



I comportamenti limite

Sommario della lezione 5

- Esperimenti replicati e media
- La VA media del campione
 - Proprietà sintetiche
 - Come è distribuita
- Comportamenti limite
- Il Teorema del Limite Centrale

Richiami

- Abbiamo già visto che gli esiti di un esperimento replicato non coincidono tra loro per l'aleatorietà degli errori di misura
- Un esperimento replicato un certo numero di volte fornisce informazioni su tendenza centrale e variabilità della corrispondente variabile aleatoria.
- Esamineremo le repliche come sequenze di VA indipendenti tutte caratterizzate dalla stessa distribuzione incognita
- Il mediare riduce la variabilità nella misura

Media e Repliche

- Abbiamo già enfatizzato in una delle prime lezioni che le proprietà che traiamo da un campione sperimentale non coincidono in generale con le proprietà della corrispondente popolazione.
- Possiamo cominciare ad affrontare con gli strumenti di cui ora disponiamo il problema della caratterizzazione delle proprietà della popolazione.
- Supponiamo di avere $X_1, \dots, X_n$ caratterizzati da una certa P e di voler stimare la media della popolazione $\mu = E(X_i)$.
- Per fare ciò effettuiamo gli esperimenti ed osserviamo gli esiti $x_1, \dots, x_n$ di $X_1, \dots, X_n$, e quindi calcoliamo

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

Media del campione

- Questa procedura è matematicamente equivalente a definire una nuova variabile aleatoria e quindi calcolarne il suo valore atteso:

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$$
$$\bar{x}_n = E(\bar{X}_n)$$

Trasformazione di
una VA vett.

- Sia la variabile aleatoria $\bar{X}_n$ che $\bar{x}_n$ sono dette **media del campione** ma non dobbiamo assolutamente confondere le due cose.
 - Perciò qualche tempo fa avevamo distinto con maiuscole e minuscole le VA e gli esiti sperimentali.
- **La VA media del campione è caratterizzata da una distribuzione di probabilità.**
 - **Quindi possiamo anche caratterizzare questa VA con media, varianza etc.**

Repliche

- Consideriamo di poter eseguire un numero qualunque di repliche dello stesso esperimento in modo **indipendente**
- Descriviamo questa situazione attraverso una sequenza di VA scalari identicamente distribuite, $X_1, X_2, \dots$, ed presumiamo che queste VA abbiano media finita $\mu = E(X_i)$ e varianza finita $\sigma^2 = \text{Var}X_i$.
- Studiamo sotto queste condizioni le proprietà della media del campione $\bar{X}_n$ al crescere di n .
- Sulla base delle ipotesi è evidente che la media della popolazione, μ , coincide con la media di ciascuna X_i .
- Inoltre vale:
$$E(\bar{X}_n) = \frac{1}{n} \sum_{i=1}^n E(X_i) = \frac{1}{n} \sum_{i=1}^n \mu = \mu$$

Varianza della VA media del campione

- La tendenza di una VA ad assumere valori prossimi al suo valore atteso è quantificabile calcolando la sua varianza.
- Abbiamo ipotizzato: $\text{Var}X_i = \sigma^2$.
- Ma possiamo facilmente dimostrare che vale la seguente proprietà:

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

Trasformazioni di VA

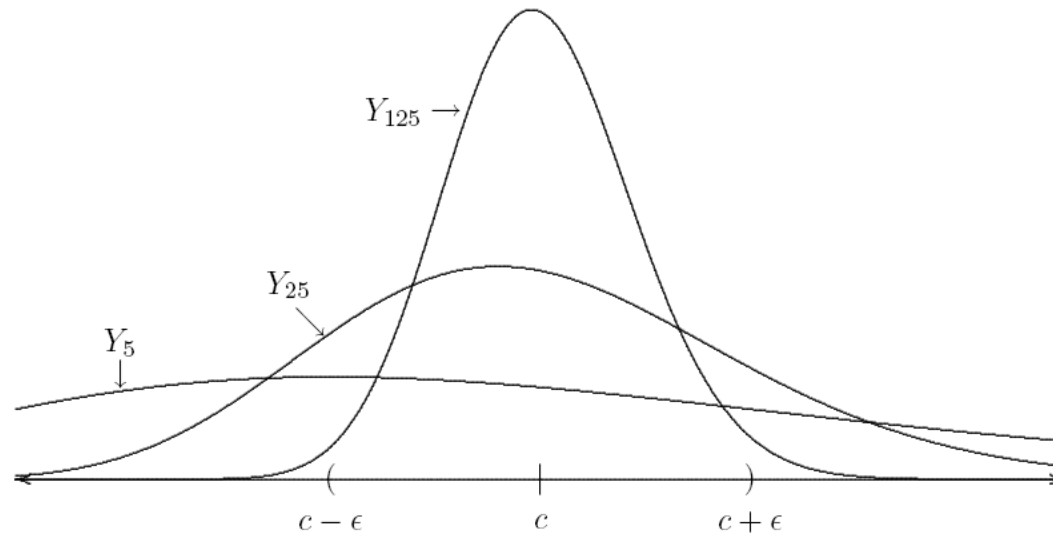
- Importante: la VA media del campione è meno variabile di ciascuna VA componente la media.
- Se $n \rightarrow \infty$ la varianza della media tende a zero: Ripetendo l'esperimento possiamo ridurre a nostro piacere la varianza della media del campione

