \mathbb{R}$,

$$\lim_{n\to\infty} \varphi_{X_n}(\omega) = \varphi_{X_\infty}(\omega).$$

- Se $d = 1$ e la variabile X_∞ ha densità continua, allora CDF_{X_∞} non ha salti. Possiamo allora chiedere che per ogni $a \leq b \in \mathbb{R}$,

$$\lim_n P(X_n \in (a,b]) = P(X_\infty \in (a,b]).$$

$$P(X_n \in (a,b]) = \mathrm{CDF}_{X_n}(b) - \mathrm{CDF}_{X_n}(a)$$



In alternativa, si può richiedere la convergenza delle funzioni caratteristiche per ogni $\omega \in \mathbb{R}$,

$$\lim_{n\to\infty} \varphi_{X_n}(\omega) = \varphi_{X_\infty}(\omega).$$

- Se $d = 1$ e la variabile X_∞ ha densità continua, allora CDF_{X_∞} non ha salti. Possiamo allora chiedere che per ogni $a \leq b \in \mathbb{R}$,

$$\lim_n P(X_n \in (a, b]) = P(X_\infty \in (a, b]).$$

- La convergenza in legge è anche equivalente alla convergenza

$$\lim_{n\to\infty} \mathbb{E}\left[g(X_n)\right] = \mathbb{E}\left[g(X_\infty)\right]$$

  per ogni funzione g continua ovunque e uniformemente limitata (ossia esiste una costante c tale che $|g(x)| \leq c$ per ogni $x \in \mathbb{R}^d$).

Legge dei grandi numeri

La legge dei grandi numeri è un teorema che giustifica l'intepretazione della **probabilità** di una affermazione come **frequenza relativa** con si realizza in una successione di esperimenti ripetuti, sotto le stesse condizioni, ma tutti indipendenti tra loro.

- Ne diamo una dimostrazione usando la convergenza in media quadratica (e quindi in probabilità).

Legge dei grandi numeri

La legge dei grandi numeri è un teorema che giustifica l'intepretazione della **probabilità** di una affermazione come **frequenza relativa** con si realizza in una successione di esperimenti ripetuti, sotto le stesse condizioni, ma tutti indipendenti tra loro.

- Ne diamo una dimostrazione usando la convergenza in media quadratica (e quindi in probabilità).
- Cominciamo con il caso più semplice delle estrazioni con rimpiazzo da un'urna.

Modello dell'urna

La frazione delle palline rosse è $r \in [0,1]$ e si effettuano n estrazioni con rimpiazzo.

- il numero R_n di palline rosse estratte ha densità binomiale di parametri (n, r), con

$$\mathbb{E}[R_n] = n \cdot r \quad \text{e} \quad \mathrm{Var}(R_n) = n\, r(1-r)$$

Modello dell'urna

La frazione delle palline rosse è $r \in [0,1]$ e si effettuano n estrazioni con rimpiazzo.

- il numero R_n di palline rosse estratte ha densità binomiale di parametri (n, r), con

$$\mathbb{E}[R_n] = nr \quad \text{e} \quad \mathrm{Var}(R_n) = nr(1-r)$$

- Passando alla **frequenza relativa** R_n/n, possiamo scrivere informalmente:

$$\frac{R_n}{n} \approx r \pm \frac{\sqrt{r(1-r)}}{\sqrt{n}}$$

Modello dell'urna

La frazione delle palline rosse è $r \in [0, 1]$ e si effettuano n estrazioni con rimpiazzo.

- il numero R_n di palline rosse estratte ha densità binomiale di parametri (n, r), con

$$\mathbb{E}[R_n] = \qquad \text{e} \quad \mathrm{Var}(R_n) = \qquad .$$

- Passando alla **frequenza relativa** R_n/n, possiamo scrivere informalmente:

$$\frac{R_n}{n} \approx r \pm \frac{\sqrt{r(1-r)}}{\sqrt{n}}$$

- Al tendere di $n \to \infty$ la frazione R_n/n converge verso r, la frazione di palline rosse sul totale.

Legge dei grandi numeri per esperimenti ripetuti

Partendo dal modello delle estrazioni dall'urna possiamo estendere ad una qualsiasi situazione in cui vi siano un grande numero, potenzialmente illimitato, di esperimenti ripetuti, tutti indipendenti tra loro, e ciascuno con probabilità di sucesso $p \in [0, 1]$.

- La *frequenza relativa del numero di successi sul totale degli esperimenti* converge quindi alla *probabilità di successo di un singolo esperimento*.

