cioè in altre parole i valori (deterministici) presi dal polinomio trigonometrico si distribuiscono come una Gaussiana standard: facendo un plot di questi valori si otterrebbe la classica curva a campana. (Per alcune referenze relative a questo risultato, si veda ad esempio il classico articolo R. Salem and A. Zygmund. "Some properties of trigonometric series whose terms have random signs". In: Acta Math. 91 (1954), pp. 245–301, oppure i più recenti Jürgen Angst and Guillaume Poly "Variations on Salem–Zygmund results for random trigonometric polynomials: application to almost sure nodal asymptotics". In: Electronic Journal of Probability 26 (2021), pp. 1–36, e Louis Gass, Almost-sure asymptotics for Riemannian random waves. In Bernoulli 29, (2023), no.1, 625–651.) Un risultato analogo, ma forse più semplice da enunciare, è il seguente: sia n_k *una sequenza di numeri interi tale per cui* $1 < n_{k+1} - n_k = O(1)$; *definiamo il polinomio trigonometrico deterministico*

$$q_n(x) = \frac{\sqrt{2}}{\sqrt{n}} \sum_{k=1}^{n} \cos(n_k x) \ .$$

Allora quando $n \to \infty$ *si ha*

$$\frac{1}{2\pi} Misura \{x : c_1 \leq q_n(x) \leq c_2\} \to_{n\to\infty} \int_{c_1}^{c_2} \frac{\exp(-x^2/2)}{\sqrt{2\pi}} dx = \Pr\{Z \in [c_1, c_2]\} \ .$$

Quindi i valori deterministici del polinomio $q_n(x)$ *si distribuiscono seguendo una curva Gaussiana standard. Per questo risultato si veda ad esempio Katusi Fukuyama, A central limit theorem to trigonometric series with bounded gaps, Probab.Theory Related Fields, 149, no. 1-2, 139–148, (2011).*

7 Il teorema del limite centrale

La formulazione più semplice del teorema limite centrale che si incontra in un primo corso di probabilità è la seguente.

Theorem 64 *Sia* $X_1, ..., X_n$ *una seuenza di variabili aleatorie indipendenti ed identicamente distribuite con valor medio* μ *e varianza (finita)* σ^2. *Allora abbiamo la convergenza in distribuzione*

$$\frac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n\sigma^2}} \to_d Z \sim N(0,1) \ , \text{ per } n \to \infty \ .$$

Proof. Ricordiamo innanzitutto la definizione di funzione caratteristica di una variabile aleatoria X :

$$\psi_X(t) := E[\exp(iXt)] \ .$$

Ricordiamo altri due fatti dai corsi elementari di probabilità: Il Teorema di continuità di Lévy ci garantisce che se la sequenza di funzioni di caratteristiche $\psi_{X_n}(t) = E[\exp(iX_n t)]$ converge a $\psi_X(t) = E[\exp(iXt)]$, necessariamente si ha la convergenza in distribuzione $X_n \to_d X$ (e viceversa). Ovviamente nel nostro caso è sufficiente dimostrare che

$$\frac{\sum_{i=1}^n \widetilde{X}_i}{\sqrt{n}} \to_d Z \sim N(0,1) \text{ , per } n \to \infty \text{ , con } \widetilde{X}_i := \frac{X_i - \mu}{\sigma} \ .$$

Abbiamo dunque, usando le proprietà standard della funzione caratteristica

$$\begin{aligned} \psi_{\frac{\sum_{i=1}^n \widetilde{X}_i}{\sqrt{n}}}(t) &= E[\exp(i\sum_{i=1}^n \widetilde{X}_i \frac{t}{\sqrt{n}})] \\ &= \psi_{\sum_{i=1}^n \widetilde{X}_i}(\frac{t}{\sqrt{n}}) \\ &= \psi_{\widetilde{X}_1}(\frac{t}{\sqrt{n}}) \times ... \times \psi_{\widetilde{X}_n}(\frac{t}{\sqrt{n}}) \\ &= \left\{\psi_{\widetilde{X}_1}(\frac{t}{\sqrt{n}})\right\}^n , \end{aligned}$$

dove nel secondo passaggio abbiamo usato il fatto che le variabili sono indipendenti e nel terzo il fatto che sono identicamente distribuite. Ricordiamo ora che

$$\psi_{\widetilde{X}_1}(t) = 1 + itE[\widetilde{X}_1] + \frac{(it)^2}{2}E[\widetilde{X}_1^2] + o((it)^2) \ ,$$

da cui

$$\begin{aligned} \left\{\psi_{\widetilde{X}_1}(\frac{t}{\sqrt{n}})\right\}^n &= \left\{1 + i\frac{t}{\sqrt{n}}E[\widetilde{X}_1] + \frac{(it)^2}{2n}E[\widetilde{X}_1^2] + o((\frac{it}{\sqrt{n}})^2)\right\}^n \\ &= \left\{1 - \frac{t^2}{2n} + o(\frac{t^2}{n})\right\}^n \to \exp(-\frac{t^2}{2}) \ . \end{aligned}$$

La dimostrazione è conclusa ricordando che la Gaussiana standard ha proprio funzione caratteristica

$$\psi_Z(t) = \exp(-\frac{t^2}{2}) \ .$$

■

Nell'ambito di questo corso, avremo bisogno di un Teorema del Limite Centrale di portata ben più generale. A tale fine, introduciamo la cosiddetta condizione di Lindeberg.