L'operazione di media
riduce la variabilità

Sequenze di VA

- Definizione: una sequenza di VA $\{Y_n\}$ converge in probabilità ad una costante c ($Y_n \rightarrow c$), se e solo se per ogni $\varepsilon > 0$

$$\lim_{n \rightarrow \infty} P\{Y_n \in (c - \varepsilon, c + \varepsilon)\} = 1$$



Legge dei grandi numeri

- Legge dei Grandi Numeri (forma debole): Sia $X_1, X_2, \dots$ Una sequenza di VA indipendenti ed identicamente distribuite con media μ e varianza entrambe finite. Allora:

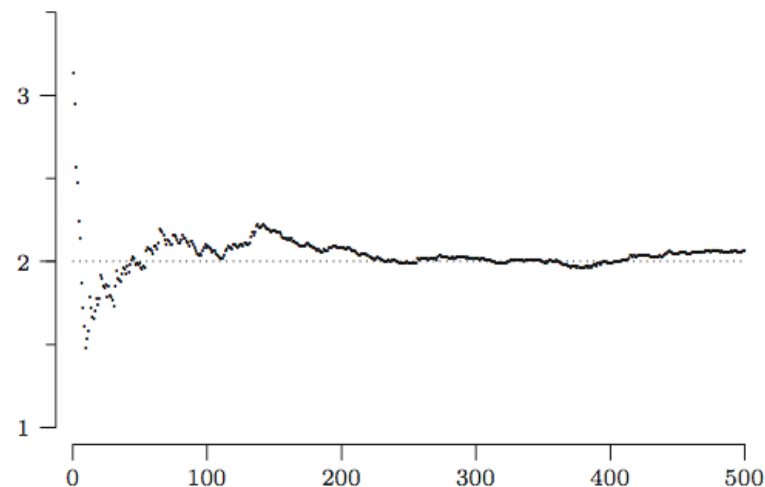
$$\lim_{n \rightarrow \infty} P(|\bar{X}_n - \mu| > \epsilon) = 0 \quad \Leftrightarrow \quad \bar{X}_n \xrightarrow{P} \mu$$

- Questa legge stabilisce che se mediamo progressivamente X_i , i valori delle medie che osserviamo tendono a distribuirsi più vicino al valore medio teorico di X_i .
 - Questa proprietà della media del campione rafforza la nostra aspettativa che il comportamento di $\bar{X}_n$ fornisca informazioni migliori sul vero valore di μ al crescere di n .
 - Se la media della popolazione non è nota si può usare la media del campione estratto per stimarla. Più è grande il campione più accurata è la stima
 - Si può anche dire che $\bar{X}_n$ converge in probabilità a μ

Legge dei grandi numeri

Interpretazione dagli esperimenti

- Immaginiamo di eseguire una serie di esperimenti.
- Facciamo n esperimenti replicati e mediamo i risultati
- Ad ogni replica misuriamo la media. Si ottiene il diagramma in basso.



Legge dei grandi numeri

Interpretazione dagli esperimenti

- Al crescere di n la probabilità che la media sia ad una certa distanza dal valore atteso cresce (Chebyshev inequality) fino a tendere ad 1 nel limite di infinite repliche.
- In pratica al crescere del numero di repliche la media tende al valore “vero”.
 - In che senso?

Teorema del Limite Centrale

- La Legge dei Grandi Numeri stabilisce un modo preciso in cui la distribuzione dei valori della media del campione collassa sulla media della popolazione genitrice al crescere della dimensione del campione.

Domande inevase:

1. Quanto rapidamente la media del campione tende alla media della popolazione?
 2. Come cambia il tipo di VA media del campione al tendere verso la media della popolazione.
- Per rispondere a queste domande conviene trasformare tutte le VA in forma standard.

Teorema del Limite Centrale

VA	E()	StDev	VA in forma St.
X_i	μ	σ	$(X_i - \mu) / \sigma$
$\sum_{i=1}^n X_i$	$n\mu$	$\sqrt{n} \sigma$	$(\sum_{i=1}^n X_i - n\mu) / (\sqrt{n} \sigma)$
$\bar{X}_n$	μ	$\sigma / \sqrt{n}$	$(\bar{X}_n - \mu) / (\sigma / \sqrt{n})$

- Consideriamo la somma e la media di $X_1, \dots, X_n$.
 - Notate che una volta in forma standard queste quantità sono identiche

$$Z_n = \frac{\sum_{i=1}^n X_i - n\mu}{\sqrt{n} \sigma} = \frac{\bar{X}_n - \mu}{\sigma / \sqrt{n}}$$

- E' sulla VA Z_n su cui ora dobbiamo lavorare un po'.