Legge dei grandi numeri per esperimenti ripetuti

Partendo dal modello delle estrazioni dall'urna possiamo estendere ad una qualsiasi situazione in cui vi siano un grande numero, potenzialmente illimitato, di esperimenti ripetuti, tutti indipendenti tra loro, e ciascuno con probabilità di sucesso $p \in [0, 1]$.

- La *frequenza relativa del numero di successi sul totale degli esperimenti* converge quindi alla *probabilità di successo di un singolo esperimento*.
- Attenzione: può essere limitante intepretare la probabilità solo come frequenza.

Applicazione: stima della probabilità

Supponiamo che inizialmente la frazione di palline rosse non sia nota, e si supponga una variabile aleatoria R a valori in $[0,1]$ ad esempio uniforme.

$$P\left(\left|\frac{R_n}{n}-r\right|\leq\varepsilon\right)=1-P\left(\left|\frac{R_n}{n}-r\right|>\varepsilon\right)\geq 1-\frac{r(1-r)}{n\varepsilon^2}$$

chebychev

Allora,

$$P(|R_n/n-R|\leq\varepsilon)=\int_0^1 P(|R_n/n-r|\leq\varepsilon|R=r)dr$$
$$\geq 1-\frac{\int_0^1 r(1-r)dr}{n\varepsilon^2}=1-\frac{1}{6n\varepsilon^2}\to 1$$

per $n\to\infty$, ossia con alta probabilità la frequenza relativa R_n/n è vicina alla variabile R (precisamente abbiamo mostrato la convergenza in probabilità).

- La probabilità calcolata sopra è rispetto all'informazione a priori, ossia prima di effettuare le estrazioni (o prima di essere informati dell'esito).

Istogrammi e densità teorica

Possiamo spiegare perché l'istogramma relativo ad n osservazioni di variabili indipendenti, tutte con la stessa densità debba essere molto vicino al grafico della densità teorica.

- Siano $(X_i)_{i=1}^n$ variabili indipendenti tutte con la medesima densità (ad esempio una gaussiana).

Istogrammi e densità teorica

Possiamo spiegare perché l'istogramma relativo ad n osservazioni di variabili indipendenti, tutte con la stessa densità debba essere molto vicino al grafico della densità teorica.

- Siano $(X_i)_{i=1}^n$ variabili indipendenti tutte con la medesima densità (ad esempio una gaussiana).
- Consideriamo un rettangolo di base $a < b \in \mathbb{R}$:

Istogrammi e densità teorica

Possiamo spiegare perché l'istogramma relativo ad n osservazioni di variabili indipendenti, tutte con la stessa densità debba essere molto vicino al grafico della densità teorica.

- Siano $(X_i)_{i=1}^{n}$ variabili indipendenti tutte con la medesima densità (ad esempio una gaussiana).
- Consideriamo un rettangolo di base $a < b \in \mathbb{R}$:

1. l'istogramma delle frequenze (assolute) avrà altezza $H(a, b)$ pari al numero delle X_i tali che $a < X_i \leq b$,

Istogrammi e densità teorica

Possiamo spiegare perché l'istogramma relativo ad n osservazioni di variabili indipendenti, tutte con la stessa densità debba essere molto vicino al grafico della densità teorica.

- Siano $(X_i)_{i=1}^n$ variabili indipendenti tutte con la medesima densità (ad esempio una gaussiana).
- Consideriamo un rettangolo di base $a < b \in \mathbb{R}$:

1. l'istogramma delle frequenze (assolute) avrà altezza $H(a, b)$ pari al numero delle X_i tali che $a < X_i \leq b$,
2. quello delle densità è ulteriormente diviso il numero delle osservazioni n e per la lunghezza della base $(b - a)$.