Condition 65 *(Lindeberg) Sia $\{X_i\}$ una successione di variabili inidpendenti (non necessariamente identicamente distribuite) definite su uno spazio di probabilità $\{\Omega, \Im, P\}$ e con momento secondo finito; scriviamo $\mu_i = E[X_i]$ e $\sigma_i^2 = Var[X_i]$. Questa sequenza soddisfa la condizione di Lindeberg se per ogni $\varepsilon > 0$ si ha*

$$\lim_{n\to\infty} \frac{1}{\sum_{i=1} \sigma_i^2} \sum_{i=1}^n E[X_i^2 I_{\left\{X_i^2 > \varepsilon \sqrt{\sum_{i=1} \sigma_i^2}\right\}}] = 0 \ .$$

Theorem 66 *Sia $\{X_i\}$ una successione di variabili aleatorie che soddisfa la condizione di Lindeberg. Allora*

$$\frac{\sum_{i=1}^{n} X_i - \sum_{i=1}^{n} \mu_i}{\sqrt{\sum_{i=1}^{n} \sigma_i^2}} \to_d Z \sim N(0,1) \ .$$

Remark 67 *La dimostrazione di questo risultato si basa su approssimazioni successive della funzione caratteristica e non è riportata per brevità. E' però interessante capire il significato della condizione di Lindeberg, che intuitivamente tende ad escludere due casi:*

a) quello in cui la somma delle varianza NON diverga, cioè $\sum_{i=1} \sigma_i^2$ vada ad una costante. In questo caso stiamo in pratica sommando solo un numero finito di variabili aleatorie e pertanto il TLC non può valere, a meno che le variabili non siano già in partenza Gaussiane

b) quello in cui la varianza nella coda delle ultime variabili aleatorie sommate sia dello stesso ordine di grandezza di tutte le altre; anche in questo caso il TLC non può valere perché è come se sommassimo due singole variabili, una identificata dalla somma delle prime $n-1$ e l'altra costituita dall'ultima.

Remark 68 *La condizione di Lindeberg è difficile da verificare in pratica e si usa più spesso una meno generale, ma di più facile verifica, la cosiddetta condizione di Lyapunov.*

Condition 69 *(Lyapunov) Sia $\{X_i\}$ una successione di variabili aleatori indipendenti ed a media nulla e definiamo*

$$\widetilde{X}_{i,n} = \frac{X_i}{\sqrt{\sum_{i=1}^{n} E[X_i^2]}} \ ,$$

in modo tale che $\sum_{i=1}^{n} E[\widetilde{X}_{i,n}^2] = 1$ per ogni n. La sequenza soddisfa la condizione di Lyapunov se

$$\lim_{n\to\infty} \max_{i=1,\ldots,n} E[|\widetilde{X}_{i,n}|^{2+\delta}] = 0 \text{ per qualche } \delta > 0 \ .$$

Remark 70 *Anche la condizione di Lyapunov può essere vista come la richiesta che non ci sia nessuna variabile aleatoria con varianza non trascurabile rispetto alla somma di tutte le altre.*

Remark 71 *E' importante discutere la relazione esistente tra teorema del limite centrale e legge dei garndi numeri. Consideriamo il caso di variabili indipendenti X_i, e normalizziamole in modo che abbiano valor medio nullo e varianza 1. Sappiamo che la somma di questa variabili aleatorie ha varianza $Var[\sum_{i=1}^{n} X_i] = n$, e pertanto $\sum_{i=1}^{n} X_i = O_p(\sqrt{n})$; in altre parole questa sequenza di somme se divisa per $\sqrt{n}$ vivrà su un compatto con probabilità che può essere resa arbitrariamente vicina ad uno. la legge dei grandi numeri ci dice che dividendo per n abbiamo*

$$\overline{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i \to_p 0 \ ;$$

in un certo senso, il Teorema del Limite Centrale si può pensare come uno zoom sulle fluttuazioni che diventano infinitesimali della media intorno allo zero: riscalando per un fattore $\sqrt{n}$ otteniamo

$$\sqrt{n}\overline{X}_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i \to_d N(0,1) \ .$$

Quindi le fluttuazioni nella somma di n variabili aleatorie indipendenti a valor medio nullo sono di ordine $\sqrt{n}$; le fluttuazioni della media aritmetica sono di ordine $\frac{1}{\sqrt{n}}$; le fluttuazioni nel teorema del limite centrale sono di ordine $O(1)$.

Example 72 *(L'uso del Teorema del limite centrale nei sondaggi di opinione.)...*