Teorema del Limite Centrale

- Possiamo facilmente notare che: $VarZ_n = 1$
$$(\bar{X}_n - \mu) \xrightarrow{P} 0$$
- Non possiamo dire molto circa il comportamento di Z_1 , ma vedremo che possiamo prendere posizione su Z_n al crescere di n
- **Il teorema che stiamo per enunciare è uno dei più importanti ed utili risultati di tutta la matematica!**

Teorema del Limite Centrale

TEOREMA DEL LIMITE CENTRALE

Sia $X_1, X_2, \dots$ una qualunque sequenza di VA indipendenti ed identicamente distribuite con media finita μ e varianza finita σ^2 . Sia

$$Z_n = \frac{\bar{X}_n - \mu}{\sigma / \sqrt{n}}$$

denotiamo con F_n la CDF di Z_n , e con Φ la CDF GAUSSIANA IN FORMA STANDARD. Allora, per ogni valore fissato di $z \in \mathbb{R}$, al divergere di n

$$P\{Z_n \leq z\} = F_n(z) \rightarrow \Phi(z)$$

Teorema del Limite Centrale

- Il Teorema del Limite Centrale (TLC) stabilisce che il comportamento della media (o della somma) di un grande numero di VA indipendenti ed identicamente distribuite assomiglierà a quello di una VA Normale in forma standard
- Questo accade indipendentemente dal tipo di VA che viene mediata.
- **Quindi il TLC ci permette di approssimare una varietà di probabilità che altrimenti sarebbero intrattabili.**
- La *normalità* viene raggiunta nel limite ma ai fini pratici può bastare $n \geq 30$.
 - Se le distribuzioni sono simmetriche anche meno.
- Il teorema viene esteso anche al caso in cui le distribuzioni non siano identiche.

Teorema del Limite Centrale

- Altre formulazioni del TLC

$$\sum_{i=1}^n X_i := N(n\mu, n\sigma^2)$$

$$\bar{X}_n := N\left(\mu, \frac{\sigma^2}{n}\right)$$

- Il TLC ha una lunga storia
 - Una versione speciale è di De Moivre (1730).
 - Una tentativo generale Laplace (1810)
 - La versione definitiva è del secondo quarto del 20° secolo

Teorema del Limite Centrale

- Quando il TLC non può essere applicato:

Se

1. Le VA non sono indipendenti.
2. Le VA non hanno varianze ben definite. L'integrale che determina la varianza diverge (Levi's walk). Senza varianza non è possibile avere la sequenza di VA con eguale varianza.

Teorema del Limite Centrale

- Abbiamo spesso affermato che in molte situazioni gli errori casuali di misura sono almeno approssimativamente Gaussiani. Il TLC in un certo senso garantisce che sia effettivamente così
- Possiamo anche intuirlo. Pensiamo alla misura dell'altezza della corda di un pendolo con un metro flessibile.
 - La misura può avere molti problemi.
 - Il metro può flettersi o piegarsi.
 - La corda può a sua volta piegarsi o muoversi.
- Tutte queste cause di errore si sommano contribuendo all'errore totale.
- L'errore medio di ciascuna causa è nullo (non ci sono errori sistematici)
- Le cause di errori sono statisticamente indipendenti.. QUINDI

Teorema del Limite Centrale

- Quindi, sulla base del TLC, la media dell'errore totale è grosso modo distribuita Gaussianamente.
- Generalizzando: ogni volta che abbiamo effetti **additivi** di molte cause indipendenti ciascuno con effetti relativi piccolo possiamo sospettare la presenza di Gaussiane.

Esempio I TLC

- Si vuole determinare la conformazione di una molecola. Per misurare la distanza tra una coppia di atomi di idrogeno si usa la spettroscopia NMR. Sappiamo che la misura ha un valore atteso pari alla distanza reale (assenza di errori sistematici) ed una deviazione standard di 0.5 angstroms. Se la misura è replicata 36 volte quale è la probabilità che la media dei valori misurati cada in un intervallo di ± 0.1 angstroms del valore vero?
- Con X_i denotiamo la VA della replica i , per $i = 1, \dots, 36$.
- Sappiamo che $\mu = E(X_i)$ è la distanza reale tra gli atomi e che $\sigma^2 = \text{Var}X_i = 0.5^2$.
- Sia $Z := N(0, 1)$.

Esempio I TLC

- Applicando il TLC:

$$\begin{aligned}P\{\mu - 0.1 < \bar{X}_{36} < \mu + 0.1\} &= P\{\mu - 0.1 - \mu < \bar{X}_{36} - \mu < \mu + 0.1 - \mu\} \\&= P\left\{\frac{-0.1}{0.5/6} < \frac{\bar{X}_{36} - \mu}{0.5/6} < \frac{0.1}{0.5/6}\right\} \\&= P\{-1.2 < Z < 1.2\} \\&= \Phi(1.2) - \Phi(-1.2) = 0.769\end{aligned}$$

Sommario

- La media del campione
- La legge dei grandi numeri
- Il Teorema del Limite Centrale