$$\frac{\#\{i : X_i \in (a,b)\}}{n\,(b-a)}$$

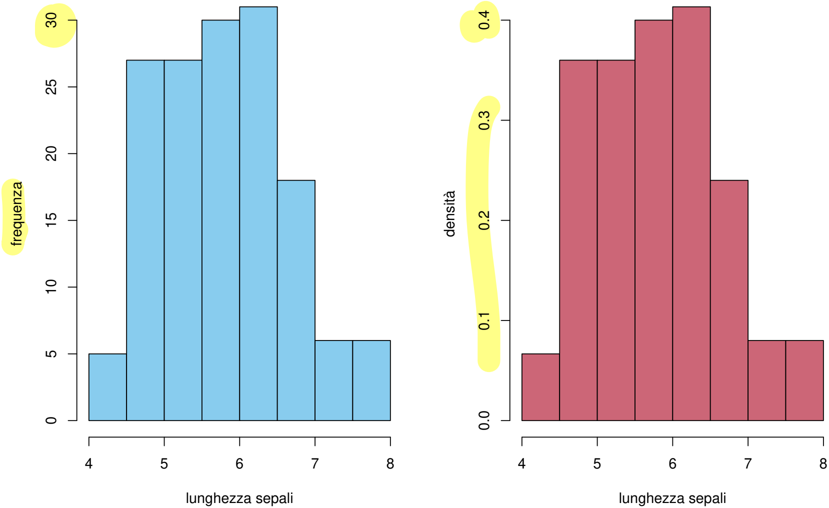


Figure 1: istogramma delle frequenze (a sinistra) e delle densità (a destra)

Per costruire l'istogramma, possiamo pensare ad un "successo" se $X_i \in (a, b]$, che avviene con probabilità

$$r = P(X_1 \in (a, b]) = \int_a^b p(X_1 = x)dx \approx p(X_1 = a)(b - a),$$

dove nell'ultima approssimazione supponiamo la densità regolare e $b - a$ piccolo.

- Considerando n esperimenti indipendenti si avrà

$$\frac{H(a, b)}{n} \approx r \pm \sqrt{\frac{r(1 - r)}{n}} \approx \boxed{p(X_1 = a)}(b - a),$$

e quindi l'istogramma delle densità, che ha altezza $H(a, b)/(n(b - a))$, è, con alta probabilità vicino alla densità teorica.

Legge dei grandi numeri generale

Verso una legge dei grandi numeri generale

Estendiamo a situazioni in cui l'esito di ciascun "esperimento" sia una variabile aleatoria X_i a valori reali.

- la frequenza relativa dei successi è sostituita dalla media empirica/campionaria

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

Verso una legge dei grandi numeri generale

Estendiamo a situazioni in cui l'esito di ciascun "esperimento" sia una variabile aleatoria X_i a valori reali.

- la frequenza relativa dei successi è sostituita dalla media empirica
$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i.$$
- Per dedurre la convergenza in media quadratica delle frequenze relative, avevamo usato che la legge della somma $\sum_{i=1}^{n} X_i$, ossia il numero di successi, ha densità discreta binomiale di parametri (n, p).

Verso una legge dei grandi numeri generale

Estendiamo a situazioni in cui l'esito di ciascun "esperimento" sia una variabile aleatoria X_i a valori reali.

- la frequenza relativa dei successi è sostituita dalla media empirica
$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i.$$
- Per dedurre la convergenza in media quadratica delle frequenze relative, avevamo usato che la legge della somma $\sum_{i=1}^{n} X_i$, ossia il numero di successi, ha densità discreta binomiale di parametri (n, p).
- Tuttavia ripercorrendo l'argomento, basta conoscere molto meno: infatti è sufficiente che il valor medio $\mathbb{E}\left[\bar{X}_n\right]$ converga a una costante m e la varianza $\mathrm{Var}\left(\bar{X}_n\right)$ sia infinitesima per $n \to \infty$.

Teorema

Siano $(X_n)_{n=1}^{\infty}$ variabili aleatorie non correlate, tutte con lo stesso valor medio e varianza

$$\mathbb{E}[X_n] = m, \quad \mathrm{Var}(X_n) = \sigma^2 < \infty.$$

Allora, si ha la convergenza in media quadratica (e quindi in probabilità)

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} X_i = m.$$

Dimostrazione

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \qquad c = \mathbb{E}[X_i] = m$$

- $\mathbb{E}[\overline{X}_n] = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[X_i] = c$

- $\mathrm{Var}(\overline{X}_n) = \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right)$

$$= \frac{1}{n^2}\left[\sum_{i=1}^{n}\mathrm{Var}(X_i) + \cancel{2\sum_{i<j}\mathrm{Cov}(X_i,X_j)}\right]$$

$$= \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n} \Rightarrow \sqrt{\mathrm{Var}(\overline{X}_n)} = \frac{\sigma}{\sqrt{n}} \to 0$$

Dalla dimostrazione segue che la deviazione standard della variabile $\bar{X}_n$ è

$$\sigma_{\bar{X}_n} = \sqrt{\mathrm{Var}\left(\bar{X}_n\right)} = \frac{\sigma}{\sqrt{n}},$$

e quindi informalmente possiamo scrivere

$$\boxed{\bar{X}_n = m \pm \frac{\sigma}{\sqrt{n}}.}$$

Applicazione: convergenza della varianza campionaria

Possiamo usare la legge dei grandi numeri anche per mostrare la convergenza della varianza campionaria

$$\boxed{\bar{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2.}$$

- Supponendo ad esempio che le $(X_i)_i$ siano tutte indipendenti, tutte con le stessa legge e dotate di momento quarto finito, quindi in particolare i momenti sono tutti uguali:

$$m_1 = \mathbb{E}\left[X_i\right], \quad m_2 = \mathbb{E}\left[X_i^2\right],$$

vale la convergenza in media quadratica e in probabilità

$$\lim_{n\to\infty} \bar{\sigma}_n^2 = \sigma^2,$$

Dimostrazione

$$\bar{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}_n\right)^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i^2 + (\bar{X}_n)^2 - 2X_i\bar{X}_n\right)$$

$$= \underbrace{\frac{1}{n}\sum_{i=1}^{n} X_i^2}_{LGN} + \underbrace{(\bar{X}_n)^2}_{LGN} - 2\bar{X}_n \cdot \underbrace{\frac{1}{n}\sum_{i=1}^{n} X_i}_{\bar{X}_n}$$

$$\downarrow \qquad\qquad \downarrow \qquad\qquad \downarrow$$

$$= \mathbb{E}[X_1^2] + \left(\mathbb{E}[X_1]\right)^2 - 2\,\mathbb{E}[X_1]^2$$

$$= \mathbb{E}[X_1^2] - \mathbb{E}[X_1]^2 = Var(X_1) = \sigma^2$$

Sull'ipotesi di indipendenza

Senza l'ipotesi di indipendenza (o non correlazione), la varianza della media campionaria è in generale la somma di n^2 termini

$$\mathrm{Var}\left(\bar{X}_n\right) = \boxed{\frac{1}{n^2} \sum_{i,j=1}^{n} \mathrm{Cov}\left(X_i, X_j\right),}$$

e quindi non segue necessariamente che sia infinitesima, anche tenendo in conto del denominatore n^2.

- Può accadere tuttavia che $\mathrm{Cov}\left(X_i, X_j\right)$ sia piccolo per "molte" coppie di indici, ad esempio per opportuni processi stocastici.

Media temporale di un processo

Dato un processo stocastico reale $(X_t)_{t=1}^{\infty}$, la media sui primi T tempi è $\bar{X}_T = \frac{1}{T}\sum_{t=1}^{T} X_t$.

- Se l'insieme degli stati E del processo non è un sottoinsieme di $\mathbb{R}$, si può considerare una qualsiasi *osservabile* $g : E \to \mathbb{R}$ e considerare la media

$$\overline{g(X)}_T = \boxed{\frac{1}{T}\sum_{t=1}^{T} g(X_t).}$$

Media temporale di un processo

Dato un processo stocastico reale $(X_t)_{t=1}^{\infty}$, la media sui primi T tempi è $\bar{X}_T = \frac{1}{T}\sum_{t=1}^{T} X_t$.

- Se l'insieme degli stati E del processo non è un sottoinsieme di $\mathbb{R}$, si può considerare una qualsiasi *osservabile* $g : E \to \mathbb{R}$ e considerare la media

$$\overline{g(X)}_T = \frac{1}{T}\sum_{t=1}^{T} g(X_t).$$

- Se g è la funzione indicatrice di un qualsiasi stato $x_0 \in E$,

$$g(x) = \begin{cases} 1 & \text{se } x = x_0 \\ 0 & \text{se } x \neq x_0. \end{cases}$$

Allora $\overline{g(X)}_T$ è la frazione di tempo trascorsa dal processo sullo stato x_0, dal tempo $t = 1$ al tempo $t = T$.

Teorema ergodico

Se il processo $(X_t)_t$ è stazionario, si vuole indentificare il limite (se esiste)

$$\lim_{T\to\infty} \overline{g(X)_T}.$$

come il valor medio di g rispetto alla legge marginale in un qualsiasi istante, ad esempio nel caso di E discreto

$$\mathbb{E}\left[g(X_i)\right] = \sum_{x\in E} g(x)P(X_i = x).$$

$\left(P(X_i = x) = \mu(x) \right)$

distr. invariante

- Se g è l'indicatrice di uno stato x_0, il valor medio è la probabilità $\mathbb{E}\left[g(X_i)\right] = P(X_i = x_0)$.

Teorema ergodico

Se il processo $(X_t)_t$ è stazionario, si vuole indentificare il limite (se esiste)

$$\lim_{T\to\infty} \overline{g(X)}_T.$$

come il valor medio di g rispetto alla legge marginale in un qualsiasi istante, ad esempio nel caso di E discreto

$$\mathbb{E}\left[g(X_i)\right] = \sum_{x\in E} g(x)P(X_i = x).$$

- Se g è l'indicatrice di uno stato x_0, il valor medio è la probabilità $\mathbb{E}\left[g(X_i)\right] = P(X_i = x_0)$.
- Risultati che garantiscono tale identificazione sono storicamente detti *teoremi ergodici*.

Enunciato del teorema ergodico

Sia $(X_t)_{t=0}^{\infty}$ un processo stazionario sull'insieme degli stati E e sia $g : E \to \mathbb{R}$ una osservabile. Se

$$\lim_{t\to\infty} \mathrm{Cov}\left(g(X_0), g(X_t)\right) = 0,$$

allora vale la convergenza in media quadratica e in probabilità

$$\lim_{T\to\infty} \overline{g(X)}_T = \mathbb{E}\left[g(X_0)\right].$$

(per semplicità abbiamo specificato X_0, ma un qualsiasi altro tempo X_t sarebbe lo stesso, essendo il processo stazionario).

Dimostrazione

Chiamo $Y_t = g(X_t)$ $\qquad$ Y è stazionario (in senso largo)

$$\overline{Y}_T = \frac{1}{T}\sum_{t=1}^{T} Y_t$$

- $$\mathbb{E}[\overline{Y}_T] = \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}[Y_t] = \mathbb{E}[Y_0]$$

- $$Var(\overline{Y}_T) = \frac{1}{T^2}\sum_{s,t=1}^{T} Cov(Y_s, Y_t)$$

$$= \frac{1}{T^2}\sum_{s,t=1}^{T} C(0, |t-s|) \le \frac{1}{T^2}\Big(T\,C(0) + 2T|C(1)| + 2T|C(2)| + \dots + 2T|C(T-1)|\Big)$$

$$\le \frac{2\cancel{T}}{T^{\cancel{2}}}\Big(C(0) + |C(1)| + \dots + |C(T-1)|\Big) \longrightarrow 0 \quad \underline{\text{Cesàro}}$$

Applicazioni: catene di Markov

Sia $(X_t)_t$ una catena di Markov stazionaria e irriducibile su un insieme di stati finito E. Allora per ogni $g : E \to \mathbb{R}$ vale la convergenza

$$\lim_{T\to\infty} \overline{g(X)}_T = \sum_{i\in E} g(i)\pi_i,$$

dove $\pi = (\pi_i)_{i\in E}$ è l'unica distribuzione invariante per la catena.

- In particolare, la frazione di tempo trascorsa dalla catena su uno stato è, nel limite, pari alla probabilità che la catena si trovi su quello stato.

Applicazioni: catene di Markov

Sia $(X_t)_t$ una catena di Markov stazionaria e irriducibile su un insieme di stati finito E. Allora per ogni $g : E \to \mathbb{R}$ vale la convergenza

$$\lim_{T\to\infty} \overline{g(X)}_T = \sum_{i\in E} g(i)\pi_i,$$

dove $\pi = (\pi_i)_{i\in E}$ è l'unica distribuzione invariante per la catena.

- In particolare, la frazione di tempo trascorsa dalla catena su uno stato è, nel limite, pari alla probabilità che la catena si trovi su quello stato.
- Un teorema analogo vale per processi di Markov a salti, dove la media nel tempo è intesa come integrale.

$$\overline{g(X)}_T = \frac{1}{T}\int_0^T g(X_s)\,ds$$

Applicazioni: processi ARIMA

Sia $(X_t)_t$ un processo ARIMA$(p, 0, q)$ stazionario. Allora

$$\lim_{T\to\infty} \overline{X}_T = 0,$$

e per ogni $t \in \mathbb{N}$,

$\mathrm{Cov}(Y_s, Y_{s'}) \to 0$ se $|s - s'| \to \infty$

$$\lim_{T\to\infty} \frac{1}{T} \sum_{s=1}^{T} \overbrace{X_s X_{s+t}}^{Y_s} = C(0, t).$$

- La funzione di autocovarianza empirica converge quindi a quella teorica. Questo fatto vale per processi stazionari anche più generali degli ARIMA.